UNIVERSITY OF SOUTHAMPTON

Towards an Understanding of Web Growth: An Empirical Study of Socio-Technical Web Activity of Open Government Data

by

Ramine Tinati

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Physical and Applied Sciences
Web and Internet Science

December 2013

UNIVERSITY OF SOUTHAMPTON

Abstract

Faculty of Physical and Applied Sciences
Web and Internet Science

Doctor of Philosophy

by Ramine Tinati

This thesis proposes a new interdisciplinary approach to understanding how the World Wide Web is growing, as a socio-technical network, co-constructed by interrelationships between society and technological developments. The thesis uses a longitudinal empirical case study of Web and offline activity surrounding the UK Open Government Data community to explore the Web as a socio-technical 'networks of networks'. It employs a mixed methods framework, underpinned by sociological theory but also drawing on computer science for technical approaches to the problem of understanding the Web. The study uses quantitative and qualitative sources of data in a novel analysis of online and offline activities to explore the formation and growth of UK Open Government Data and to understand this case, and the Web itself. The thesis argues that neither technology nor 'the social' alone is sufficient to explain the growth of this network, or indeed the Web, but that these networks develop out of closely co-constructed relationships and interactions between humans and technology. This thesis has implications not only for how the Web is understood, but for the kinds of future technological design and social activity that will be implicated in its continued growth.

Declaration of Authorship

I, Ramine Tinati, declare that this thesis titled, **Towards an Understanding of Web Growth: An Empirical Study of Socio-Technical Web Activity of Open Government Data** and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
- Parts of this work have been published as:

Tinati, Ramine, Halford, Susan (2013) Big Data: Methodological Challenges and Approaches for Sociological Analysis. Journal of Sociology (Accepted, Expected 2013)

Tinati, Ramine, Carr, Les, Halford, Susan and Pope, Catherine (2013), Exploring the Use of #OpenData in UK Open Government Data Community. Digital Economy 2013, Salford, UK

Tinati, Ramine, Carr, Les, Halford, Susan and Pope, Catherine (2013), The Promise Of Big Data: New Methods For Sociological Analysis. The World Social Science Forum, Montreal, Canada

Hall, Wendy, Tiropanis, Thanassis, Tinati, Ramine Booth, Paul, Gaskell, Paul, Hare, Jonathon and Carr, Leslie (2013), The Southampton University Web Observatory. In the Building Web Observatories Workshop, Fifth Annual ACM Web Science Conference (WebSci2013). Paris, France

Ramine, Tinati, Tiropanis, Thanassis , Carr, Les (2013) An Approach for Using Wikipedia to Measure the Flow of Trends across Countries. In the Web Observatories Workshop (WOW), WWW2013: 22nd International World Wide Web conference, 2013. Rio de Janeiro, Brazil

Tinati, Ramine, Carr, Les, Halford, Susan and Pope, Catherine (2013), The HTP Model: Understanding the Development of Social Machines. In the SOCM Workshop, WWW2013: 22nd International World Wide Web conference, 2013. Rio de Janeiro, Brazil

Tinati, Ramine, Carr, Les (2012) Understanding Social Machines. Doctorial Consortium, International Conference on Social Computing (SocialCom), 2012. Amsterdam, Netherlands

Tinati, Ramine, Carr, Les, Halford, Susan and Pope, Catherine (2012) Exploring the Impact of Adopting Open Data in the UK Government. Digital Futures 2012. Aberdeen, Scotland

Tinati, Ramine and Halford, Susan (2012) Interrogating Big Data for Social Scientific Research - An Analytical Platform for Visualising Twitter. Internet, Politics, Policy 2012: Big Data, Big Challenges?, Oxford, UK

Tinati, Ramine, Halford, Susan, Carr, Les and Pope, Catherine (2012) Mixing Methods and Theory to Explore Web Activity. Third Annual Web Science Conference (WebSci2012), Evanston, US

Tinati, Ramine, Carr, Les, Hall, Wendy and Bentwood, Johnny (2012) Scale Free: Twitter's Retweet Network Structure. Network Science 2012, Evanston, US

Tinati, Ramine and Carr, Leslie (2012) ReFluence: A Real-Time and Historic Visualization Application for Twitter Conversations. in the Social Media Visualisation (SocMedVis) Workshop, International Conference on Web Logs and Social Media, 2012. Dublin, Ireland

Tinati, Ramine, Halford, Susan, Carr, Leslie and Pope, Catherine (2012) Using Mixed Methods to Track the Growth of the Web: Tracing Open Government Data Initiatives. in the Web Science Track, at 21st International World Wide Web conference (WWW2012), 2012. Lyon, France

Tinati, Ramine, Carr, Leslie, Hall, Wendy and Bentwood, Jonny (2012) Identifying Communicator Roles in Twitter. In Mining Social Network Dynamics (MSND 2012), at 21st International World Wide Web conference (WWW2012), 2012. Lyon, France

Tinati, Ramine, Carr, Leslie, Hall, Wendy and Bentwood, Jonny (2012) Identifying User Roles in Twitter. In, Information Communication Technology in Development (ICTD) 2012, Atlanta, US

Tinati, Ramine, Carr, Les, Halford, Susan and Pope, Catherine (2011) Digital Innovation Investigating the Sustainability of New Kinds of Web. In, Digital Engagement 2011, Newcastle, UK

Tinati, Ramine (2011) Tracking the Sustainability of Open Government Data. Open Government Data Camp 2011, Warsaw, Poland

Tinati, Ramine, Halford, Susan, Carr, Leslie and Pope, Catherine (2011) Conceptualising a Web of Linked Actors. In, Second Annual Web Science Conference (WebSci2011), Koblenz, Germany

Tinati, Ramine, Halford, Susan, Carr, Leslie and Pope, Catherine (2011) Exploring the UK Open-PSI community. In, SharePSI: Removing the Roadblocks to a Pan European Market for Public Sector Information Re-Use, Brussels, Belgium

Tinati, Ramine, Pope, Catherine, Halford, Susan and Carr, Leslie (2010) (re) Configuring the Web User in the Digital Economy. Digital Futures 2010: Digital Economy All Hands Meeting, Nottingham, UK

Signed: ______________________________

Date: ______________________________

Acknowledgements

I would like to thank my PhD supervisors, Les, Susan, and Cathy as well as the other academics, researchers, and fellow PhD students in WAIS for helping me along this journey, your guidance has been invaluable and without your support this thesis would have not been possible. You have all been instrumental in my Web Science training, teaching me skills that have proved to be essential during the last three years. I want to thank my family and friends (too many names to list, but I'm sure you know who you are) for putting up with my constant and never-ending (one-way) discussions about my work, especially for the times when I kept on about analytics, networks, and how everything can be understood as an 'Actor-Network'.

I would like to thank all the participants that took part in my study; thank you for spending time responding to my emails, forms, and providing me with valuable knowledge during the interviews.

Finally, I would also like to thank the Web Science Doctoral Training Center (or the 'DTC') for providing me the opportunity (and funding) to undertake this research, and also all the staff that work at the DTC 'HQ' for helping me out with anything and everything that I required, no matter how trivial!

Thank you.

Contents

List of Figures

List of Tables

Abbreviations

ANT	**A**ctor **N**etwork **T**heory
API	**A**pplication **P**rogramming **I**nterface
AJAX	**A**synchronous **J**avascript **A**nd **X**ML
BA	**B**arabasi-**A**lbert Model
CERN	**E**uropean Organisation for **N**uclear **R**esearch
CKAN	**C**omprehensive **K**nowledge **A**rchive **N**etwork
CSV	**C**omma **S**eparated **V**alue
GMOD	**G**reater **M**anchester **O**pen **D**ata
HTTP	**H**ypertext **T**ransfer **P**rotocol
HTML	**H**ypertext **M**ark-up **L**anguage
IP	**I**nternet **P**rotocol
ISR	**I**nformation **S**ystems **R**esearch
OD	**O**pen **D**ata
OGD	**O**pen **G**overnment **D**ata
OPP	**O**bligatory **P**assage **P**oint
PDF	**P**ortable **D**ocument **F**ile
SCC	**S**trongly **C**onnected **C**entre
SCOT	**S**ocial **C**onstruction **O**f **T**echnology
SNA	**S**ocial **N**etwork **A**nalysis
SNS	**S**ocial **N**etwork **S**ystem
STS	**S**cience and **T**echnology **S**tudies
TCP	**T**ransmission **C**ontrol **P**rotocol
UK	**U**nited **K**ingdom
US	**U**nited **S**tates
URI	**U**nique **R**esource **I**ndicator

URL	**U**nique **R**esource **L**ocator
USB	**U**niversal **S**erial **B**us
W3C	**W**orld **W**ide **W**eb **C**onsortium
WWW	**W**orld **W**ide **W**eb
XML	E**X**tensible **M**ark-up **L**anguage

Chapter 1

Framing the World Wide Web

1.1 Preamble

In the last 20 years, the World Wide Web has grown from zero to 3.8 billion pages (Kunder 2013), used by an estimated 2 billion people (Alexa Internet 2013). It is now an integral part of human activity and society; from helping humans choose a suitable restaurant, to supporting billion dollar transactions between global corporations. Yet despite its integration with multiple social activities, there is very little understanding of the Web beyond a technical understanding of its architecture, technology, and infrastructure. This thesis seeks to answer a number of critical questions about the growth of the World Wide Web and aims to develop a theoretical approach and analytical framework to explore how the Web has grown, and continues to grow.

This thesis asks challenging and critical questions about an important and influential innovation of modern times. These questions require critical thinking about the complex relationship between the social and the technical. The Web is not only a technical infrastructure, but also a complex set of social practices. Thus the roots of this thesis are necessarily situated at an interdisciplinary crossroad labelled Web Science. Web Science draws upon the theories, concepts, techniques, and capabilities of a range of academic disciplines, in order to develop an understanding of the Web that transcends disciplinary boundaries and their respective analytical and theoretical capabilities. Despite the challenges of interdisciplinary research, through practice and discussion, this thesis aims to demonstrate how interdisciplinarity offers a richer understanding than individual disciplines alone are able to achieve.

This thesis strives to contribute to Web Science research, by asking the fundamental questions: what is the Web and how does it grow? To begin, this chapter will provide an introduction to establish an understanding of the Web and its core technologies, and its relationship with society and the individual.

1.2 An Introduction to 'The Web'

The 'World Wide Web', 'The Web' or WWW, transcends linguistic, geographic, social, cultural, and political borders, and despite being underpinned by common technological principles and protocols, it does not prescribe what it can or cannot be used for. The Web is a ubiquitous technology; it may be accessed via a desktop computer, a smart phone, or embedded into the latest home kitchen appliance. Along with being technologically platform independent, it is unrestricted in terms of purpose and use: it might be used for research and innovation, for finance and business, for helping support political revolutions, or even for criminal and unlawful behaviour. In the last two decades, the Web has been become a technology that has been used by over 2 billion humans; this has certainly outgrown its origins from the physics laboratories where it was first conceived and reaches far beyond the original community of scientific users (Berners-Lee 1988).

Society is permeated with digital technology (Lankshear & Knobel 2008). Human interactions are responsible for disrupting and redefining the functionality and capabilities of technology; how it is used, modified, adopted, or dismissed, which as a consequence directly shape it future. The Web shares many similarities with other technologies, its development is part of this on-going evolution, and as more humans interact with it, and as the networks of technologies that surround and support it change, so does the Web. Yet unlike other technologies, the scale and impact of the Web has happened in ways unexpected (Hall et al. 2012).

The impact of the Web can be witnessed at both the micro and macro level. The Web is both an individual and collaborative system that reaches far beyond just online interactions. At the micro level, the Web provides new ways to conduct day-to-day activities, whether as an individual or as a collective. Similarly, at the macro level, the Web is integral to the running of global economies, the operation of industry, the success of commerce, and the powers of government (Yu & Robinson 2012). Yet, this is all part of one system, there are no barriers or borders between the individual and the collective, between the micro and the macro, it is just a matter of its configuration. As a consequence of this boundary-free capabilities, information can now be exchanged faster, knowledge can be found (almost) everywhere (Weinberger 2012), and communications between humans (and machines) are no longer restricted by physical boundaries of location or time. The Web is changing the way humans interact at both the level of the individual and the collective.

The Web is not one thing, but many; yet despite this, it is defined by the World Wide Web Consortium (W3C) simply as "an information space" with global identifiers and Uniform Resource Identifiers (URIs) (World Wide Web Consortium (W3C) 2009). The W3C's abstract technical definition does not encapsulate any of the aspects of the interaction between humans and the Web. How can this narrow definition of an information space explain the Web's rapid growth and adoption by humans within the last two decades, and how can just an "information space" have such an unpredictable impact on the activities and operation of modern society (Hall 2011)? Technology alone was not responsible for this, the Web is more than just a set of protocols and documents; it is a space of social activity exempt from traditional physical constraints.

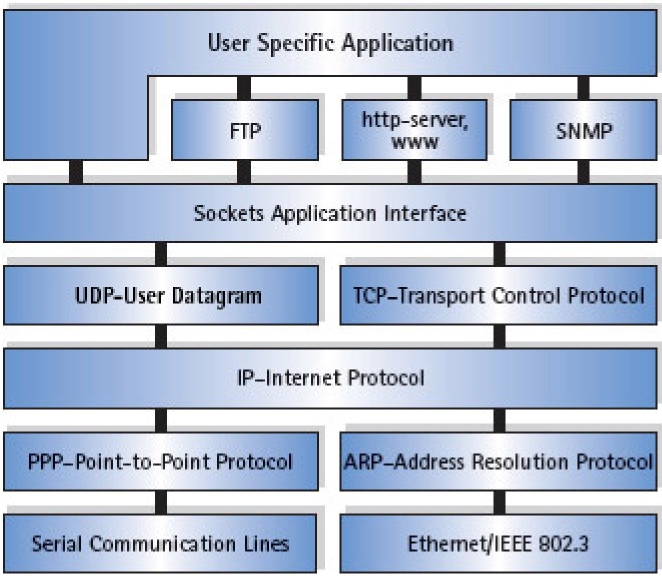


FIGURE 1.1: TCP Service Stack

Whilst the Web is used for many different activities and purposes, technologically, the original Web's architecture and protocols developed in 1989 remain to the present day (Berners-Lee 1988). The technical architecture that the Web uses is simple yet elegant, harnessing TCP/IP of the Internet as a technical layer for data exchange. The Transmission Control Protocol (TCP) and the Internet Protocol (IP) provide the technical end-to-end standards and formats for how data should be transmitted and received. Within this protocol there are four different layers which are used to pipe and control different network traffic, and sitting within the top 'application later' is the core Web protocol known as the Hypertext Transfer Protocol (HTTP), which enables information, written in Hypertext Mark-up Language (HTML), to be transmitted between a client and server. As Figure 1.1 illustrates, the HTTP protocol sits on top of the TCP/IP layer, and offers a new communications protocol to transmit documents. Based on HTTP, marked-up information (the resource)in the form of HTML can be hosted on a server, and then accessed by a client via the use of specific Web browsing software. The Web was developed to be scalable and technically heterogeneous, independent of operating systems or hardware architectures (Berners-Lee 1988). Initially, the Web provided a platform for a known number of scientists in an enclosed and controlled environment, with a known set of technologies to share scientific documents between their computer terminals. Twenty years later, the Web provides a platform for a worldwide network of billions of different electronic devices (computers, phones, televisions, etc.), enabling individuals (not just scientists) to share, view, create and modify diverse sources of information. The technology that underpins all this might be shared, but its use and purpose, is not.

1.2.1 The Prerequisites for a Web

The motivation to develop the Web can be accredited to Sir Tim Berners-Lee during his employment at the European Organisation for Nuclear Research (CERN); an international organisation supporting various large-scale, multi-national collaborative scientific research projects. CERN was a focal point for a complex laboratory featuring scientists from across the world, and exchanging data over a network (which at the time used their own variation of the Internet) was the only practical solution to sharing their research. During the late 1980's, the social and working conditions within CERN meant that lots of data was being created by different teams working on the same experiments, however sharing information was challenging. Unlike computing systems and technologies that are available as of today, there were no USB Flash Drives, no wireless, no Bluetooth, even CD's were not in use. Coupling this with diverse and incompatible Operating Systems, hardware architecture and programming languages led to a working environment with many constraints. Aware of these limitations, Tim Berners-Lee worked within the constraints of CERN and created a technical blueprint which provided a simple yet scalable solution to share information (Berners-Lee 1988). This system, which later became known as the World Wide Web, provided a technical solution to solve CERN's information problem; however, unknowingly, it also had application far beyond a scientific community of users (Berners-Lee 1999).

The (CERN-based) Web was not just an "information space". This Web was both a response to social requirements of collaborators on large-scale projects and the technological capabilities and skills of CERN's infrastructure. These social and technical requirements helped the Web at CERN become a functioning computer-based, networked information system. It was a collection of technologies and a set of new social practices of the scientists around sharing their data. As the technologies developed and improved, so did the Web's social reach, with non-scientists adopting the social practices pioneered at CERN. The Web grew, both technically and socially. The Web was no longer contained at CERN; other communities such as academic researchers and e-commerce used it for their own purposes, which led to the emergence of their own set of social and technical practices such as sharing scholarly research and papers, or adapting traditional offline social practices of shopping and selling products as a Web activity. These new social activities were also supported by the development and advances in technology.

During the era of the 'CERN-Web', the transmission of information was limited to text and images (of limited size), not least because of connection bandwidth and computer hardware limitations. However, since this time, computer hardware has increased in processing speed and decreased in cost for both industry and consumers. As a consequence, the availability of Internet-enabled computational equipment has increased, the price of high-speed internet connections has decreased, and connection bandwidth has improved dramatically. The technical stack shown in Figure 1.1 is now able to transmit far richer forms of information, including: videos, music, games, and TV. All these factors have contributed to the growth of the Web. In line with the improvements in computer software and hardware performance, new frameworks, developer tool kits and programming languages have emerged, offering software developers tools to create Web services with features that the original specification did not provide. Web technologies such as HTML has matured to HTML5, offering new features that give more power to the developer and more functionality for the user. Alongside this, other frameworks such as 'Asynchronous JavaScript and XML' (AJAX) have helped reshape the functionality of a Web page, the browser, and the user's interactions. These technical enhancements enhance the user experience, and the functionality and capabilities of a Web service. This has been a springboard for a new technological landscape and set of affordances for users and developers to create new ways to visualise, share, and broadcast information.

Methods of interacting with the Web have also changed; transitioning from being a technology limited to the confines of a desktop computer connected via a hardwired network, to a technology accessible across many different platforms and hardware devices. Telecommunications technologies such as Wi-Fi and 3G now make it possible to access the Web, with no need for a designated internet connection or cable. The Web is now a technology that is natively supported by (smart) mobile phones, game consoles, and even (smart) televisions. The Web no longer resides as a technology in isolation, but features as a process or component within and between other technologies.

The Web is changing how individuals connect, communicate, and operate. Technologies in past did similar things; for instance, the printing press enabled the distribution of information and knowledge. Yet the Web is different, it does not suffer from temporal and physical limitations, information can now be shared virtually instantly and potentially to a global audience. News which once could only be broadcast to a local audience, is now geographically unbound. Knowledge, which was once buried in encyclopaedias, can now be obtained via Web search engines and collaborative driven knowledge-bases, and everyday activities such as shopping, banking, and entertainment are all possible on the Web. The Web has become a "do-everything tool" (Beauvisage et al. 2007).

The capabilities of the Web are facilitating existing human activities and also providing a technological platform for new forms of interaction and activities to emerge and grow. However, all these activities are united by their use of the underlying Web architecture which is not restrained by geography or scale.

1.3 A Social and Technical Web

From examining the emergence and development of the early Web, it has become apparent that it is not just a technology and to define it in such a way does not capture its social and technical complexity. The development of the Web is a result of process involving technological development and human interaction. Even the development of the Web's architecture underwent these processes; Tim Berners-Lee, the physicists, CERN, the data, they are all part of complex set of interactions. Tim Berners-Lee may have created the HTTP because there was a problem with sharing data, but this was only due to the creation of large sets of data and documents, which was a result of a multinational collaboration between scientists, which itself was a consequence of a shared interest in advancing scientific knowledge. The Web was also subjected to policies and administrative decisions made by the directors of CERN; the same decisions that provided Tim Berners-Lee with the extra resources to work on developing the HTTP. A whole chain of events and interactions lead to 'the Web' being developed.

The Web is part of a highly intricate network of technological developments while at the same time it is part of a larger network of social developments. The Web has not developed in isolation from other technologies and scientific innovations. The Web is a "social machine" (Berners-Lee 1999) (Hendler et al. 2008), which has developed as part of the changes in society. From one perspective, the Web can be considered as the reflection of human creativity and change. From another perspective, the Web can be considered as a technical network of electronic devices communicating and sharing bits of digital data. Whilst both perspectives for describing the Web are correct and required, neither alone help explain what the Web is, or how it has grown.

Hopefully by now, the complexity of the Web and the difficulty in describing it has become apparent. The Web and society mutually shape each other (Halford et al. 2009), and its development has not been in separate from social factors and change, but because of, and constitutive of them. The Web needs to be understood from both perspectives; on the one hand being able to comprehend its technical design, on the other hand, comprehending its social interactions. However, these are not parallel activities, they need to be integral to each other. Taking this into consideration, what theories, methods and approaches can be used to help understand the Web as a co-constituted phenomenon, where the social and technical can be considered simultaneously?

1.4 What is the Web? Framing the Question

Understanding the Web is a complex task, it requires an understanding of intricacy of human and machine interactions. Even with the lightest of touches, describing the processes and changes that have surrounded the growth of the early CERN-Web to the current Web is a somewhat overwhelming task, which in turn raises more questions concerning how this growth occurred, and how the interactions between the social and technological shaped outcomes. However, whilst there are numerous questions that can be asked about the relationship between the Web and society, before these can be considered the fundamental questions of what is the Web and how does it grow, need answering. Understanding these core questions which are also the core drivers for the emerging field of Web Science (Hendler et al. 2008) provide the foundations of a co-constructed perspective of Web growth for studies to build upon. Once these questions are answered, practical questions about the Web can be asked, questions which will be able to challenge the current understanding of the relationship between the Web and increasingly relevant topics such as governance, policy, politics, economics, communications and collaboration.

This chapter has argued that the Web is not something that can merely be described or understood as an "information space" of documents connected to each other via URIs, and nor can it just be understood a social phenomenon. It is too simplistic to suggest that the Web managed to acquire 2 billion 'users' and over 3 billion Web pages simply because people 'used' it. There is something more fundamental to understanding its adoption, and this thesis wishes to uncover it. This therefore raises the question which this thesis wishes address: what is the Web and how does it grow. To answer these questions the thesis needs to also know what theories and methods to use, and therefore will explore what, if any, theories and methods can be used to deconstruct and explain the Web in terms of the social and technical processes that have enabled it to grow.

1.5 Thesis Outline

This chapter has begun questioning the current understanding and conceptualisation of the Web. However, this has only touched upon the surface. In order to explore theory thoroughly of what is the Web and how it grows, a more detailed exploration into the understandings offered by different disciplinary approaches to the Web is required. As a way into this, chapter 2 will investigate what computer science and social science perspectives offer and demonstrate that, although they have their own strengths and analytical advantages, in isolation, they are not adequate to understand the Web as both social and technical.

Based upon the issues identified in chapter 2, chapter 3 will argue for the adoption of an interdisciplinary approach to understanding the Web, and demonstrate that through the lens of an interdisciplinary perspective, social and technical processes that were once hidden can be revealed and exposed.

Chapter 4 will build upon what has been learnt about understanding the Web through an interdisciplinary perspective and discuss the methodology and case design which will provide the foundations for answering the research questions. Based upon the case design, chapter 5 will introduce the chosen case study, and the methods and analytical techniques which make use of computer science and social science approaches in order to provide an analytical framework to understand the Web and how it grows.

In order to demonstrate the analytical capabilities of this interdisciplinary approach for understanding the Web and how it grows, chapter 6 will provide the first level of analysis drawing on the case study data. The analysis provides the foundations for chapter 7 and 8 to examine in greater depth the underlying processes that are embedded within the growth and development of the Web.

Reflecting on the findings of the analysis, chapter 9 will synthesise the arguments about the underlying social and technical processes in Web growth and develop a theoretical model which helps describe Web growth, and Web structures.

Finally, chapter 10 reflects on the implications of this research for the computer science, social science, and Web science communities, and addresses the debate about the drive towards interdisciplinarity, and what affordances will the findings of this thesis provide for future research.

Chapter 2

A Current Understanding of the Web

2.1 Disciplinary Perspectives

This chapter will explore literature that has helped create and define our understanding of the Web, and will demonstrate that an interdisciplinary approach is required to understand both its social and technical components. This will initially draw on studies from computer sciences and which have focused on analysing and interpreting the structures and characteristics of the Web and then will examine theories from the social sciences that offer a way to understand human interaction.

The previous chapter framed the Web as complex, not only in terms of the technological (both software and hardware) infrastructure, but also in terms of the human interaction and activity that uses and develops this. The Web (as a social and technical entity) interacts with traditional social structures, social practices or 'fields' (Bourdieu 1984) within and across different societies and countries; these are all part of the network. In light of this, describing the Web requires an understanding of its complexity in terms of both the human and the technical components of the Web.

As a starting point, the following section will examine computational science literature concerned with understanding the structure and growth of the Web. This chapter will also explore how more recent computer science research has drawn upon concepts and theories from other academic disciplines including network science and sociology, in order to capture the idea of the Web as a social and technical phenomenon.

2.2 Understanding the Web from a Computational Perspective

From a computational perspective, the Web represents a network of documents, data, protocols and interfaces (physical and virtual); it also represents a technical feat of designing scalable and reliable systems. This discussion will focus on the literature that has pioneered techniques for understanding the Web, building on mathematical research in graph theory to describe and analyse network structures at different scales (Fiedler 1973) (Kernighan & Lin 1970) (Pothen et al. 1990). Early on in the Web's existence, Broder et al. (1997) explored ways to group coherent Web pages based on similarities on text content. This research was one of the earliest studies to discuss the Web in terms of groups of syntactically related documents. Effectively Broder *et al.* attempted to map the Web as a single network of documents, rather than discrete pieces of information and in doing so revealed a network which was not uniform in structure, or document content. This exposed the scale of the Web in terms of the number of documents and links being created. Building upon this research, Gibson et al. (1998) adopted this approach and developed it further; creating techniques and models to understand the structure and growth of the Web in respect to the hyperlinks formed between Web documents, reconfirming the distributed network structure and the varying content, size, and links of Web documents.

Although these studies were ground-breaking, their aim was to understand and model the network structures that emerged from the connections of Web documents, rather than the human interactions that were responsible for the Web graph's structure. Indeed, observations of the social were non-existent. The Web was studied as a system where documents were linked together by hyperlinks, and the groups of documents that these links formed led to the identification of hyper linked communities. The extent to which the social was implicated in the Web was ignored.

Studies continued to examine the Web's structure from a quantitative perspective, identifying features such as the bow-tie shape of the Web (Broder et al. 2000), a finding which demonstrated that the Web graph could be split into a dominant strongly connected centre (SCC), and groups of nodes (Web documents) which are connected together by the SCC. Similarly, research at the same time were interested in the structures that formed within the Internet routing topology, which led to the discovery of power-law distribution (Faloutsos et al. 1999). Similar to these findings, the Web's structure was explored in more detail, revealing a topology which had an uneven distribution of links between Web documents; and that the structure was not as chaotic as once previously thought (Gibson et al. 1998). These discoveries led to the notion of hubs and authorities, which represented the Web pages that either had a high or low proportion of inbound hyperlinks (in-degree) (Kleinberg 1999) (Kleinberg & Lawrence 2001). The identification of these new structures suggested that this type of quantitative analysis could provide an insight into the communities that are forming on the Web, particularly networks around specific topics and interests.

The findings of Web structure provided the basis for the development of search algorithms which identified highly connected documents and determined whether they are hubs or authorities Kleinberg (1999). These findings also led to the development of the popular PageRank algorithm which used the Web's structure to rank pages according to their reputation, calculated by a number of parameters related to the Web graph (Page et al. 1998). Effectively, research in this field helped establish a quantitative path for Web-based analyses, providing the groundwork for studies to develop algorithms and approaches to investigate the Web's topology and structure as a quantitative phenomenon.

In a bid to step beyond understanding the Web as just a set of digital documents, Chakrabarti et al. (1999) developed an alternative description of the Web graph, describing the hyperlinks between pages as the social links that represented different topics of interest in the Web graph. Clustering algorithms such as 'Clever' (Chakrabarti et al. 1998) and others (Huang & Lai 2003) were developed to explore the Web graph as social links, providing a method of information discovery and categorisation, which effectively produced a taxonomy of different online Web activities. The classification relied on network metrics such as the number of inbound and outbound links, yet paid little (if no) attention to what these communities represented in terms of their content. Effectively the social links that Chakrabarti et al. and Huang and Lai described were illustrating the Web as a system of multiple kinds of activities, which were connected and interconnected by manually created hyperlinks.

Considering that the aforementioned research was uncovering phenomena that was the result of the social and technical interaction, the research did not attempt to address or reflect on these findings; instead it focused on quantitative metrics and technological developments. Whether this was due to not being equipped with theories that could offer to expose the social elements of their observations or simply due to disciplinary interests is unknown, however, human interactions are central to these hubs and authorities. They emerge not from technical design or restraint, but due to the activities and interactions of the humans using the technology. Underlying the observations and hierarchical rankings of Web pages are different social processes that enabled these pages to gain differentiated popularity.

The 'shape' of the Web structure offered an observational window to expose human and technical interaction even if there was little attempt in computer science to understand the social components that enabled these processes to occur. The social links that Chakrabarti described were the outcomes of humans connecting Web pages together in virtual (hyper) space. The observations of the hyperlinks were traces of the physical social actions of actors interacting and working together.

2.2.1 Web Growth Modelling

Whilst computer science pursued the development of algorithmic techniques to categorise and provide a structural understanding of the Web, scholars were seeking other methods to understanding the growth of the Web, developing random graph models such as the Barabasi-Albert (BA) model, that help replicate network structures such as the Web (Albert et al. 1999) (Barabasi & Albert 1999) (Watts & Strogatz 1998). Derived from mathematics and physics, these approaches leant towards understanding the Web using mathematical techniques such as topological modelling to expose the patterns and metrics that emerged from its hyperlinked document structure (Erdös & Rényi 1959). The BA model proposed that the growth of the Web exhibited a number of properties in terms of its linking structure, such as linear preferential attachment (Albert & Barabasi 2000) and a power-law degree distribution in terms of the in-bound links of Web documents. Simply put, only a small proportion of Web pages have many URLs, whereas most have only a few.

Based on these findings, other Web growth models were developed such as the random walk model (Jari & Kaski 2004) (Vazquez 2002), the common neighbour model (Girvan & Newman 2002), and the finite memory model (Klemm & Eguíluz 2002), of which all three share the characteristic of network growth and structure introduced by the BA model. Although approached from a different disciplinary perspective, this research was similar to the observations made by Kleinberg (1997) and others which revealed the hubs and authorities structures and the existence of weak ties (Granovetter 1973) that connected different Web communities together. These studies also draw upon network structures, comparing the Web's growth to structures that occur in nature, and contributed enormously to understanding the mathematical structural properties of the Web (Barabasi & Albert 1999). They argued that the Web was a scale-free structure (Albert et al. 1999), and further research by Watts discovered that the Web's structure and growth was similar to other discoveries of small-world networks (Watts 1999).

Despite being able to replicate the properties of scale-free and preferential attachment via algorithmic approaches this research did not reveal the underlying human interactions that formed these structures. These models exposed the processes of human interaction, yet did not comment or pay attention to how these interactions form. Understanding the social structures is not something that can be achieved by just providing statistical measures of a network. However, given the disciplinary perspective associated with these studies, it is understandable why these researchers were not interested in human interaction focused on the models and mathematical forms of the Web's structure.

Undoubtedly, the research conducted was essential to further developing the technical understanding of the Web, which eventually led to the emergence of a discipline know as Network Science (Barabási 2009) (Brandes et al. 2013). Although grounded in mathematical theory, greater emphasis has been placed on studying systems which have implications beyond just technical curiosity (Barabási 2009), albeit, through mathematical and network theory (Brandes et al. 2013). This interest in networks was closely associated with existing research on large scale systems, both man-made and organic. Studies have been conducted on exploring the structure and formation of biological networks such as human protein chains, the formation of friendship networks (Ball & Newman 2013), the fluctuations in financial markets (Eguiluz et al. 2003) (Girvan & Newman 2002) and trading networks (Ward et al. 2013). The studies using these models to understand the Web were far less interested in the specifics of classification and clustering; rather, they were interested in letting the data show the topology and shape of the Web, i.e. revealing the structural properties that underpinned network growth. Network science techniques have provided techniques to observe and model the emergence and quantitative characteristics identified during the growth of a network. Whilst these approaches provide a quantifiable means to explore the growth of the Web, yet they still do not take into consideration the human interactions which have led to the formation and structures they were observing.

2.2.2 Web Communities

The detection of communities is another important area of Web graph research, which draws upon graph theory mathematical theory and techniques, as well as techniques from the social sciences (Kernighan & Lin 1970) (Pothen et al. 1990). Typically, the community detection algorithms work by splitting the network into sub-components, and this process is repeated until the desired number of communities are detected or reached. These types of community detection algorithms either work by examining the graph structure as a whole and find communities by partitioning by removing the 'most important' links between them (Girvan & Newman 2002) (Tyler & Wilkinson, Dennis M. Huberman 2003), or performing local community detection, which use an already defined network of nodes to identify the communities within it (Bagrow & Bollt 2004) (Clauset 2005).

In addition to this work, community detection algorithms have also been developed to examine the textual context of documents and hyperlink structures of the Web graph the activities which they represent (Dourisboure et al. 2009) (Jo et al. 2011). These approaches are similar to the earlier research by Chakrabarti et al. (1998) who detected specific topics within the Web graph by examining the linking structure and content of Web documents. More recent methods used to identify communities used a seeded (starting with a pre-defined list) and unseeded (starting from a blank search) approach, which used Web graph mapping and community detection algorithms, and also concepts and analytical techniques from graph and network theory (Flake et al. 2000) (Flake et al. 2002). By using these techniques, cyber-communities have been labelled and ranked based upon the number of hyperlinks that a Web page obtains and produces(Dourisboure et al. 2009) or by using existing clustering algorithms such as HIT (Imafuji 2003). Other techniques include examining the distribution of words and topics within the Web graph were identified (Jo et al. 2011), which has provided a method to label these topics as different 'Web activities', categorised by terms such as 'General shopping', 'News, 'Sports' and 'Government'. The techniques of studying the Web graph also helped study the 'boundaries' and 'borders' of the Web. Research has not only focused on the complete Web graph (macro), but also the Web graph within geographic or (IP) network boundaries (Han & Lee 2007) (Liu et al. 2005) (Somboonviwat et al. 2008). These studies offer insight into the variance in networks structures and how the Web graph at the micro differs by geographic location or country (Han & Lee 2007) (Zhu et al. 2008).

However, despite these studies describing what the Web is being used for in terms of what content is contained on a Web page, they did not address the issue of what were the social processes that enabled this content to be produced, and what effect this had on the Web. By only engaging with computational methods to understand the observations of the Web's structure, no clear understanding of how the Web is being shaped and is shaping human activity (at the micro and macro level) is offered. The focus and direction of these studies push the study of the Web deeper into the technologically deterministic rabbit hole, yet inadvertently they are describing a Web which is tightly woven by the interactions of humans and technologies. Despite what a macro analysis of the Web graph may show, the Web is not a single coherent network; country-to-country, micro Web graph to micro Web graph, it is structurally and contextually different. The legislation, government, or social conditions within one society may directly affect the use, thus structure of the Web (Zhu et al. 2008). The networks observed all share one common Web protocol, HTTP, yet as findings show, this is only one part of the Web graph.

These findings of social links and varying network structure reinforce the argument that if the Web is to be understood as a co-constructed phenomenon, then methods are needed that can appreciate the social as well as the technical and requires an approach that steps beyond only computational methods. Whilst these methods are essential to understanding the Web at the macro, as the studies have shown, they offer little insight into the social processes or human interactions that occur at the micro level, yet it is these interactions and human activity that create the structural patterns that have been observed in the Web graph.

2.2.3 Web Social Networks

Following the rise in social networking platforms available on the Web, there has been considerable interest from academic researchers. These platforms boast exciting figures with regards to the number of users interacting, offering an alternative domain to apply existing Web related techniques to study human activity and interaction at a scale previously impossible, or impractical to achieve. There exists several threads of research that apply computational techniques to study social networks, including community detection borrowed from Web graph community detection studies (Luo et al. 2006) (Wakita & Tsurumi 2007). These studies use detection techniques to examine the formation of communities of users (rather than documents) within the social networks. Whilst there has been emphasis on developing community detection techniques for use on large-scale social networks (Gargi et al. 2011) (Steinhaeuser & Chawla 2008), detection is problematic due to unpredictable behaviour of users (Leskovec et al. 2010). Another thread of social network research is concerned with the observation of communities, studies borrow techniques from the network sciences, and earlier social network analysis (SNA) techniques from research investigating the associations and communications of offline social networks (Milgram 1967) (Wasserman & Faust 1994). Using these techniques social network studies can be distilled in to research concerned with structure, growth, the activities and characteristics of users (humans and machines), and the observation and modelling of information diffusion.

Early studies of online social networks applied the same techniques of Web graph research and found social network platforms exhibit similar properties as the Web such as small-world characteristics (Adamic et al. 2003) a Strongly Connected Core (SCC) (Meeder et al. 2011), and the formation of strongly connected groups (Girvan & Newman 2002). The structure of social networks were also examined in comparison to the activity patterns of users, illustrating that the structural properties of the static friendship network is far denser than the activity of interactions between users (Hyunwoo et al. 2008) (Wilson et al. 2009). Research also examined the structures within and between different social networks by comparing the connectivity and strength of ties between social networks (Mislove et al. 2007) (Ahn et al. 2007), and the characteristics of these networks in terms of their size and topology (Ahn et al. 2007) (Kwak et al. 2010). These studies tend to adopt network science studies, exploring the topological characteristics of networks, applying the same mathematical models to communications between individuals as the network scientists did between Web documents and hyperlinks. More recent studies have concentrated on user behaviour within and between social networks (Gao et al. 2012) (Hogg & Szabo 2009) (Honeycutt & Herring 2009), and also developing techniques to observe and model information diffusion between communities of users (Leskovec et al. 2009) (Simmons et al. 2011) (Weng et al. 2012). These studies examine how information within a social network is spread amongst different communities, exploring both rate and content of the information (Luu et al. 2011), which have also helped develop techniques for the detection of spam (Chu et al. 2010) (Yardi et al. 2010).

The findings of these studies have been critical in discovering the structural properties of large-scale online social networks and also for discovering and tracing user behaviour and information diffusion. However, as a result of the quantitative focus of this type of research, these studies steer away from thinking critically about the relationship between the social and the technical, and instead focus on the network as metrics, as a structure, or as a static object. Although there are more recent studies that try and engage with understanding human activities, their approaches still are limited to quantitative methods of sentiment analysis and word counting, as seen in studies that try to use social networks to identify political opinion and activism (Zhou et al. 2010) (Conover et al. 2011) (Park & Lee 2011), the identification of public health disorders (De Choudhury et al. 2013) (Paul & Dredze 2011), or identifying the mood of a population (Bertrand et al. 2013) (Lansdall-welfare & Lansdall 2012) (Ritter & Hernandez 2013). Exceptions exist, and these offer a rare glimpse into the underlying social processes that are part of the Web's growth (Boyd et al. 2010) (Christensen 2011). These studies push beyond the quantitative methods and deterministic perspectives often associated with understanding the Web and ask more fundamental questions about why and how users are interacting with each other, and the wider implications of these in terms of their impact on their day-to-day activities as well as society. Yet these studies are certainly not the norm.

In short, within computer science the word 'social' simply highlights the existence of the non-machine components within an information system. By not embracing the networks at the granular level of humans interacting with technology, this type of research can appear to adopt a technically deterministic view of the Web, where the network structures are only understood in metrics, not context. Instead, viewed through a lens that considers the underlying social processes; the associations between nodes, edges, and the network structures, offer an insight into the activities of humans which are forming these complex, dynamic networks.

In fact the methods used in modern SNA are far from their original roots which predate the emergence of computer science. The study of social networks was well-established before computational approaches were available. Previous SNS studies focused on the social characteristics of social networks (Milgram 1967) (Mitchell 1969) (Granovetter 1973). These studies were part of longitudinal ethnographic studies of human interaction and associations, and tended to work with smaller sets of qualitative data sources comprising of interviews and observations (Burt 1978) (Peay 1980). This social science inspired research attempted to understand the social dynamics and formations of networks, emphasising the processes and reasons for human interaction. Unlike computational studies, describing the macro level was only possible through the collection, and analysis of the activities of the humans at the micro level. The small-scale (in comparison to recent studies) qualitative data offered only a partial picture of the network being observed, yet because of this, the data collected and used was highly detailed and provided a rich context to the interactions of humans. However, with the advances in computational techniques and availability of large data sources (Boyd & Crawford 2011) (Manovich 2011), these traditional techniques of SNA have been forgotten in favour of quantitative approaches, which tend to not have the same richness and detail in terms of the context of their findings. Whilst computational methods provide a way to examine the network as a whole, understanding human activity becomes reduced down to quantifiable measures and statistics.

2.3 From a Technical to a 'Socio-Technical' Perspective

Computational methods and techniques provide a way to observe the large-scale structures and mathematical properties of interactions between humans and technologies. However, studying the interactions of humans and technologies at the micro level appears to be essential to understand the underlying social and technical processes.

Recognising that understanding the Web requires a handle on both the technical architecture and social processes between humans and technologies, Web Science scholars have turned to interdisciplinary approaches (Berners-Lee et al. 2005) (Hendler et al. 2008) (Halford et al. 2009). This has led to an acknowledgment of the suitability and need of methods and theories from the social sciences. However, engaging with other disciplines is far more complex than just using the language and methods. Not least embedded within methods are the fundamental principles that often separate the beliefs and underlying epistemological and ontological values of the 'hard' and social sciences (Orlikowski & Baroudi 1991). The 'hard' sciences are often associated with positivism, whereas the social sciences typically are associated with an interpretivist position (Guba & Lincoln 1994), offering different perspectives on how the world is understood, thus requiring different methods and techniques to generate this understanding.

Despite this, recognition of the importance of the interaction of humans and technologies within the Web has led computer science research to adopt the concept of 'socio-technical' in order to describe various aspects of the Web. Socio-technical has been used to describe human activity in areas such as the design and development of technology (Baxter & Sommerville 2011) (Georgas & Sarma 2011), to describe the adoption of Web technologies (Bird 2011) (Hester 2012), and to describe Web systems with human elements (Eschenfelder & Chase 2002) (Slattery 2009) (Whitworth & Liu 2009). This has also been key in Web Science research where understanding the process of the Web (Berners-Lee 2007) requires understanding both the social and the technical, and how they affect and contribute to the 'lifecycle' of Web engineering. As Figure 2.1 demonstrates, in terms of its current application within Web Science literature, at its most basic, the concept of socio-technical offers a way to 'label' and study the activities and interactions of humans and technologies recognised as components of the Web.

As part of the shift towards interdisciplinarity, being aware of the underlying epistemological and ontological differences between the computational sciences and the social sciences is essential to avoid a misunderstanding of the term socio-technical. Web Science increasingly recognises and appreciates the importance of considering the Web as social and technical, whether labelled as a "Social Technology" (Berners-Lee et al. 2006), a "Social Machine" (Berners-Lee 1999), or as a "socio-technical" system (Halford et al. 2009) (Hall & Tiropanis 2012). However, to understand the Web as a socio-technical phenomenon requires more than labelling the social and technical; it is a way to frame

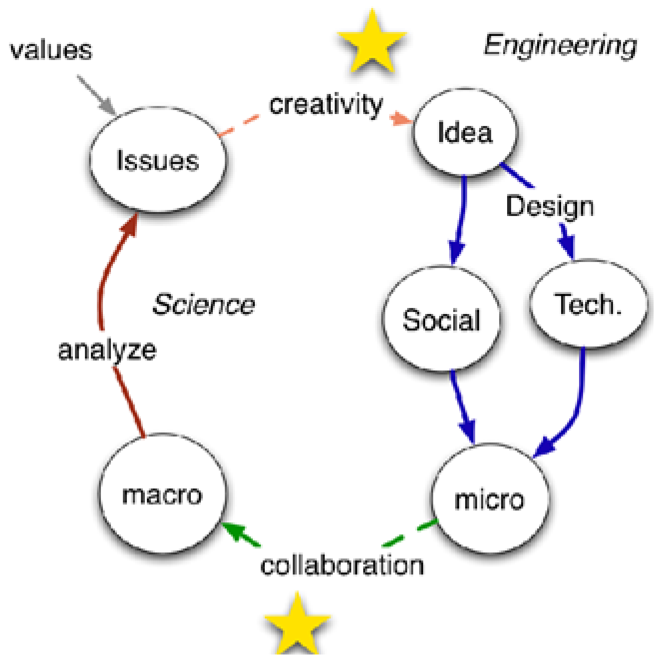


FIGURE 2.1: The 'Process of the Web'

and investigate the interactions between of human and technologies (Hanseth & Monteiro 1998), thus if Figure 2.1 was redrawn from this socio-technical perspective, the social and the technical would need to be considered as one rather than two separate components. In fact, as shown in Figure 2.2, the process of the Web would need an entire reconfiguration, as adopting a socio-technical perspective encompasses the entire process, not just the stages between the formation and development of an idea. Throughout the entire process of the Web, the concept of socio-technical applies, whether at the micro or the macro, human and technologies are present.

The adoption of the socio-technical concept within the computational sciences raises questions about its purpose and use, particularly towards understanding the underlying concepts and perspectives that are associated with what constitutes as a 'socio-technical' system. Used superficially, it does not provide any additional insights; it only becomes useful when supported by concepts and theory that can provide insight into the relationship and interactions between humans and technologies. The use of concepts such as 'socio-technical' opens up the possibility for interdisciplinary collaborative working environments and analytical findings that a single discipline may not be able to achieve. However, Web Science is still searching for approaches which allow for harmonious cross-pollination of theory and methods, and to establish common grounds for understanding the Web as a socio-technical phenomenon. If the concept of socio-technical is to be

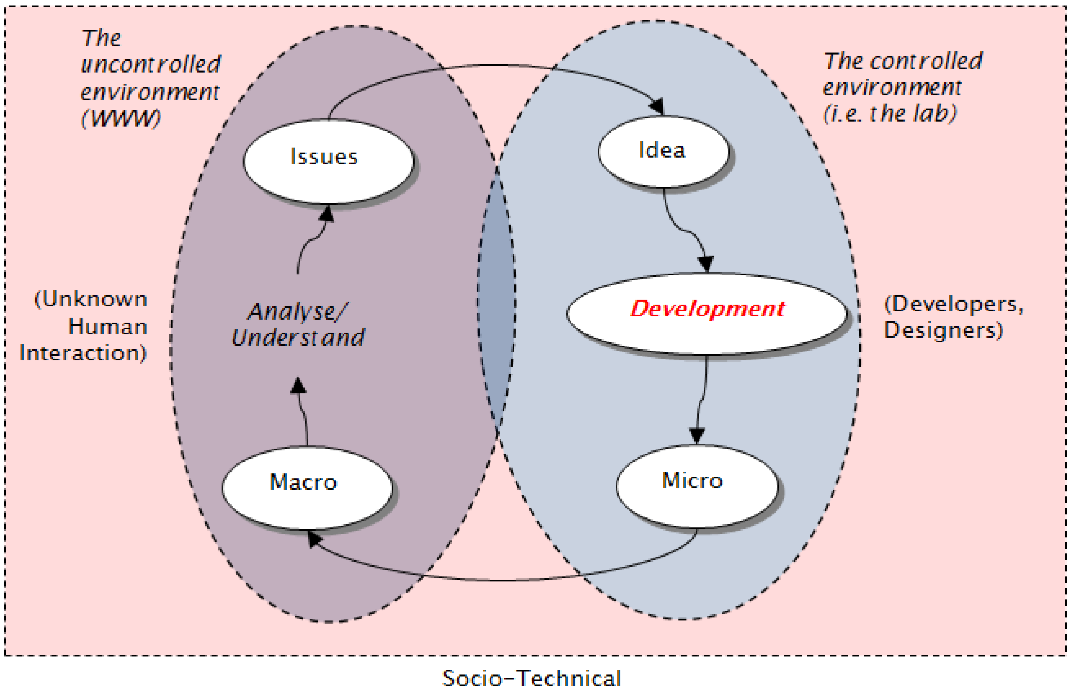


FIGURE 2.2: The 'The Socio-Technical 'Process of the Web'

adopted to understand the Web, then it is imperative to step beyond just computational techniques of analysis and perhaps more importantly, a realisation that studying human and technology interaction is not something that is limited to the micro level where development occurs, but rather as a continuous process which encompasses the controlled micro - and uncontrolled macro environment. Considering this, it is therefore necessary to develop an understanding of what a socio-technical perspective is and provides for understanding the Web.

2.4 Socio-Technical: Developing a Perspective

This section will examine the socio-technical perspective, drawing upon theory and concepts from disciplines not typically associated with Web or Computer Science discourse.

The roots of this concept can be first accredited to early research in the 1950's studying people and technology in workplaces (Trist & Bamforth 1951), and then later adopted and developed into well-established and empirically-grounded literature within sociological studies of the relationship between science, technology and society (Sismondo 2010*b*) Underpinning this approach is an epistemological perspective which is radically different to the positivistic approach associated within the computational sciences, (Orlikowski & Baroudi 1991). Through empirical-based qualitative studies, the concept of 'socio-technical' was developed to support the observations of the on-going relationship between the interactions of society, technology, and science, and more recently, the relationship between the human interactions and technological design.

Relevant to this is the sociological school of thought commonly known as Science and Technology Studies (STS). Research within STS has contributed to the development of methods and theoretical concepts which has helped advanced a sociological understanding of the relationship between science, technology and society (Law 2008) In order to understand the principles and concepts that underpin the current conceptualisation of the relationship between society and technology, it is essential to first understand how STS emerged.

2.4.1 Studying Science, Technology, and Society

The emergence of STS can be placed in a longer term history of the sociology of scientific knowledge, stemming back to the Vienna circle in the in 1930's (Popper 1963). Such research was deeply interested in challenging views of the relationship between science and technologies. Traditional beliefs that technology was simply the application of science were contested by the likes of Merton who questioned a technology's impact on science, drawing in aspects of the social and its involvement with the development of both technology and science (Merton 1973).

Before this, technology was seen as merely the product and outcomes of scientific endeavours; it was not considered that the development of technology influenced or impacted the path of scientific knowledge, or techniques. In hindsight, there are many examples where the development of a technology provided a gateway for new scientific discoveries to be possible; take for example the development of the Large Hadron Collider at CERN, as a result of its development, new particles such as the Higgs Boson have been discovered. Merton's work drew upon broader ideas about social function, and argued that science was a well-regulated social activity, and that existing and developing social structures had a direct influence on the outcomes of scientific studies.

Technology was also explored as a social phenomenon, with scholars such as Mumford (1967) and Heidegger (1977) investigating the influence of society on the development of a technology, and also considering the scientific limits to the advancements in technology. In turn, this sparked debate about determinism, does society drive the path of technological development, or does technology determine social structures? Asking these questions led to studying the development of science and technology, drawing upon scholars from multiple disciplines, whose theoretical positions were polarized by the line of determinism. By advancing the philosophical debate between science, technology and society, this eventually converged into the emergence of Science and Technology Studies (Sismondo 2010*b*).

Within STS there exist a range of theoretical positions regarding the relationship between science and technology and society (Sismondo 2010*b*). As more studies began to investigate the relationship between society and technology, technological determinism was further put under scrutiny, and by the mid-1980s, STS was beginning to offer a better understanding of the role of the social within technological development (MacKenzie & Wajcman 1985). This was also in response to the availability and growing integration of specific technologies in the workplace, public space, and home, STS studies started to focus on the role of technology; a period that became known as the technological turn (Bijker et al. 1989).

There was a shift in focus in STS from examining the relationship between science and society towards the relationship between technology and society. Sociologists such as Steve Woolgar began exploring this and describing a socially constructed relationship between technology and society (Woolgar 1991), which eventually led to the formation of a theory known as the Social Construction of Technology (SCOT). Underpinning this theory was the idea that humans shape the progress and direction of technological development (Bijker et al. 1989), and that technology is shaped by human actions. To be able to understand a technology, it must be understood in its social context. SCOT introduced and worked with concepts to explain how society was shaping the development of technology. At the core of these concepts, which also was a motivation for SCOT is the idea of the 'Strong Programme' (Bloor 1976) (Bloor 1981), which led to the development of Symmetry, which helps explain the origins, success and failures of science and technologies. Concepts such as interpretive flexibility and design flexibility, which offered a way to understand how different social groups viewed and used technology differently, and how these uses were inscribed into the technology's future development (MacKenzie & Wajcman 1985). SCOT also offered a way to discuss how a technology reaches closure, which represents the process of a technology losing its interpretive flexibility as a result of specific 'closure' mechanisms through the actions of the social groups that use a technology. A technology that reaches closure becomes blackboxed (Pinch & Bijker 1984), considered as a system that is functioning by its own success, and emphasis should be on its inputs and outputs, not its internal complexity (Latour 1999).

Whilst SCOT lies at the social deterministic end of the spectrum of the relationship between technology and society, it sparked important debates about technological development, and how human interaction (or society) is part of this. This was an important turning point for understanding the relationship between society and technology, and offers a suitable point to situate the current discussion of developing a socio-technical understanding of the Web.

Considering the Web as a socially constructed technology changes the understanding of its development and structure. Rather than seeing the Web as a technology that has developed through technological advancement, it can be understood as a socially driven system. A SCOT approach would argue that its creation, structure, features, and growth was a result of the needs and desires of the social groups that use it, which originally were the scientists at CERN. The inscription and eventual development of the Web was a product of the social structures that existed within the collaborative scientific projects, and their changing requirements.

How does this perspective address studies which examine the structures and characteristics of the Web graph discussed at the beginning of this chapter? The empirical studies introduced in earlier sections have led to mathematical explanations such as preferential attachment (Barabasi & Albert 1999); a process which leads to the condition where the distribution of links connecting documents together is not evenly shared. This can be represented by visualisations such as that shown in Figure 2.3 which help distinguish 'Web communities', and through the application of specific algorithms (Kleinberg 1999), can be used to identify 'Authorities' (Web pages that share a higher proportion of the links within the graph) (Dourisboure et al. 2009). However, whilst this provides a mapping of Web pages and the topological structure of the Web, under the lens of SCOT, this network and distribution of links observed, mean more than simply statistics and metrics. They represent the activities of humans and the acceptance and interpretive behavior of social groups using the Web (MacKenzie & Wajcman 1985). The questions and findings that Barabasi and others are engaging with are not the same as those that scholars such as Mackenzie, Pinch, or Wajcman are concerned with. An STS perspective alerts us to the need to understand the underlying social processes, and for this, appropriate theory and methods are required and used.

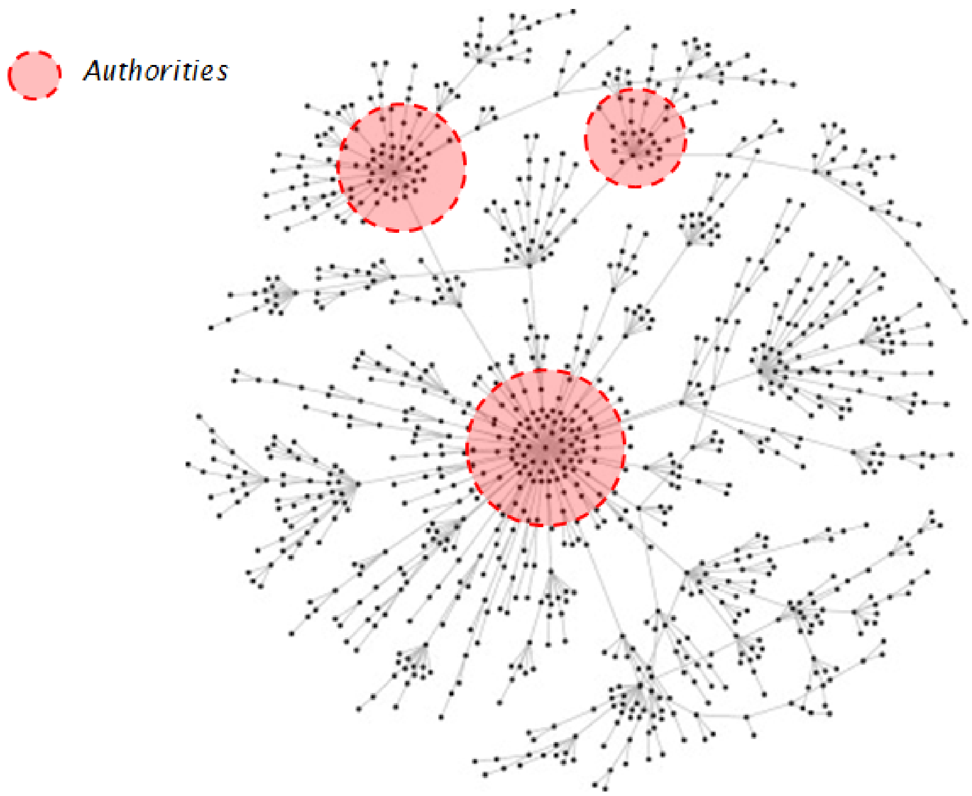


FIGURE 2.3: Example of Authorities within a representation of the Web Graph

2.4.2 Limitations of a Socially Constructed Perspective of the Web

SCOT offers a different perspective to understand the social processes that underpin the formation and structure of the Web. However, whilst it exposes the role of human activities, it fails to recognise how the development of a technology also has an effect on the social groups that are interacting with it. Whilst SCOT helps explain how a technology arises and is shaped by the acceptance of relevant social groups, it does not offer an understanding of the consequences after the technology has emerged, and it does not say anything about how it matters in the broader context (Winner 1993).

Unhappy with the way that socially deterministic theories such as SCOT underplay the role of technology within the relationship between society and technology, scholars thought harder about their co-evolution and interplay (Russell 1986) (Winner 1993). A new perspective has been proposed to capture the idea that society and technology are equally responsible for each other's development. Labelled as the 'co-construction' (Taylor 1995), this approach offered a way to examine both the social and the technical from an even-handed perspective. By adopting a co-constructional perspective, empirical studies of technologies innovations in science and technology argued that the emergence and use of a technology occurs though the interactions of humans and technologies (Callon 1986*b*) (Callon 1986*a*) (Latour 1987) (Law 1987). Co-construction offered a way to look at technologies beyond a deterministic position by providing a platform in which the social and the technical could be analysed together, rather than as processes that are independent of each other's actions.

2.4.3 A Co-Constructed Web

As a consequence of the growing attention towards co-construction, a new theoretical and methodological approach was developed which become known as Actor-Network Theory (ANT). The epistemological and ontological positions that this theory proposed were radically different from existing theories associated with STS or even within the broader context of sociology (Akrich & Latour 1992). ANT aimed to provide a new perspective on the fragile and dynamic relationship between the social and technical (Cordella 2003). ANT provided an alternative perspective for explaining not only the relationship between science, technology, and society, but perhaps more importantly, the relationship between humans (actors) and machines (actants) (Latour 2005). ANT insisted on a co-constructive approach which examined the symmetry between the development of society and technology, and also between actors (humans) and actants (machines). The emergence and acceptance of a technology is considered as a collection of interactions and associations between heterogeneous networks of actors and actants and only through the process of their interaction are outcomes produced. The most radical position that ANT makes is that any a priori knowledge cannot be taken into consideration when studying a phenomena, Thus constructs such as 'power' are viewed as the product of the interactions and emergent structures that the actors produce not as pre-existing or independent phenomena (Latour 2005). ANT's relational ontology decentres a technology and insists that it acquires its meaning and effects through relations with other entities.

ANT's move away from mainstream sociology and STS played an important role developing its perspective on agency and how it placed equal weight on the social and the technical, and on the actions of humans and non-humans. This was a radical step in forming ANT's distinct, and at times, contentious position. By readdressing the ontological boundaries of humans and non-humans, ANT placed radical symmetry between actors and their agency. An ANT perspective considers the development of the social and the technical are the direct result of the interactions between both actors and actants and on this basis, no single actor or actant should receive specific attention, they are all part of a network of associations to be studied (Latour 2005). This offers a different perspective and way to understand the Web from previous approaches discussed, which also requires appropriate methods to accommodate this ontological position. The traditional approaches that have been used to understand the Web as a graph of metrics and statistics or as a social construct are not adequate to understand the Web from an ANT perspective, methods are required which can understand the emergent phenomena as a product of both social and technical interaction.

2.5 The Emergence of a Socio-Technical Understanding

ANT made it possible to conceptualise and describe the development of a technology and change in society as a mutually constituted process, rather than either social or technical. Essentially, ANT offers a way to consider the social and the technical within one platform, not only at the epistemological layer, but also empirically. Effectively, this is a socio-technical perspective, and unlike Figure 2.1 , the social and technical components are not separate entities that can be understood in isolation of each other but are a mutual and co-constructed set of interactions between the humans and non-humans that form heterogeneous structures, which are complex in structure and constantly evolving and reshaping. From an ANT perspective, the Web can be understood as a complex and dynamic network of humans and technologies.

2.5.1 Interdisciplinarity: Re-stating the Aims

This chapter has shown that the Web can be defined and conceptualised in a variety of ways. Within the computer sciences; research tends to examine the Web in terms of its technical architecture, as is driven by quantitative methods and mathematical models. As a consequence of this, their definition of the Web is defined as “a complicated graph” (Wang & Reiter 2010), “a distrusted network” (Kang et al. 2005), or as “a huge network of Web pages and hyperlinks” (Murata 2007). The social sciences offer a different perspective, with far greater emphasis on observing the Web through a social lens, resulting definitions such as “a dead mechanical entity” (Fuchs 2003), “a dynamic medium of communication” (Pant & Srinivasan 2009), or even “a potential treasure trove of data” (Gong 2011).

Computer science and social science do not yet provide a perspective that appreciates the co-constructive and mutually shaping relationship between humans and technologies; instead, they work in isolation, separating the social from the technical. As this thesis aims to understand the Web and how it grows, it is clear that this is an appreciation of the social processes and human interactions and agency as well as a technical understanding. Understanding the design, development, growth and structure of the Web are both a social and technical concern, crucially it requires a way of understanding how the social and technical are co-constructed.

Actor-Network Theory appears a promising solution for understanding and conceptualise the Web and its growth from a socio-technical perspective, and therefore the following chapter will explore it in greater detail, examining its underlying theoretical position, concepts and capabilities, and its application for understanding the Web through the interaction of humans and technologies.

Chapter 3

The Web as a Socio-Technical Phenomenon

3.1 The Transition to a Socio-Technical Perspective of the Web

This chapter will explore the capabilities of the socio-technical perspective that Actor-Network Theory provides for understanding and conceptualising what the Web is, and how it grows.

The previous chapters have argued that in order to understand the Web, both technical architecture and social processes need to be considered. Current approaches tend to separate the social and the technical, and discuss them in isolation from each other, which has led to a deterministic perspective for describing and understanding the Web. Interdisciplinary thinking has been suggested as a solution to help tackle questions about the Web (Hendler et al. 2008) (Hall et al. 2012) and although the emerging field of Web Science has begun to cross the disciplinary borders between the social and technical perspectives and draw on concepts such as socio-technical to describe human and technological interaction, there has been little attempt to seriously engage with theories, methods and concepts associated outside computer science. However, in order to engage seriously with an interdisciplinary perspective at the theoretical and empirical level, adopting concepts such as socio-technical need to be considered in respect to what insights it provides for understanding the relationship and processes between human and technological interaction.

ANT has been shown to offer a social-technical perspective on human and non-human interaction. In order to understand how ANT can help conceptualise the Web, the following chapter will explore the theoretical perspective in more depth, examining the epistemology and ontology.

3.2 An Actor-Network Socio-Technical Perspective

The core principle of ANT is that the world is a network of associations between actors (Latour 2005). ANT aims to describe how associations enable networks to form and develop (Latour 2005). A network in ANT is heterogeneous, consisting of both humans (actors) and non-humans (actants), and the associations between them are not tangible links as such, but rather dynamic and changing ties that are established and lost through the interactions and actions of actors and actants.

To understand how the activities and associations between actors lead to the development of a network, ANT adopts a framework known as translation (Callon 1986*a*), which consists of four stages: Problematisation, Interessement, Enrolment, and Mobilisation, with the ultimate goal of producing a stabilised, operating network. Practically, translation offers a way to frame how the interests, actions and formation of associations between actors enable a socio-technical system to emerge and develop. At a theoretical level, translation provides a lens to expose the underlying processes and interactions between humans and technologies that produce network outcomes.

ANT emphasises the importance of considering translation as a fluid process (Law & Singleton 2005) which reflects the relationships and interactions between actors. A network which translates only holds its state (within or between stages of translation) as long as the actors involved continue to interact. Therefore, stabilisation of a network is only ever a temporary state, thus the network is able to change and reconfigure at any point.

The world through the lens of ANT is neutral and only through the interactions and configuration of actors that a network and phenomenon is produced. This perspective has allies with understanding the Web. In theory, the Web by design is a hierarchical-free structure; it provides no advantage or precedence to any particular actor, yet as Web studies discussed in the previous chapter has shown (Kleinberg 1999) the Web (graph) is unevenly distributed in respects to the links of Web documents and users. It is then fair to say that this uneven distribution is not due to technical design, but due to the activities of humans interacting with the technology. Indeed an ANT perspective may provide a richer understanding of the Web's structure by considering it as a evolutionary process that forms its structure through these interactions, such an understanding steps beyond the mathematical methods used to understand the Web. Therefore the 'process of translation', a concept concerned with the formation and growth of heterogeneous networks, may be able to help understand this evolutionary process.

3.2.1 Core Concepts

Underpinning the process of translation are two critical concepts: inscription and heterogeneous networks (Law 1992) (Latour 2005). Inscription provides a way to conceptualise how technical actors or artefacts that are part of a network take shape and embody patterns of use (Hanseth & Monteiro 1998). The patterns of use of a technical artefact emerge from the combination of the agency of an actor and the network, what the technology is for, and how it is used. As Akrich describes:

> *"Technical objects simultaneously embody and measure a set a set of relations between heterogeneous elements"* (pg.205) (Akrich & Latour 1992)

Where the capabilities and functionality of a technical object is related to, and can be observed by, the associations between actors.

3.2.1.1 Inscription

Inscription offers a way to describe how the development of patterns of technological usage and functionality is partly the result of the desires, capabilities, expectations and limitations of the network of actors. Human actors may define technological artefacts with specific competencies, features, motives and limitations but ANT also considers how other factors such as the use of the technology, the development of science, and its place in the network will evolve the technology in ways that are not solely prescribed by human actors (Akrich & Latour 1992). ANT disposes of ideas of social order and argues that stability relies on the network's ability to adapt over time, which is achieved by the continuous interactions between actors. Translation presupposes inscription, in that translations are "embodied in texts, machines, [and] bodily skills" (Callon 1992).

As a network translates, scenarios of how the technological artefact will be used are inscribed by the human actors. Within an inscription, users are given 'programs of actions', which define the expected role and use of the technology by the user. Through this process, implicit and explicit assumptions are made with regards to the users of the system, delegating roles and competencies to the actors of the socio-technical network; by inscribing programs of actions, the technological artefacts become an actor and impose its capabilities and limitations on other actors (Latour 1991). However, the inscribed patterns may not succeed, the interpretative flexibility and agency of actors may mean that an artefact deviates from its intended use. In principle, inscriptions are flexible, and can offer either strong or weak programs of action. This is an important dimension to how inscription offers an understanding of the development of a network; different strengths and flexibility of inscription is dependent on the reversibility of the network, and depending on this, the network's translation may last, become stronger or fade away.

Inscription provides a concept to understand the relationship between technical design and human interaction in socio-technical systems, such as the Web. It focuses attention on how the interests and decisions of specific human actors influence and determine the development of technological artefacts, and provides a way to discuss this in terms of the strength of their inscriptions, and how these evolve throughout the network's growth.

3.2.1.2 Heterogeneous Networks

Heterogeneous networks provide a theoretical basis to represent different actants and artefacts within a network. These networks are not static (or in fact, tangible), but rather a product of the interactions and transient relationships that form between a set of actors and actants (Latour 2005). Through these interactions, artefacts of the network emerge and social outcomes are assembled. Rather than working with an *a priori* assumption of social structures, ANT takes the position that the interactions, negotiations, and strengthening of ties between actors within the network enable social structures to form and a network to emerge. As the relations between actors are transient, social structures, thus the roles of actors, are only temporarily held stable.

Heterogeneous networks operate at various levels of scale and granularity, providing a way to observe interaction between actors depending on the intended observational focus. For instance, a television in a room with several people could represent a coherent, single actor within a larger network of actors, however depending on its context and its mode of operation, it could just as easily represent a complex actor-network of electronic components and human interventions; as Law (1992) contends "an actor is also, always, a network". Exposing or collapsing an actor is achieved via the process of punctualisation, rendering the actor-network as either a black-box or as a network of actor-networks (Callon 1986*b*). The use of this concept provides different levels of granularity for observation, offering an approach capable of studying phenomena as a whole, which is also the same as studying it's parts (Latour et al. 2011).

Unlike social theories such as SCOT, ANT places equal agency to both technologies and humans (Sismondo 2010*a*). Whilst ANT does not intend to downplay the self-agency of a human, it argues is that all actors are 'social agents', and because of this, agency is "never located in bodies and bodies alone"; only through the interactions of actors is agency really understood (Law 1992). ANT defends it's position by suggesting that to truly understand a socio-technical phenomena, an analytical perspective is required which provides a platform for both human and non-human actors to be explored together (Latour 2005).

ANT's stance on agency has received criticism for limiting its capabilities for observing alternative and unobserved social processes that may have contributed to the overall outcomes (Whittle & Spicer 2008). It is suggested that by giving an equal share of agency to all actants ANT does not take into account the mind or self-driven decisions of human beings (Lee & Brown 1994). However, ANT's approach to agency is analytical, rather than epistemological; it is not aimed at reducing humans to objects or machines. Instead, it attempts to emphasise that action (of the network) is not only through humans, but is an effect caused by the network (Law 1992). In building (or constructing) the world (and indeed, the Web), human actors are not the only actors involved in its construction and development, and perhaps more critically, the emergence and growth of a network is the result of the sum of its parts. Even if actors strongly inscribe technical artefacts, these artefacts still have agency which can affect the actions of human actors, thus despite being manipulated and constructed by humans, they are still able to resist (i.e. a technology that fails) (Latour 1990).

Some studies that have engaged with ANT have downplayed non-human agency (Miettinen 1998), reducing the capabilities of ANT's analytical perspective. Instead, by embracing the balance of agency between humans and technologies, it is possible to uncover the social and technical processes in which humans and technologies interact with each other, therefore by denying or ignoring technological agency reduces the explanation to a socially constructed perspective. This provides no benefit for this thesis as it has already been argued that the Web comprises of both human and technological interactions. Whilst humans shape technologies, technologies also affect the way humans interact and operate within a network. Paying attention to all actors helps uncover and establish the socio-technical processes of the Web.

3.2.1.3 Translation

Whilst heterogeneous networks offers a way to conceptualise the interactions and relationships between of human and technologies within a single network; the process of translation helps to understand the underlying processes that enable these networks to emerge, function and become stabilised. Changes in the network are the process of translation, defined by ‘‘a relation that does not transport causality but induces two mediators into coexisting” (Latour 2005). Therefore, translation is the process whereby actors align to form a heterogeneous network.

Translation comprises of four stages:

1. Problematisation involves the realisation and definition of a problem by an actor (not necessarily a single actor) and the relevant actors who are required to overcome it. As part of this process the actor (labelled as the focal actor) identifies the interest of the other actors which will need to be captured in order to obtain their help. The activities of the focal actor lead to obligatory passage points (OPP) being set, which represent the tasks that other actors need to pass through in order for the network to translate and produce outcomes. By setting these OPPs,

the focal actor becomes indispensable within the network and is required for the network to successfully translate (Callon 1986*a*).

2. Interessement is the process of obtaining the interests of other actors and negotiating their involvement. Network objects such as incentives or negotiations are essential to gain the interest of actors; these objects are labelled 'interessement devices' and are used to help develop and strengthen ties between actors, and disrupting potential competing actor-networks which may inhibit the success of the network. Effectively, devices of interessment help form ties with actors that are needed for a network to translate, and help "construct a system of strong alliances between actors" (Callon 1986*a*).

3. For the network to be successful, the interested actors must align together, forming alliances, and accept to take on their roles. This leads to the third stage of translation, enrolment. Given the successful use of incentives and other types of interessement devices, alliances between actors form, which can be described as a "group of multilateral negotiations" (Callon 1986*a*). As part of the process towards the enrolment of actors, it is essential to translate the questions identified during the interessement of actors into descriptive statements; these form the goals that need to be achieved for network success.

4. The fourth stage of translation is reached when the alignment of the enrolled actors starts to produce network outcomes, and the network goals set during the previous stages start to be achieved (Callon 1986*a*). At this point, the network operates in a 'mobilised' state of translation, and the primary actors assume a the role of a spokesperson for the passive actors (and actants) within the network (Callon 1986*a*). By taking this role, the primary actors try to mobilise the passive actors into action and retain and support actors that are producing network outcomes. Based upon the focal actor's problematisation during the initial stages, and their ability to attract and interest actors, a heterogeneous network emerges from the interessement, alignment and mobilisation of actors. The mobilised actors attempt to achieve the obligatory passage points set, and to produce network artefacts and outcomes.

The process of translation involves various negotiations between actors and actants, and successful negotiations represent an actor-network that has committed actors and aligned interests (Latour 2005). The process of translation describes how an initial set of ideas and problems becomes a 'functioning' network, yet this only remains stable and producing outcomes as long as the commitment of the actors are not threatened or interested by another networks' goals or interests.

Translation provides a framework to describe how socio-technical outcomes emerge, yet criticisms have been raised concerning the rigidity of the translation process, and how the prescribed stages may miss hidden layers of action (Sismondo 2010*a*). These criticisms position themselves to suggest that translation is a fixed process, however, even though translation can be described as a set of rules, they are not considered as strictly defined unit of process. Contrary to this, the stages of translation, and to this point, the framing of translation are due to the activities and actions of the actors. Translation does not define the actor-network; it is the result of the interactions and associations of the actors.

Objects and artefacts within the network are considered a product of the interaction and alignment of actors. Although ANT's position is that entities only exist due to their configuration and relations with other entities, it appreciates that some objects remain the same, independent of the network they are in. ANT considers these as *immutable mobiles*, objects that may remain stable, across more than one network (Sismondo 2010*a*). Immutable mobiles are produced like other network artefacts and outcomes, they emerge through the associations and interactions of actors, however, they have resisted change to emerge as order (Law 1992). Whilst immutable mobiles are objects that do not change within the network, they depend on their usage and configuration within the network; they are not inevitable or (necessarily) permanent. Immutable mobiles and are subject to "tinkering" and the judgement and acceptance of actors (Latour 2005). Therefore objects, like translation, are not prescribed in ANT, they are fluid concepts and entities, produced by the associations between actors.

Enticed by the capabilities of a theory that can help describe the interactions between humans and technologies, Information Systems research (ISR) has employed ANT to understand technology innovation, use, failure (Tatnall 1999). The following section will examine to what extent ANT has been used in Information Systems Research to consider it's utility for examining the Web.

3.2.2 ANT in Information Systems Research

One of the driving factors for the use of ANT within Information Systems Research (ISR) is the growing interest in understanding the interplay between technologies and society (Cordella 2003). The study of understanding technology and society within information systems has previously borrowed theories from other sociological work, with previous studies working with theoretical models from structuration theory (Orlikowski 2000) and hermeneutics (Lee & Brown 1994). However, as the direction of Information Systems Research has become more interested in how the design of technologies are inscribed by actors, ANT has been poised as a theory which could provide both methodological and theoretical insights to understand the relationship and interaction of huamns and technologies within information systems (Walsham 1995). The interest towards ANT as a solution has also been driven by the ontological position that ANT takes in regards to a socio-constructivistic explanation of technology design (Cordella 2003), that being, (paraphrasing Akrich (1992)) how the activities of the designers have specific interests, desires and agendas, and that these are inscribed within the shaping of technology. ANT has been used to examine questions of technical design, technical flexibility and the shaping of technologies (Monteiro & Hanseth 1995)(Hanseth & Monteiro 1997). At the same time, ANT has been used as a framework to understand the innovation, adoption, and implementation of information systems. Concepts such as translation have been used to help tease out how 'users' accept or resist technological change (Tatnall 1999) (Hannemyr 2003).

ANT has been applied across many different areas of interest in Information Systems Research, from examining the implementation and use of technical systems within organisations and emergency centres (Mähring et al. 2004) (Brooks et al. 2008), to understanding the reason for global adoption of technologies such as email (Lee & Brown 1994), the internet (Hannemyr 2003) or the Web browser (Faraj et al. 2004). Within these studies, ANT has been used to explore the interplay of humans and technologies that led to either the success of failure of the implementation or use of a technology, specifically focusing on how conflicts between actors led to outcomes which were neither foreseen nor envisioned as part of the initial design process.

Using this approach has also helped understand the dynamics and adoption, and standardisation of technology within and across organisations (Baygeldi & Smithson 2004)

(Cordella 2011), and the events and actions of the actors that led to the failure of technologies within large organisations (Sarker et al. 2006). By using ANT as a theoretical framework, these studies have shown how the design and use of technology is not separate from business-level decisions, and that the design and adoption of a technology is a dynamic process; failure of technologies are a consequence of rigid, non-flexible organisations and processes. These studies share similarities with ANT's use in studying the adoption of E-commerce systems (Tatnall & Burgess 2004) (Lee 2005) and studies examining the implementation and failure of digital government systems (Ayyad 2009) (Gunawong 2010). These studies have described how the acceptance and growth of a technology cannot be attributed to solely its capabilities and functionality or the development and deployment of social processes and/or policies. Successful adoption is revealed as a complex process of negotiations, driven by alliances between humans and technologies. These studies have demonstrated that ANT has application within Information Systems Research. More recent research has used ANT as a lens to study the rise and fall of Web technologies such as MySpace (Rossi 2008) the characteristics of Web 2.0 (Depauw 2008), or the wider cultural and global impacts of adopting western naming conventions on the Web (Zimmermann 2008). These empirical studies have helped strengthen ANT's position as a theory that can shed new light on the interplay between technologies and humans, and how these shape the adoption and failure of socio-technical systems. However, whilst these studies provide a detailed account of the success and failure of a technology, an organisation, or a social process, they do not attempt to ask questions beyond the descriptive layer. Whilst description is necessary, this thesis strives to provide an understanding of the Web beyond just a descriptive dialog.

These studies have used traditional data sources such as documents and ethnographic data but advances in computation techniques and the increasing amount of digital data is providing new ways for ANT to observe the interactions and associations between humans and technologies. This thesis will build upon the success of previous Information Systems Research studies' that have harnessed ANT, but seeks to go beyond these by drawing on new sources of data and analytical techniques. In order to describe this further, the next section examines how these new sources of data can help provide an additional perspective for observing how associations between actors produce networks and artefacts, and how networks are configured.

3.2.3 Harnessing Digital Traces through ANT

Traditionally, ANT has been aligned with the use of qualitative sources of data; observations and interviews were essential tools for providing insight into the interactions between human and technological actants. Prior to the availability of digital data recording human interaction, the collection of qualitative data was one of the only methods to gather evidence on how actors were interacting. However, since the development of global communication technologies and advances in computing techniques, quantitative data can provide an additional way to describe the "intangible social phenomenon" produced by human and technological interaction (Latour 2011).

There is growing acknowledgement of the analytical capabilities of 'big data' (Boyd & Crawford 2011), a source of digital data, albeit with specific characteristics: volume, velocity, and variety (Beyer & Laney 2012). Big data is not only large, but it can represent a real-time view of activities and human interaction; and perhaps the most important aspect, it varies in content and is linkable between data sources. Big data offers the potential to uncover interactions of humans and technologies, the socio-technical processes involved, and the network outcomes produce. Arguably:

> *"it would be a pity to miss this occasion to explore such a powerful alternative that may provide another way to render the social sciences empirical and quantitative"* (Latour 2011)

This digital data provides a rich resource to observe the interactions between actors, and to uncover the processes that help form heterogeneous networks. The availability, richness and variety of digital data resources have not gone unnoticed by ANT scholars (Latour 2011). It is now possible to collect vast amounts of 'social data', which promises useful insights into understanding the interactions between humans and technologies, at scale (Manovich 2011).

> *"As a sociologist of sorts, I have been especially interested in what this revolution does to social theory. And what it does is truly amazing: It dissolves entirely the individual versus society conundrum that has kept social theorists and political scientists busy for the last two hundred years. To sum up a long argument: We have the social theory of our datascape. If*

you change this datascape, you have to change the social theory." (pg. 7) (Latour 2011)

By harnessing both quantitative and qualitative sources, ANT is offered a richer set of resources and tools to observe, describe, and analyse the interactions between human and non-human actors which form heterogeneous networks.

ANT presents itself as a theory that has concepts to understand socio-technical phenomenon from a co-constructed perspective. Its epistemological and ontological position takes a balanced view of humans and technological agency, and consequently ANT tears down the deterministic boundaries once faced in STS. This is not to say that ANT is STS, but it is a radical move away from the traditional positions and theories has dominated STS discussions.

3.3 A Socio-Technical Web

The thesis is now in a position to consider how the Web can be understood from a socio-technical perspective. Argued throughout these chapters, the Web has been described as more than just a technological development, the Web is a "social machine" (Berners-Lee 1999) which develops via the interactions and negotiations between humans and technologies. ANT offers an analytical lens that appreciates the complexity and interrelatedness of social and technical processes that lead to the development of the Web. The ANT concept of translation can provide a framework to analyse how the Web grows, and how heterogeneous networks form. Whilst previous computer science literature has described the Web by its technical architecture, the structures between Web documents, and developed quantitative approaches to detect, observe and model communities via hyperlinks or user connectivity, this thesis offers a socio-technical perspective on the Web, led by a theory that is grounded in exposing the interactions between humans and technologies which produce socio-technical outcomes.

Thinking of the Web as 'socio-technical' is complex. It is derived from disciplines outside the hard sciences, and as the previous chapter has shown, the perspective that it offers challenges the positivistic position and findings that is typically found within computer science Web research. ANT offers a way to capture the essence of what being socio-technical means, and by extension to this, what is a socio-technical Web. The argument in this thesis has moved beyond the question of what is a socio-technical Web; instead, this has been pushed up another epistemological, ontological, and theoretical rung on the ladder. This thesis now takes the position that the Web is a socio-technical phenomenon; it has become clear that the Web has developed as a consequence of human and technological interactions, which has been part of a co-constructed process. However, there is very little, if any, understanding, of what this co-constructive process actually is, and how it happens. Literature presents concepts such as socio-technical to label this process, yet do not provide an in-depth understanding, explanation or analysis of what it actually means or how it happens. Therefore, this thesis asks the questions: what is a co-constructed Web and how does it grow?

3.3.1 The Web: More than an 'Information Space'

By asking what is the Web and how does it grow, this thesis will need to question how the past methods used to understand the Web can contribute towards a socio-technical understanding. A socio-technical lens reveals that technically-driven approaches provide only one perspective of a highly complex network (of networks) of heterogeneous actors. From this perspective, the Web is a large scale network which can be reduced to various smaller networks, identifiable by geographic location, topics, or common interests. These networks are structurally different, yet architecturally, they are the same. However, from a socio-technical perspective, the different networks emerge as a result of the interactions of actors and actants, and the Web's variation in structure is produced by the different configurations of humans and technologies.

The Web is a heterogeneous network: it is the mix of technologies, documents, information, human activities, governing standards, and cultural practices, that act to reify the abstract 'information space' that the W3C formally identify as the Web (World Wide Web Consortium (W3C) 2009). As a result of this, capturing, or even conceptualising how the Web grows is a challenging task. Not only are the interactions between humans and technologies occurring within the online environment, but also across the blurred boundaries of the online and offline.

The Web is a system that spans between being online and offline, it is simultaneously a collection of interconnected documents within an 'information space', whilst also a heterogeneous network of humans and technologies. There is a co-constructive relationship between online and offline activities and understanding the Web must take into account the activities that occur both online and offline. Being aware of the Web as a socio-technical phenomenon is important for both analysis and design, and understanding the Web at the micro is just as important as studying the macro. Both are essential to understand the Web's growth and structure.

New technologies provide ways for large-scale communications and broadcasting of information; enabling the diffusion of information to be captured, tracing the interactions of and between individuals, and by analysis, understand how technology is used, and how individuals and groups communicate and overcome social and technical barriers, such as those during times of political uprising (Rahimi 2011) (Velasquez 2012) (Zhou et al. 2010), health pandemics (Ginsberg et al. 2009) (Ortiz et al. 2011) and the spread of disease (Paul & Dredze 2011) (Sadilek & Kautz 2012).

In light of this, in order to answer the questions about the Web's growth that this thesis proposes - which are closely aligned with the fundamental questions that has led to the emergence of Web Science (O'Hara & Hall 2012) - interdisciplinary research has shown to be necessary. This requires serious consideration and engagement from both the computer and social sciences (Savage & Burrows 2007) (Savage & Burrows 2009), stepping beyond single disciplinary perspectives and methods. This thesis builds upon this, and argues that in order to understand how the Web grows, research must look beyond just statistical measures and metrics of network structures and take into consideration the dynamic interactions of humans and machines embedded within the data (Scott 2008) (Barabási 2009). The use of these methods will help conceptualise the Web at the micro and macro level in terms of its socio-technical structure, and unpack the underlying interactions between humans and technologies of Web growth.

3.4 Research Questions: Conceptualising a Socio-Technical Web

This chapter has argued that the Web needs to be understood in socio-technical terms. By examining existing studies of the Web, this thesis has demonstrated that understanding the Web from either a technical or social perspective does not provide an adequate approach to understanding to what is the Web, or how it grows.

The Web is a heterogeneous network of human and non-human actors and as such is co-constituted. This description had moved beyond the original concepts of the Web as a "Social Technology" (Berners-Lee et al. 2006) and pushed it further, arguing that understanding what it means to describe the growth of the Web as a socio-technical phenomenon needs further thinking. The Web is neither social nor technical; one does not exist without the other. Yet current perspectives are polarized and deterministic; even within the social science, approaches assume that there exists a Web 'out there', and that human interaction makes it function:

> *Without human beings, the Web is a dead mechanical entity that is not self-organising* (Fuchs 2009)

This thesis argues that a technological deterministic view provides no benefit to understanding the Web. Technology does not drive social change, it is part of a co-constructed process. To that end, the Web is not something that 'just exists', the 'mechanical entity' in fact is a result of the continuous engagement of human interactions.

A socio-technical perspective adds valuable knowledge to understanding what the Web is: by illustrating that the development of the Web is as much a social process as technological evolution, it takes the position that humans and technologies rely on each other and do not exist or develop in isolation. This thesis therefore asks: how can this perspective be used to understand what the Web is and how it grows. Indeed it re-asserts that these are questions which form the fundamental drivers for the emerging discipline of Web Science (O'Hara & Hall 2012).

3.4.1 Characteristics of the Web

Conceptualising what the Web is and how it grows is a broad challenge. The previous chapters have been used to develop the argument that the Web can be conceptualised as a system of human and technological interaction. Based upon this, the current understanding of what the Web is can be reconceptualised and interpreted through a socio-technical lens:

1. The structural observations of the Web (graph) and the detection of communities suggest that it has emerged through the interaction of different communities which are associated with different social activities and interests.
2. The identification of different communities within the Web using the global and local algorithmic techniques developed in the computer sciences indicate that the Web is developing as a result of the interactions between humans and technologies at the micro and macro level.
3. The observation of the degree disruption of hyperlinks between Web pages in the Web (graph) suggest that Web documents which obtain more popularity may reflect different forms of social interaction and could be associated with activities that have different popularity (e.g. Online shopping v.s. E-Government websites).
4. The varying size of the Web communities identified in the Web (graph) indicates that the growth and structure of these communities are different and develop over different periods of time, which may be a consequence of the different underlying social and technical processes of a community's growth.

These characteristics of the Web help demonstrate that the Web is a social and technical phenomenon and that the adoption of a socio-technical lens reveals processes of human and technological interaction. Thus, based on these characteristics, this thesis is challenged by the question of how can the growth of the Web be understood?

3.4.2 Research Question

Based on the discussion presented in this and previous chapters, this thesis has identified that there is currently no way to understand the Web from a socio-technical perspective. Consequently, this thesis asks the following question:

How to investigate, understand, and analyse the growth of the Web?

Building upon the already established thematic line of socio-technical enquiry, this thesis will aim to explore how the Web is growing via socio-technical processes, and the characteristics that have been identified above. In order to achieve this, what will follow is the development and application of a methodology and empirical-led analysis grounded with a theoretical position and set of concepts offered by Actor-Network Theory that will enable the socio-technical growth of the Web to be explored.

Therefore, in order to answer the research question, the following challenges are faced:

A The development of a methodology to describe and analyse the growth of the Web, which harnesses the epistemological and ontological position of a socio-technical perspective.

B The development of a suitable research design and case study to apply the methodology in order to demonstrate how the Web grows and structures emerge.

C The development of a theoretical model to conceptualise the growth and structure of the Web at varying levels of granularity, applicable at the micro and macro level.

In order to answer the research questions of this thesis, the next chapter will introduce the research design for describing and analysing the growth of the Web.

Chapter 4

Choosing a Case Study Design and Mixed Methods

4.1 Introduction

This chapter provides the bridge between the theoretical discussions and empirical analyses of a co-constructed socio-technical Web by outlining the study design and methodology required to answer the research questions proposed in this thesis. The previous chapters have developed a theoretical position to help examine the co-constitutive growth of the Web. It was argued that the questions being asked in this thesis lie at an interdisciplinary crossroads between computational and sociological approaches so the research design is necessarily interdisciplinary and embraces methods from both disciplines.

This chapter will begin by introducing the overarching methodology and study design employed within this thesis, and how this helps answer the research questions proposed in the previous chapter.

4.2 Case Study Design

The following section provides an account and justification of the single case study design used within this thesis. Different study designs were considered and are briefly discussed as are the challenges and implications of conducting case study and mixed methods research when studying the Web. This section also explores how the study design is informed by Actor-Network Theory and how the theory and methods inform each other.

This section will introduce and examine the case study research design used within this thesis. A research design provides "a logical plan from getting from the initial set of questions to a set of conclusions" (Yin 2003). The choice of the design should relate to the research questions and the availability of resources. Decisions made about the design will influence the direction of the research and ultimately, the findings. As an approach to help decide the most appropriate research design, Yin (2003) asks three questions, what type of research question is it? Does it require the behaviour of events to be controlled? Finally, what type of events (historic or contemporary) will be focused on? These questions offer a loose framework to help choose a suitable research design from the different research strategies available.

In addition to the questions Yin proposed (Yin 2003), it is worth considering the type of events and phenomena being studied when choosing a research strategy. Three broad research strategies are delineated by Yin: namely histories, experiments, and case studies (Yin 2003). The first is used to retrospectively examine events which typically are not under the control of the researcher, effectively providing an account of the documented activities which have previously occurred. Experiments are concerned with contemporary events and analysing the impact of interventions and so require a high level of control over the setting and opportunities to change what happens within a defined group or setting. Case studies provide a descriptive or explanatory approach to uncover and understand the underlying principles of the observed phenomenon and can use retrospective or prospective data. Case studies can be longitudinal and they often combine different sources of evidence to study people, events, systems, and/or policies (Thomas 2011).

Based upon the research questions asked in this thesis, a case study design appears to offer a suitable approach to explore the characteristics, development and processes responsible for the growth of the Web. An historical study was ruled out as retrospectively analysing Web activity might not adequately capture its dynamic features, and so might not help answer the question of how the Web grows. The availability dynamic Web data which provide (near) real-time understanding of the phenomena unfolding seemed to require methods which were not aligned with historical studies. The use of experiments as a research strategy for this study was excluded as it was not possible (nor within the aim of the thesis) to intervene or control variables within the large scale, dynamic, global network of the Web. Whilst experiments within a self-contained environment (micro level) which simulated the Web are possible, this approach was felt to offer less insight into the growth of the Web than a case study which incorporated naturally occurring (non-experimental) data.

Case studies are a common research strategy within numerous disciplines, partly because they are helpful for understanding complex social phenomena. This design also neatly straddles the gap between technical and social disciplines as it is commonly used in both. That said, whilst case studies offer a way to explore specific phenomena, they have been criticised for limited generalizability of findings and outcomes (Tellis 1997), as, unlike experimental studies, case studies are by their nature often smaller scale may not enable inference or prediction. However case studies make up for this by offering a high level of detail and because of this case study analyses may develop transferable explanations and analytical concepts which can be applied in other similar settings and cases (Yin 2003).

The next section will describe key components of a case study and its capabilities and limitations and demonstrate how this approach was used to frame the thesis research.

Yin outlines five components of a suitable case as a way to begin designing a case study. These are:

1. the research questions being asked,
2. its propositions (if any),
3. its unit(s) of analysis,
4. the logic linking the data to the propositions,
5. and the criteria for interpreting the findings (Yin 2003).

A case study design should reflect the research questions being asked which typically fall into five different but not exclusive types of questions: who, what, where, why, and how (Hedrick et al. 1993).

This thesis is driven by the question of what is the Web and how it grows. This means empirically analysing the Web, or more practically, some part of the Web to unpack and understand the co-constructive relationship between humans and technologies. The intention is to provide a novel conceptualisation of understanding the growth of the Web, and as a consequence of this, demonstrate a methodology grounded in interdisciplinary theories and approaches for the emerging Web Science discipline. This therefore requires choosing a case, and specifically units of analysis that reflect key features of the Web, but are manageable, and practically analysable within the timeframe of a doctoral study.

The choice of unit(s) of analysis is informed by practical considerations (i.e. what is feasible and manageable in a doctoral research thesis) but also relate directly to the research questions and propositions. In order to investigate the Web at the micro and macro level, it was important to be able to explore the individual social and technical processes and interactions at a granular level, for example looking at individual actors in a network whilst at the same time being able to take a wider more abstract view of a whole network and set of interactions. In addition to these considerations a choice between using a single or multiple cases was necessary. Single case studies are often used in research that seeks to critically evaluate an existing theory and for longitudinal studies where lots of data can be amassed over time on a single case. They are also suitable in situations where there are no comparators (Tellis 1997). Alternatively, multiple case studies allow comparison across cases to prove or disprove theory or hypotheses (Yin 2003). Both types of case study design offer their own benefits and limitations, not only in terms of their theoretical and analytical capabilities, but also with regards to the resources required, including time.

A multiple case study design was considered for the thesis research. However this immediately raised questions as to how many cases should be studied, and what the scale and scope of each case should be. In addition to introducing this kind of complexity it was felt that using multiple case studies for understanding the Web would not offer advantages over a single detailed case study approach. In the next chapter, the chosen case is described in more detail, but before this, the remaining sections in this chapter look at how a mixed method can be used in a case study.

4.3 A Mixed Methods Approach

The mixed methods approach used for this thesis combines qualitative and quantitative data and their accompanying analytical methods (Johnson & Onwuegbuzie 2004). Greene et al. (1989) suggest that a mixed method approach provides a way to take advantage of quantitative and qualitative approaches without being tied to a specific research paradigm. Indeed mixed methods approaches are sometimes framed as a solution to ontological differences between the philosophical viewpoints regarding the nature of reality (ontology) , and the appropriate methods for generating knowledge (epistemology) Guba & Lincoln (1994). Mixed methods studies typically combine different methods and analytical techniques, often drawing on qualitative and quantitative methods to give a more comprehensive understanding of a phenomena or case.

Methods can be combined in various ways (Rossman & Wilson 1994) (Tashakkori & Teddlie 1998). Quantitative and qualitative data can be analysed either concurrently, or sequentially (Creswell 2003) (Greene et al. 1989). A concurrent mixed methods approach requires the researcher to collect and analyse the quantitative and qualitative data simultaneously. In a sequential mixed methods approach the researcher may start with either the quantitative or qualitative approach and use this to inform subsequent data collection. Unlike the concurrent approach where the boundaries of the data paradigms are blurred, a sequential approach tends to separate the qualitative and quantitative sources and may help preserve the paradigmatic underpinnings and associated ontological assumptions.

There are additional arguments for combing a concurrent or sequential approach, consideration needs to be given to how and at what stage of the study are the data sources are linked together (O'Cathain et al. 2007). This may be achieved during a distinct analytical phase or during the data collection. A key concern is often the need to balance between the quantitative and qualitative components in the sequence of collecting and analysing the data so that one paradigm is not dominant.

In order to examine this in more detail, the following section will discuss the quantitative and qualitative paradigms and the implications for a mixed methods approach.

4.3.1 Quantitative and Qualitative Paradigms

Quantitative research is often underpinned by positivism (Orlikowski & Baroudi 1991). In this paradigm phenomena can be reduced to empirical indicators. The ontological position taken is that there is a single truth, and that reality exists independent of human observation. The goal of positivist research is therefore to measure and analyse causal relationships between variables within a value-free framework (Denzin & Lincoln 1994). Quantitative studies tend to use larger sample sizes than the qualitative research enabling findings to be analysed using statistical methods (Carey 1993). If the aim of a study is to predict and/or control natural phenomena (Guba & Lincoln 1994) this is useful. Quantitative research in the positivist tradition involves the verification or falsification of an idea or hypothesis, based on the assumption that it is possible to independently test theory. A core test of a positivist study is that its results are verifiable and repeatable.

By contrast, qualitative research is associated with an interpretivist paradigm (Altheide & Johnson 1994) (Kuzel et al. 1991) (Secker et al. 1995) and often with social constructionist perspectives (Guba & Lincoln 1994). Ontologically these approaches argue that there is no single truth but instead there are multiple realities which derive from the researcher's construction of reality (Berger & Luckmann 1967). Epistemologically these approaches do not see reality as independent of the researcher's perception and there is no external referent to compare truths (Smith 1983). Thus the definition of reality is constantly changing and socially constructed. The emphasis in qualitative research is on process and meaning rather than prediction or modelling. This requires the use of various data sources which utilize both direct contact with the subjects being studied via interviews, observations, participatory involvement, and less obtrusive methods such as reviewing texts and documentation they produce. Reflecting this different nature and purpose of qualitative studies are often small-scale, micro-level studies that focus on detail, and do not address macro-level structural concerns (Reid 1996).

The underlying ontological assumptions of the two paradigms result in differences which extend beyond methodological and philosophical debate. However these differences are rarely discussed in some disciplines due to the positivist paradigm being a dominant frame of reference, especially for the physical sciences. Reflecting on these differences between quantitative and qualitative approaches, it is nonetheless helpful to combine them when exploring micro-level and macro-level phenomena. It was also felt that combining methods would help to harness the benefits of both technical and social approaches to understanding the Web (Guba & Lincoln 1994).

Aware of the differences between the two broad categories of data, the challenge lies in appreciating the difference between paradigms. Whilst the quantitative approach may use scientific processes in order ensure findings are reliable and valuable (Smith & Heshusius 1986), qualitative approaches may draw upon data sources such as interviews, observations and focus groups, and apply methods which would appear unscientific and subjective to the quantitative paradigm. It is therefore necessary for qualitative and quantitative studies to apply parallel but distinct cannons of rigor if integration of the two approaches is to be achieved (Strauss & Corbin 1990).

4.3.2 Benefits and Challenges of Mixed Methods Approaches

Guba (1987) argues that the differences between quantitative and qualitative paradigms means they cannot be reconciled, ‘‘the one precludes the other just as surely belief in a round world precludes belief in a flat one” . However, others argue that the different approaches can be used for complementary purposes (Sale & Brazil 2002). Howe (1988) suggests that research needs to go with what works, and that truth, is a normative concept like good. Truth is “what works”; and quantitative and qualitative researchers should “embrace positivism coloured by a certain degree of interpretivism” (Howe 1988). Although Sale & Brazil (2002) suggest overlooking the construction of truth and reality is for those not aligned within a specific paradigm, this thesis will demonstrate that appreciating the ontological perspectives of both paradigms is possible. Instead of using the strengths of one method to cover the weaknesses of another, the approach taken in this thesis was to use the qualitative and quantitative approaches to complement each other and provider to a richer, comprehensive understanding of the same phenomena.

There are arguments for combing quantitative and qualitative approaches. Both approaches share a common goal of understanding the ‘world’ (Haase & Myers 1988), and, it is argued, they share a unified logic which can be extended to make sure the same rules of inference applies to both (King et al. 1994). Both the quantitative and qualitative paradigms share the tenets of facts, fallibility of knowledge, and a value-ladened process of inquiry. They also strive for rigor in the process of research (Reichardt & Rallis 1994). Casebeer & Verhoef (1997) argue that quantitative and qualitative methods can be chosen as part of the continuum of research and that specific techniques of data collection or analysis need to be based on the research questions. Indeed, a mixed methods approach reflects the complexity of the study and requires the combination of data from a number of perspectives in order to uncover and understand what is being observed (Clarke & Yaros 1988).

However, using quantitative and qualitative sources of data within a single study means that attention has to be paid to how they can be combined in order to provide valuable analytical insights. Depending on the discipline of the researcher, mixed methods approaches often lean towards relying on 'familiar methods' so that within the physical sciences this can mean downplaying or marginalising findings from qualitative analyses in favour quantitative and statistical ones (Sale & Brazil 2002). A typical example of this occurs when qualitative data is used for a pilot study to design a quantitative measure which is seen as delivering verifiable findings (Morgan 1998). The danger with this privileging of quantitative methods is that it can reduce the strength of the qualitative data in its own right, and risk losing the context, richness and insights that the data provides, and therefore limit the understanding generated.

The tendency to deduce meaning from quantitative methods causes concerns with regards to how the findings may not reflect or be meaningful to the individuals being studied. Whilst quantitative studies may be generalizable and potentially provide statistical prediction, qualitative analyses help unpack emic views which offer explanations (Glaser & Strauss 1967). Thus, qualitative data readdresses this imbalance by providing context to findings and a level of granularity which offers a rich insight into human behaviour.

The increasing use of analytical computer software can further add to the danger of reducing context and richness in qualitative analysis because these often represent the data using quantitative measures and metrics. Rather than rebalancing the two paradigms, quantification of qualitative findings undermines the qualitative research and may not allow this research to be 'taken seriously' in the world of positivist research (Sale & Brazil

2002). Working within the boundaries of the qualitative and quantitative paradigm ensure findings reflect the capabilities of the data, which is particularity important for qualitative research as it will help provide rich insights 'without slipping on the mantle of quantitative inquiry' (Smith & Heshusius 1986).

Despite these challenges, mixed method studies have many advantages. They can provide a comprehensive, richer understanding than one derived from one side of the qualitative or quantitative divide. This thesis makes use of this approach by incorporating the wealth of data available on the Web as well as drawing on human experience and activities. This was achieved by a well-structured and theoretically grounded methodology that is aware of the boundaries and limitations of the methods and techniques employed (Reichardt & Rallis 1994).

4.4 Justification of Case Design and Mixed Methods

Case studies have been shown to be an effective solution to tackling the type of research questions that this thesis is driven by (Schell 1992) (Yin 2003); they are not restrictive in scope, the kinds of data that they draw on or analytical methods. Bearing in mind the research questions and the analytical differences and practical limitations, a single case study design was adopted for this exploration of the growth of the Web. This was seen as having the ability to provide an in-depth descriptive and explanatory analysis which would offer insights that could be applied to understanding aspects of the Web beyond the chosen case study.

As the thesis wishes to answer the question of what the Web is and how it grows, methods that could interrogate large quantitative sources of Web data as well as capturing context rich qualitative sources of human activity were required. As previous chapters have argued, both micro and macro-level sources of data are required in order to understand the social and technical processes of the Web's growth. Mixed methods provide a way to follow the statistical and more qualitative (Erzverger & Kelle 2003) (O'Cathain et al. 2007) traces of humans and technologies across and between various online and offline mediums (Latour et al. 2011). A concurent mixed methods approach was chosen to avoid the risk of dominating the analysis by either quantitative or qualitative perspectives and the aim was for data sources to complement and support each other (Edwards & Crossley 2009) (Mische 2003) (Emmel 2009) (Heath et al. 2010).

4.4.1 Using Mixed Methods with Actor-Network Theory

The previous chapters have argued that the theoretical lens provided by Actor-Network Theory is helpful for understanding the Web as a socio-technical phenomenon. This section explores how mixed methods can be used with this rhetorical approach. Case studies have been used in several ANT driven studies including early research by Woolgar and Latour investigating how scientific work is conducted (Woolgar & Latour 1979), to more recent studies which look at the success and failure of technologies (Gunawong 2010), social processes (Chen 2009), and policies (Zhan 2010). As these studies have demonstrated, case studies paired with context rich qualitative data sources and methods used to collect and analyse them were aligned to the strengths of ANT. In these studies qualitative methods such interviews, observations, and documents offered a way to follow actors and provide insight into their activities and interactions at a granular level. More recently mixed methods approaches to ANT research which draw upon large-scale quantitative sources of data have been used to follow the digital activities and interactions of the actors (Latour 2010) (Latour et al. 2011) at a scale not possible with qualitative data alone. This helpfully provides a precedent for pairing qualitative and quantitative sources of data with Actor-Network Theory to examine the Web's growth combining social and technical perspectives and methods.

4.5 Summary

This chapter outlined and justified the use of a mixed methods single case design for the thesis research. The choice of a single case study was informed by the research questions and practical limitations. The mixed methods were chosen to take advantage of the strengths offered by combining quantitative and qualitative data and the apparent suitability of this approach for working with the socio-technical lens provided by ANT.

The thesis now turns to the case study itself and describes this and the data sources used to answer the research questions.

Chapter 5

The Case Study: The Open Government Data Network

5.1 Introduction

This chapter describes the case study, the data and analytical methods used within this thesis. Building up the case study design and mixed methods approach described in the previous chapter, the chapter begins by discussing the criteria used to select the case, which is then followed by a review of the methods used to collect, manage, and analyse the data.

5.2 Case Study Selection

The Web supports different types of activity which can are typically associated by common terms, i.e. online shopping, online banking, social networking, etc. (Beauvisage et al. 2007) (Dourisboure et al. 2009) (Weinreich et al. 2008). As the previous chapters have described, using computational approaches the identification and categorisation of different Web activities is possible using community clustering techniques which examine network traffic and linking structures within the Web graph. These quantitative methods provided a method to identify the existence of different Web communities, which illustrate that the Web has a 'fractal' structure (Dill et al. 2002) consisting of multiple networks of activity.

Given its scale, a study of the entire Web is impractical due to the resources and time-frame required, but it is possible to choose a discrete area of activity on the Web as a case study to explore social and technical interactions. Indeed, using case studies has been the basis for other studies of the Web, ranging from examining the coordination of nascent Web communities (Petricek et al. 2006)(Bird 2011), to the decline and failure of others (Rossi 2008) (Gunawong 2010). Based on the assumption that the Web can be studied in terms of different technical Web activities, the following section examines the criteria used to select the case used within this thesis.

5.2.1 Choosing a Suitable Web Activity

The study required data from a bounded Web community that contained enough resources to explore the co-constructive interactions of humans and technologies.

The following criteria were used:

1. The activity should be an active network with social and technical features. This would ensure that data were timely and reflected the socio-technical features of the Web.
2. The Web activity studied needs to be emerging and growing. This enables prospective analyses and avoids some of the data collection limitations surrounding access to historical and archived data sources. Studying a recently formed and growing community will increase accessibility to qualitative and quantitative data sources.
3. The Web activity should have an active community both online and offline so that it is possible to explore the links between these and their relationship to network formation and sustainability. Studying a community that produces outputs will also increase the chances of available data for analysis.
4. A final consideration is more practicality. The case needed to fit the resources available and the limitations of time.

Taking these criteria into consideration, the United Kingdom Open Government Data (OGD) community was chosen as the case study for this thesis. Other communities were examined as potential candidates for this thesis including other international OGD communities, the scholarly publishing community, and a community surrounding a human computation Web system. However given the resources available and the characteristics of UK OGD community matching the criteria, the UK OGD Web community was decided as a suitable candidate to help answer the research questions proposed. This emerging community is a current and contemporary Web activity and included the development of social practices and digital technologies. It has the added advantage of drawing on members of a range of different groups political sphere, academia, technology, broadcast media etc.

The UK OGD community is an interconnected grouping of individuals and organisations engaged in various on- and off-line practices designed to improve transparency and access to a wide variety of different types of data produced by local and national government. The activity associated with UK OGD included the development and promotion of various technologies for classifying, storing and interrogating data. The UK OGD is engaged with a variety of social agendas, driven by a common interest in 'opening up' and publishing UK government data to provide transparency to citizens nationally and potentially around the world (HM Government 2012). This process involves changing the social and technical approach of how government data is organised and archived. Disrupting practices deemed to lack transparency is part of the 'mission' of OGD, and considerable effort has been devoted to finding technical solutions to 'free' the data and transform it into formats that are open and accessible. The UK OGD community appear to engage in activities, online and offline. UK OGD involves the use of new technologies and aims to disrupt a wider set of actors, including government, media, researchers, and 'normal' users of the Web.

From a practical perspective, the emergence of UK OGD was advantageous in terms of access to different sources of data. Many of the actors involved during early development were potentially accessible online and offline (in interviews). In addition to this, there were several large repositories of documentation and communication data that the analysis could draw upon for analysis.

The UK OGD community met the criteria outlined above and appeared to be a useful case study to answer this thesis's research questions. The next section will provide an overview of the UK OGD Web activity, including the configuration of the offline and online network of actors.

5.2.2 Defining the Case: The UK Open Government Data Community

The following overview of the UK OGD community has drawn upon the empirical evidence gathered, and describes the characteristics of the community at the end of the study. The intention of this section is to provide an overview of the case and specific aspects will be explored and analysed in more detail in subsequent chapters.

The UK OGD community can be broken down into several subsections. From a top down perspective, the community consists of a number of high profile interest groups including government, academics, industry partners, civic society, organisations and also various UK based companies and businesses. Separate but not disconnected from these actors are individuals who are involved with the activities surrounding the technical implementation of data.gov.uk, which is the UK government's national data portal to store both national and local government data, and similar government-funded technology projects designed to open up data. Journalists and national media are also represented as they report OGD activities as well as their own ventures in using OGD to inform their reports. Finally, various more technically-focused actors including programmers and developers interested in Semantic Web and Linked Data technologies which include individuals and groups, who are representatives of academia, industrial and research groups, are included in this community. These actors may not be confined to the UK, and they were linked to the UK OGD case study despite their geographic location.

Examining UK Open Government Data further, a network of sub-communities and interest groups are visible; these include actors involved within the national level (or central government) OGD activities such as those responsible for enacting national by policies and initiatives. These actors include individuals working in local councils and government departments, who may be scattered geographically across the UK. Figure 5.1 illustrates the structure of the UK OGD community. These can be divided into three parts relating to the UK governmental organisational structures: national, regional, and county. Operating between these parts are other actors that use open data, including developers, businesses and smaller organisations, as well single or activists or small groups e.g. lobbyists, and Open Data.

The online presence of the actors involved in the UK OGD community can be traced via the development of Web applications and tools, and in the use of social networking sites and 'Web 2.0' services, notably Twitter and WordPress, as well as in documents such as national news articles, blogs on government personal websites. The OGD community participate in a range of online and offline activities some of which have already been described as part of the theoretical underpinning of this thesis. As the online and offline presence does not necessarily signal separate networks, and as subsequent chapters will show these groupings and activities are all part of the network of interwoven associations and digital traces (Latour et al. 2011).

Figure 5.1 provides a starting point to position the data collection, but it must be emphasised that associations and interactions are illustrative and do not portray a fixed, rigid arrangement.

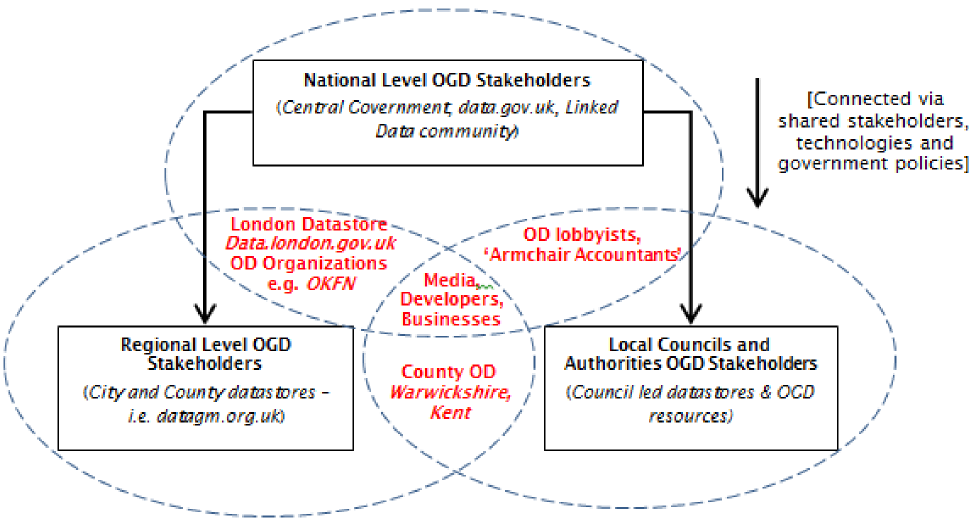


Figure 5.1: Structural Overview of the UK Open Government Data Community - Data providers at the National, Regional and Local Level, and the Data Consumers (Red Labels)

5.2.3 Quantitative and Qualitative Data Sources

This section will describe the data collected for the analysis of the UK Open Government Data community. In line with the mixed methods approach, a collection of both quantitative and qualitative sources were used.

Quantitative Data	Qualitative Data
Social Network (Twitter) Data associated with the #opendata hashtag/keyword	Interviews with 24 individuals within the UK OGD community
Data Repository Record Metadata	Documentation from Web Sources
Activity Timeline data	Observations of UK OGD related events

Table 5.1: Quantitative and qualitative data sources to support UK OGD case study

Table 5.1 shows six sources of data collected; these include three quantitative data sources and three qualitative data sources.

A considerable amount of qualitative data was potentially available for the study including: archival records and other documentation, interviews direct observations, and physical artefacts such as the development of technologies and technical standards (Yin 2003). These forms of data thus included text, pictures, and multimedia content such as videos and audio materials. The following sections describe the methods used for the qualitative and quantitative data collection.

5.2.4 Interviews

Interviews are a common source of data within qualitative research. Interviews can be designed and developed in specific ways depending on their purpose; they can be scripted or participant-led, formal, conducted either face-to-face or mediated by digital technologies telephone or Skype for example. The choice of type of interview is tailored to the to the research design used (Yin 2003).

For this study, semi-structured interviews were conducted with different individuals within the UK OGD community. These comprised of approximately one hour discussions, and were conducted face-to-face, or by a telephone or Web video chat. The interviewees were asked a series of open-ended questions regarding the emergence and growth of the UK OGD community, their role, the use of technologies, and their thoughts on the further development of UK Open Government Data. These questions are shown in Appendix A.

The selection of the interviewees was based on an initial snowball sampling initiated by the first interview conducted. Contact with the first interviewee was achieved via contacts with the researcher and research department. As the analysis in the following chapter demonstrates, the selection of additional interviewees was guided by those being interviewed, with specific questions related to asking interviewees for names of individuals who would be able to provide answers to the questions of the research. Furthermore, an association network (based on the names of individuals who interviewees talked about) was extracted and used as part of an iterative process to identify individuals who were discussed during the interview; by forming a graph of mentioned individuals between interviewees, it was possible to draw out those who were potentially influential within the discussions.

In addition to this, the use of purposive sampling informed by other data sources to try to ensure that a variety of different respondents were included in order to reflect the different sectors identified during the scoping of the UK Open Government Data community. The selection aimed to achieve an unbiased sample of the different communities within the UK OGD community which as Figure 5.1 shows, extends across government, academia, civic society, and business. As the following section on integrating data sources demonstrates, the use of the social networking data was used to support the process of finding interviewees, and was essential for discovering those that were prominent or mentioned from the interviewees performed.

Although methods have been used to reduce the bias towards choosing only individuals who are of high profile, this research recognises that there is the possibility that some individuals or organisations of the UK OGD community may not of been represented. However, taking this into account, and in an attempt to reduce the risk of missing potential valuable sources of data, the research performed a number interviewees until data saturation was reached.

5.2.5 Documents

Documents offer another source of qualitative data, and can include sources such as written communications between actors, formal and informal historical records (ranging from parliamentary sources to diaries), official documents and or less structured ones such as memos and meeting notes (Denzin & Lincoln 1994). Documentation provides another layer of information for analysis, and although textual information is common, documents can be multimedia images, videos, and audio transcripts.

For this thesis a series of documents related to events or activities were collected between January 2009 and December 2012 (Appendix C). Each document contained information about the activities and/or social and technical developments of the UK Open Government Data community. The process of identifying and collecting these data was informed by the interviews which ran in parallel to the data collection. Data extraction was supported by quantitative keyword analysis and data harvesting. Documents were collected from multiple Web sources, including blogs, online news print, national and local government Websites, discussion forums, and mailing lists. Also, as a consequence of references made by interviewees and external links identified within the collected documents, a number of additional documents were identified that predated January 2009 (Appendix C).

5.2.6 Qualitative Observations

Observations also feature as a common source of data within qualitative studies, providing an understanding and insight of the participants' behaviours and the context which may not be captured so well other sources of data (Yin 2003). Observations may reveal the 'unspoken' processes and interactions between actors. Qualitative observation can take participant and non-participant forms but in line with other ANT-inspired approaches (Tatnall & Burgess 2004) both approaches were used for this research. It was possible to participate in and remotely observe different events and activities associated with the OGD community.

Participant observation was undertaken at various events related to the activities of the UK Open Government Data community, including:

- The 'Removing the roadblocks to a pan European market for Public Sector Information re-use' workshop (May 10th 2011) - A workshop with attendees from the UK Open Government Data community discussing the current adoption of Open Data (UK and European)
- The Open Government Data Camp 2011 (October 19th 2011) - A conference with various presentations, and workshops regarding Open Government Data

- The Open Knowledge Festival 2012 (September 16th 2012) - A conference with a specific track related to the activities of Open Government Data

In addition observations were made at other workshops, tutorials, and presentations. During and subsequent to the events, detailed notes were recorded about what had happened, the topics covered and their relevance to the activities of the UK Open Government Data community.

5.2.7 Web Data Mining

In addition to these qualitative data the Web contains other data sources, typically large sets of numerical data which can be collected, transformed and analysed. This represents another form of observational data referred to here to distinguish this from qualitative data collection - as data mining. Scholars such as Barabasi & Albert (1999), Gibson et al. (1998), and Watts (1999) have harnessed large scale Web hyperlink data and paired it with appropriate analytical methods to provide an insight into the structure of the Web graph.

Web data mining is the traditional method used to obtain large-scale data (Han & Kamber 2000) (Chakrabarti & Faloutsos 2012) and has been used within different kinds of Web studies (Albert et al. 1999)(Chakrabarti 2002)(Dourisboure et al. 2009). This entails the development of computer programmed scripts to harvest the data. Bespoke programmed scripts that perform 'data scraping' were used, which either performed manual scraping techniques or connected to a structured Application Programming Interface (API), which is a service offered by different Web platforms to extract volumes of data without the need to interact with the graphical user interface. Essentially the API provides a way to gather large amounts of data in a simple and structured fashion. In addition to this, the analysis used techniques of Web scraping, which involves a series of iterative processes to download the content of a Web page (or collection of Web pages), and then the process of extracting and storing the data required in a suitable format.

Data mining techniques were used within this study to collect a number of different Web data sources that contributed as a source of Web observations for the analysis conducted. These sources included the collection of social networking communications data from the Web social network platform, Twitter, repository data collected from the UK government's national data portal data.gov.uk and other local government data portals including the London datastore, Greater Manchester datastore, Kent Connects, and Warwickshire datastore. In addition to this, data mining techniques were used to collect data related to the activities of the UK OGD community from various Web sources as listed in Appendix C. These sources were chosen for their capability of capturing the online activities of the UK OGD community, which included the communications made by individuals within the community and outcomes (data) that the Web activity produced. The following sections will describe these three sources of data in more detail.

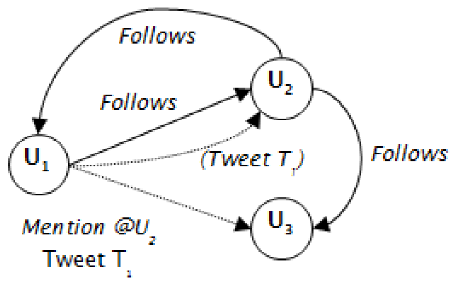
5.2.8 Social Network (Twitter) Data

For this study, data mining techniques were used to collect a dataset from the social network platform, Twitter. The data was harvested using the Twitter API and contained the communications (messages) made between Twitter 'users', as well as the 'static' social network graph data, the latter of which can be transformed into networks of online 'friendship' and 'follower' associations. This represents the static links between individuals, i.e. a friendship/follower network, which can be interrogated with statistical metrics to examine structure and network properties.

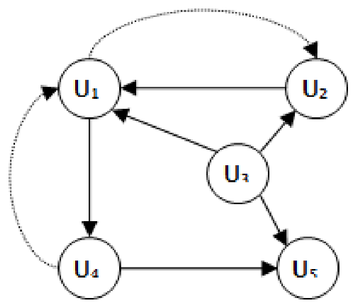
5.2.9 Twitter Messaging Architecture

The Twitter service provides users the option to post 140 character messages known as tweets to their own timeline of activity, which can be public or private. As part of the user model (and user experience), users can 'follow' each other, which creates a directed link between two users, as shown in Figure 5.2(a) The 'following' of a specific user results in the followed user's tweets appearing in the 'following' user's personal timeline. Depending on the followed user's notification settings, they may be told that they are being followed by a specific user, but it does not require them to 'follow' them back. This therefore results in a directed network of users (nodes) and followings (edges).

Twitter also enables users to write tweets with a 'screenname' of a Twitter user included as part of the message performed using the '@' symbol. As a result of this, the user that has been 'mentioned' in the tweet will view the tweet in their timeline. This



(a) 'Following' and 'Mention' Network Structure



(b) Twitter Retweet Functionality

Figure 5.2: Twitter messaging architecture Static and dynamic communication model

therefore produces another network structure of users (nodes) connected together by tweets (directed edges), as shown in Figure 5.2(a) Sharing tweets is also altered by the network of user's followers, for example, if 'user A' follows 'users B' and 'user C', and 'use B' creates a tweet mentioning 'user C', consequently, that tweet will also appear on 'user A' timeline.

Similar to the mention function, the retweet function enables users to copy tweets from the timeline of the users they follow and add them to their own timeline. Twitter defines the retweet function as:

> *"A Tweet by another user, forwarded to you by someone you follow. Often used to spread news or share valuable findings on Twitter."* (Twitter 2013)

As a result of this, followers of the user that has retweeted will see the retweet (and a link to the retweeted user) on their own timeline. This therefore creates a network where users (nodes) are connected by retweets (directed edges), but through a reverse directed graph, as shown in 5.1. For example, if U2 retweets U1, then U3's timeline will become populated with the tweet U2 has retweeted.

The collected Twitter social network dataset comprises of two forms of data. The first contains 623,061 unique records of messages sent between individuals, collected between 1st January 2009 and 31st December 2012. A second set of Twitter data with 1925 tweets between January 2006 and December 2008 was also obtained. However, these were subject to a earlier Twitter architecture, thus they were excluded from the structural analysis performed in chapter 6. The collection of tweets between January 2009 and December 2012 was achieved by using the Twitter API to filter tweets which contained the keyword '#opendata', which was identified in the interview data and Web documents as a communications channel for discussions involving the UK OGD community. Each record (tweet) within this dataset comprises of:

'TwitterID','TwitterUsername','TwitterUserID','Timestamp','TweetText'

The 'TwitterID' represents a unique numerical identifier for each record, which is generated by the Twitter Web service. Similarly, the 'TwitterUsername' and 'TwitterUserID' are unique to an individual with a Twitter account. The timestamp represents the date and time that the record was made, contained in a ISO 8601 format (YYYY-MM-DDTHH:MM:SSZ). The 'TweetText' contains the body of the communications, which included external URLs (hyperlinks) to other Web services, Twitter usernames, and hashtags (#), which are embedded linkable keywords. In order to extract specific content and community structures from the communication dataset a number of information diffusion and clustering algorithms were applied (Appendix D) to the dataset.

The second form of Twitter data represents the offline static 'friends' and 'follower' network, which was collected secondary to the initial communications data. This data comprises a comma separated value list of friends and followers generated on an ad-hoc basis depending on the requirements for the analysis. Each record within the list contains the following data:

'TwitterUsernameA','TwitterUsernameB'

This represents the friendship or following relationship from individual to another, as the relationship is not reciprocal, the list represents a directed relationship.

5.2.10 Data Repository Record Metadata

Data mining techniques were also used to collect other forms of Web data which contain evidence of human and technological activity and interaction. These data sources were collected due to their richness in terms of context and content and they included data streams from specific Web services that are related to the object of study. The datasets collected contained metadata from a collection of different government data repositories and data portals within the United Kingdom, including:

- data.gov.uk - UK's national open government data portal
- London Datastore Greater London's Government Data portal
- Warwickshire Datastore County of Warwickshire Data Portal
- OpenKent County of Kent Data Portal
- OpenGM - City of Greater Manchester Data Portal

For each of these data sources, the accompanying metadata for each record was collected, which contains different fields such as: the data and time of the published dataset, the dataset owner, the dataset content/subject, the licence (if any) used, the format(s) (i.e. CSV or PDF), and additional notes of the dataset.

5.2.11 Activity Timeline Data

By using Web scraping techniques, a seeded list which contains Web sources was used alongside a keyword filtering algorithm in order to gather specific resources, which was then transformed into different formats, which included activity timeline of events. The activity timeline data came from different Web data source which recorded offline and online activities of individuals within the UK Open Government Data community. As shown in Appendix C, each record contains a timestamp, the source, the description of the activity, and also an external link to where the activity was found (not shown in appendix). This data was collected by using keyword harvesting and crawling techniques in order to find the data sources and the individual records.

In addition to this, an additional activity timeline was constructed using data from OpenlyLocal.com, a website which contains records of the number of UK councils producing open data, and their accompanying license. For each timestamp snapshot collected, each record details regarding the council name, geographic location, and the availability of open data. Using computational techniques, this was transformed into a time series representing the activity of UK councils, as well as a geographic illustration of the data.

5.3 Working with the Data

This thesis made use of a range of different data sources, tools, and techniques, to provide rich and in-depth understanding of human and machine interactions at both the micro and macro level. Data analysis techniques play an important role in being able to exploit the various forms of collected data to reveal the socio-technical processes of the UK OGD Web activity. Using appropriate analytical methods with the data sources was essential to draw out insight into an individual set of data, and also provide a set of findings that enabled the cross-comparison and triangulation across data sources.

In order to analyse the data it must go through a series of processes to ensure that it is in a suitable format and structure. These processes included: coding, structuring, transforming, and sanitising. The terminology here differs slightly for the quantitative and qualitative analysis but the aim is to enable analysis that ensures a fair and consistent representation of the data. A mix of manual and automated techniques were used for these processes, some which were assisted by software analysis tools, which helped reduce the time required and problem of human error. The following section examines the qualitative and quantitative techniques for organising and managing the data.

5.3.1 Qualitative Data Coding

A standardised approach to organising and analysing the data is required, ideally one that allows in a mixed method study the integration of analyses.

The qualitative data comprised of various formats including interview transcripts, Web documentation, and qualitative observations. In order to manage this information, data coding techniques described by Ritchie & Lewis (2003) were employed to provide a structured and logical way to formally interrogate and categorise content.

The data coding process assigned codes to sections of text within an interview transcript or other forms of documentation and then used a framework approach to refine and develop this into categories and themes. A coding schema was developed by noting down inductively individual codes from the data. For instance, an example code labelled 'Network Interactions (Events)' was used, this represented sections of text which related to an event related to the UK OGD community. This was also informed by the theoretical categories offered by ANT so that the analysis was also deductive. In order not to neglect the strong theoretical basis guiding this thesis, as illustrated in Figure 5.3, the coding was first performed using a topic guide that helped extract content regarding the socio-technical processes driving the Web's co-construction. Based upon this initial set of codes (see Appendix B), an iterative process was used to go through the data and codes to identify common themes and concepts. These were then refined and regrouped forming a hierarchical coding schema, where codes were nested within themes and sub-themes. This provided a method to link specific coded section together, for instance within the branch of 'Network Interactions', there existed nested codes of 'Events', 'Meetings', and 'Serendipitous Encounters'.

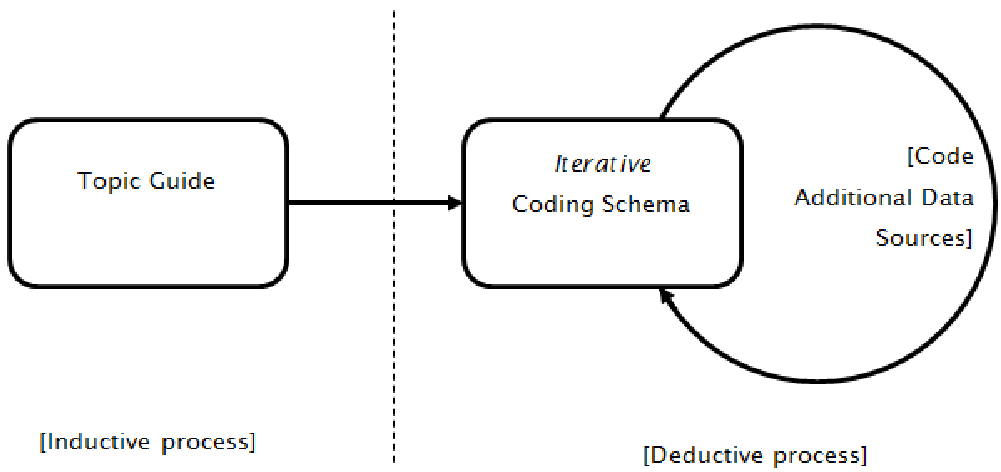


FIGURE 5.3: Inductive-Deductive Coding Schema

This was further developed by creating a meta-coding schema that mapped to the iterative coding schema, as shown in Figure 5.4. Ultimately, organising the data using this method enabled a way to compare and triangulate different data sources. This coding was the first step in extracting relevant information regarding the socio-technical processes, and offered early insights into the growth of the Web at a granular level. The early process of coding fed into data collection for example by informing interview questions. The next following section describes the techniques used for organising the quantitative data.

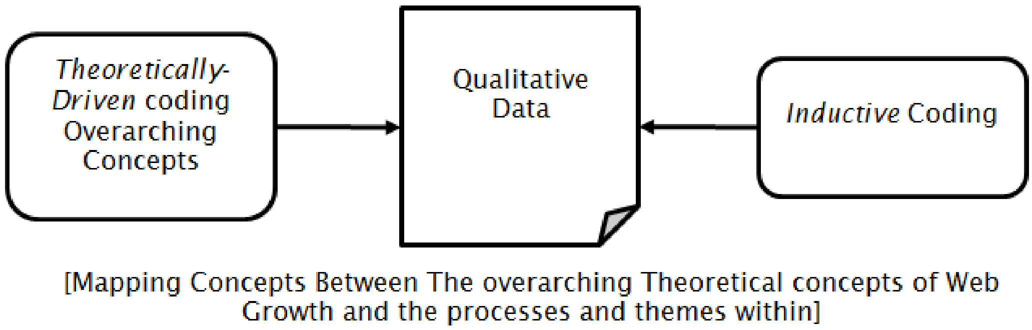


Figure 5.4: Dual-Layer Coding Schema

5.3.2 Quantitative Data Management

Quantitative data needs to be examined for consistency, structured in a suitable format, and if required, merged with other datasets. Unlike the qualitative data, the size of data collected presents obvious difficulties for manual processing. In order to deal with the scale of the data, computer programme scripts and tools were used to efficiently and reliably transform and format the data ready for analysis. The development of computer programmed scripts was used on a case-by-case basis based upon the requirements and formats of the collected data. In order to provide a consistent set of data for comparison, a specific set of formatting standards and data transformation practices was developed.

The quantitative data contained different forms of record data which were either related to specific interactions between actors, events in time, or some form of data resource. However a common attribute within the metadata (additional data fields which are associated with a specific record or dataset) of these data were timestamps (date and time) related to the time that the record was created. This was one of the most crucial fields within the records, and was critical to ensure that the structure of the data was consistent.

In order to ensure data validity during the data transformation process, any records that were missing a date or timestamp field were processed, however these records were flagged in order to indicate that the record was incomplete, missing fields, or incorrectly formatted. In addition to this, specific datasets contain their own unique fields or attributes, which when missing reduced the integrity of a single record, or possibly the entire dataset. These missing data were flagged and dealt with on a case-by-case basis. If the missing field was a unique identifier of a record (for instance, the ID of a Tweet), then a string was generated and inserted in order to enable the record to be used. However, if a timestamp is missing from a record, two techniques were used; generating the correct timestamp would be attempted, and if this failed, then a 'flagged' ISO 8601 timestamp with the value of '1900-01-01T00:00:00Z', was assigned. This provided a way to identify and exclude records if time series analysis was required.

Ultimately, the aim of transforming and formatting the quantitative data was to ensure that the data was consistent and reliable for the analysis, and by using a common identifier such as the timestamp provided a format that could be used in combination with the qualitative data sources in order to triangulate data sources. Finally, throughout the transformation and structuring process a key consideration was to ensure that the data were as readable as possible by the researcher, whilst still maintaining the data's completeness, as well as machine-readability.

5.4 Analysis

The following section will describe the qualitative and quantitative analytical techniques used in order to expose the richness and insight that the collected data for the case study contained.

5.4.1 Quantitative Analysis

The quantitative analysis of the Twitter social network data presented in the next chapter was conducted in order to understand the structure and evolution of the UK OGD community. To achieve this, two approaches were used:

1. The static analysis of an emergent network: using metrics to calculate the end-state of the network, providing the equivalent of a snap-shot of a particular data point. This approach does not concern itself with the processes which took place to form this end-state; it provides an analysis of the overall structure of the network by using metrics of network size, connectivity, degree, and modularity.

2. The dynamic analysis of a network: this offers a series of observations of the changes within a network's growth over time. The growth of a network can be examined, along with the structural changes that emerge. The incremental snap-shots of the network over time provide an indiscreet set of metrics and a dynamic visual account of the network's growth.

   In order to achieve these analyses, social network analysis techniques were employed. The following section will provide an overview of social network analysis, and how it was used within the analysis.

5.4.1.1 Social Network Analysis

Social network analysis (SNA) can be originally accredited to scholars within the social sciences who focused on 'social relations' between humans (Burt 1978). Early research in SNA built upon the methods and techniques developed from the mathematics of graph theory (Moreno 1934) and anthropological network studies, which harnessed observational, narrative, and visualisations techniques to describe the social relations (Barnes 1954) (Mitchell 1969). Recently, the development of computation techniques to map and measure the structure of the networks has meant that qualitative approaches to SNA are less popular (Carrington et al. 2005) (Heath et al. 2010). New quantitative techniques allow analysis of social networks are a greater scale (Hesse-Biber & Griffin 2012) (Manovich 2011). The uptake of quantitative approaches to SNA has led the mainstream discussion of SNA to drive towards quantitative techniques (Heath et al. 2010).

Building on the techniques and methods developed for exploring the Web graph and social networks (Barabasi & Albert 1999) (Watts 1999) (Scott 2000) (Dourisboure et al. 2009) (Tinati, Carr & Hall 2012), the analysis used graph and social network metrics to investigate the static and dynamic community structures of the social networking data collected. These techniques provided a way to understand the emergence and growth of online networks, which included the analysis of the patterns of online communications between actors and the network-like structures that were produced. These quantitative findings were used to examine the growth and structure of the activities at scale.

In addition to this, recent algorithmic techniques developed to identify different user roles within social networks (Tinati, Carr & Hall 2012) and also trace the diffusion of information and embedded content within Twitter communication channels (Tinati et al. 2013) were used in combination with the qualitative data in order to examine the formation, growth and interactions between different communities (relative to Figure 5.1) within the UK OGD Web activity. These recent methods build on traditional social networking and graph theory techniques and apply them to dynamic, real-time streams of communication data, which help filter and elicit information from large sets of time series communications data.

In order to demonstrate how the qualitative data was used in the analysis, the following section will describe the analytical techniques that were used with the qualitative data, and also the techniques used to combine the findings with the quantitative social network analysis.

5.4.2 Qualitative Analysis

Qualitative analysis techniques were used with the interviews, Web documents, and observations, to:

1. Complement and support the quantitative analysis, which includes various forms of social network analysis and statistical analysis.
2. Provide a qualitative description of a translation process; focusing on how translation occurred as a way of explaining how the Web grows.

The analysis drew upon the quantitative and qualitative data to explore processes at the micro level, and the structures that emerge at the macro level. The qualitative data was used to provide context and description alongside the quantitative findings. The interviews, Web documents and observations data were used to triangulate specific events and themes that helped explain the growth of the UK OGD Web activity. To describe the emergence of the UK OGD Web activity both quantitative and qualitative data sources were used, combining content rich interview and text resources and with quantitative social network data which contained traces of human interaction and communication.

As the previous chapters have argued, in order to achieve an understanding beyond a single disciplinary perspective, methods are required which can use the collected quantitative and qualitative sources together. For SNA, whilst quantitative techniques are good at describing structure and patterns, the combination of quantitative and qualitative methods is of particular interest to understand the interactions of and processes between humans and technologies that generate network structures (Edwards 2010). There is a growing argument for using qualitative research to complement quantitative research (Edwards & Crossley 2009) (Mische 2003) (Emmel 2009) (Heath et al. 2010), with emphasis placed on mixed data sources to understand content and context of the network's associations between humans and structures that emerge (Emirbayer & Goodwin 1994) (Mische & White 1998) (Mische 2003) (McLean 2007) (Clark 2007) (Emmel 2009) (Edwards & Crossley 2009). This also draws upon wider attempts to engage with interdisciplinary research methods to harness the capabilities of computational techniques of analysis while also engaging with sociological theory (Riles 2000) (Knox et al. 2005). The methods used in early qualitative-led SNA studies are now being revisited and considered as essential components if a richer understanding of the phenomena is to be achieved (Barabási 2009) (Scott 2010). Mixed methods approaches to SNA have emerged, drawing upon an interdisciplinary set of data analysis techniques in order to produce more than just quantitative or technical findings (Hesse-Biber & Griffin 2012) (Tinati 2012)(Metaxas & Mustafaraj 2013).

Mixed methods framed the analysis conducted, and drew on concepts from Actor-Network Theory including the process of translation to describe and analyse the socio-technical activities that enabled the UK OGD Web activity to emerge and grow. By combining the descriptive analysis with the patterns found within the initial social network analysis helped uncover how individuals were interacting both online and offline, and by drawing on the Web artefacts such as the repository data, it was possible to explore the relationship between human activity and technologies.

5.5 Integrating Qualitative and Quantitative Methods

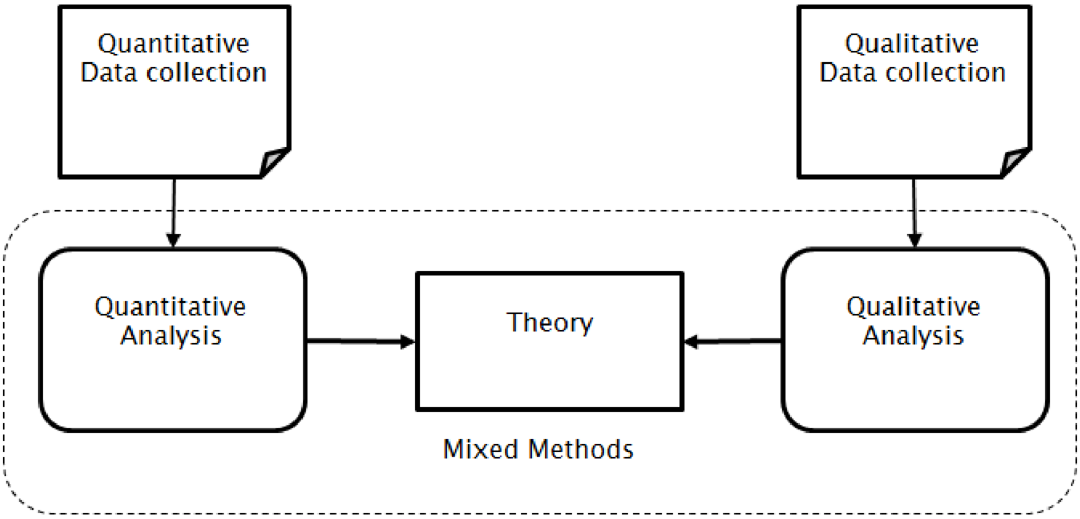


FIGURE 5.5: Mixed Methods Approach

Figure 5.5 illustrates the concurrent collection and analysis process used to combine the quantitative and qualitative analyses, informed by theory. Without sparking the debate of quantity verses quality (Boyd & Crawford 2011), the volume and scope of quantitative data collectible will be larger than the qualitative, however, qualitative data holds value in terms of its the richness and context.

Given that the study draws upon different types of data and it is important to recognise these differences but also that different data can be triangulated to provide a more comprehensive view. Triangulation is a process by which, data are examined to see if they converge (Yin 2003). Triangulation provides a way to check whether the findings agree (Denzin 1978) (Morse 1994).

> *"Without such multiple sources, an invaluable advantage of the case study strategy will have been lost"* (Yin 2003)

As Figure 5.6 illustrates, the quantitative and qualitative data were integrated at several points so that all the data could inform on-going data collection and later so that claims and interpretations could be verified. In the early phases of the analysis, integration was facilitated by comparing the interview data with the social network analysis, which explored the associations between actors within the micro and macro level of the UK OGD Web activity. This was then used to feed into the translation analysis, which deductively examined the social and technical processes that led to the translation of the Web activity. As Figure 5.5 shows, this drew on different sources of data.

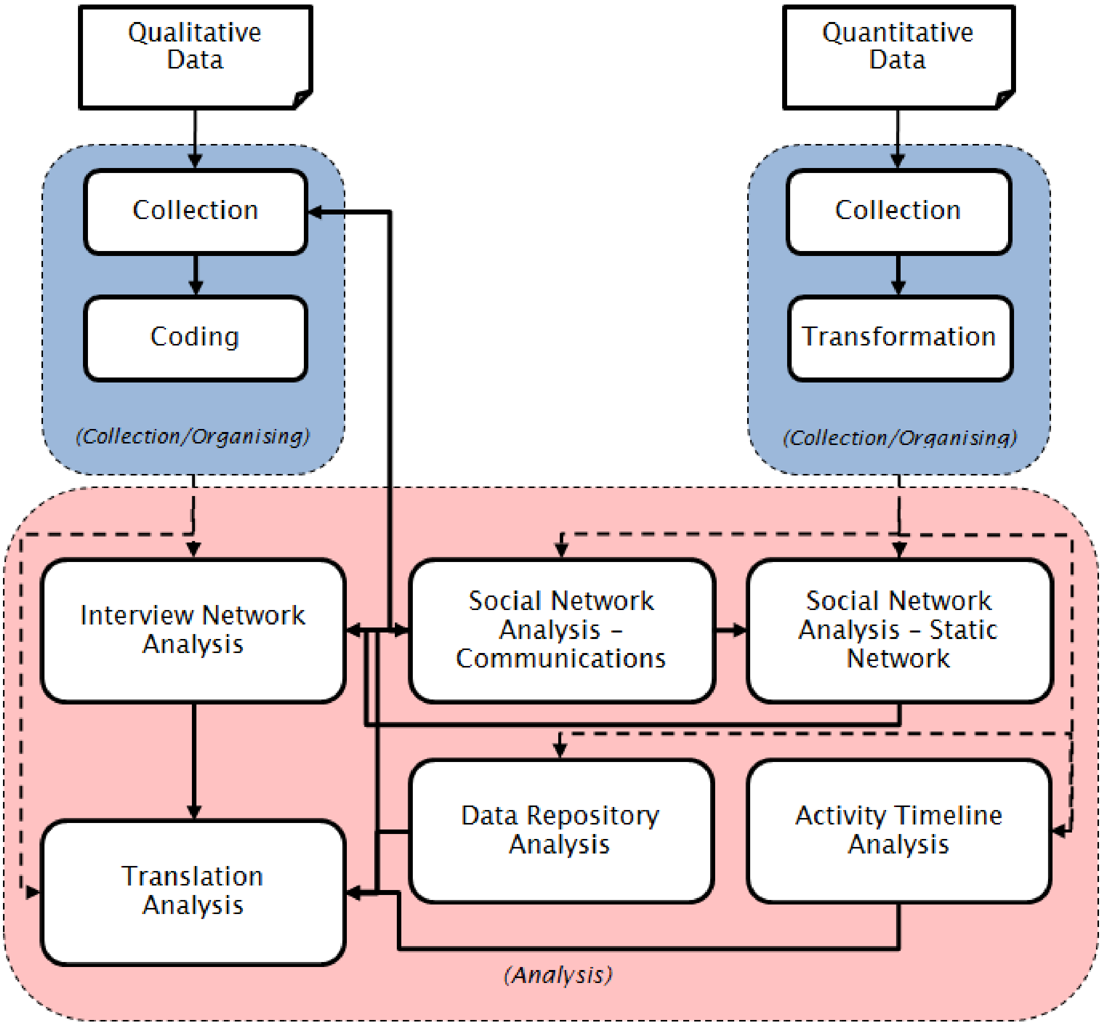


FIGURE 5.6: integrating the qualitative and quantitative data for analysis

The integration of the qualitative and quantitative analysis built on and informed the data collection process, and unlike studies which tend to reduce the qualitative data into statistical (Pahl & Spencer 2004) (Trotter 2007) in order to only explain network structure (Emmel 2009), the analysis of the social network communications helped inform and identify individuals to interview. The coded interview data informed the analysis of the social networking communications data by helping to provide context to the findings. Similarly, the analysis of the translation required the use of multiple sources of quantitative and qualitative data and based on these findings the other analyses will be fed back with additional insight.

To summarise, the mixed methods approach made use of a range of qualitative and quantitative data including: interviews, documentation, observations, social network data, Web records, and activity timelines. Data were triangulated in order to reveal processes that a single data source could not expose. Comparative analysis within and across qualitative and quantitative data sources were used to develop rich descriptions. Finally, the theoretical concepts provided by Actor-Network Theory's process of translation were used to theoretically analyse the data.

5.5.1 Ensuring Rigor and Quality in the Methods

Reliability and validity have been identified as ways to ensure that the research provides a reliable and trustworthy set of findings (Lafaille & Wildeboer 1995) (Golafshani 2003). Reliability involves the process where the research methods produce accurate and consistent results. In order to maintain reliability within this study, detailed records of the data sources were made which included: how they were collected, coded (or quantitatively transformed), and used for the various stages within the analysis. This study used tools such as NVIVO to ensure a consistent and systematic method to code and analyse the qualitative data, and analytical tools such as Gephi, NodeXL and bespoke network analysis tools (Tinati, Halford, Carr & Pope 2012) were used to process and transform the quantitative data to ensure consistency between all datasets. Records of these processes should enable the study to be replicated to check the findings (Mays & Pope 1995).

In order to ensure the validity of the research findings, the analysis was triangulated to check agreement. The use of interviews and observations raise questions about the potential for bias and subjectivity (Becker 1958), inadvertently tainting the evidence collected as the researcher's participation may influence or change behaviour of the actors within the study. Indeed, the participation of the researcher is also problematic as objectivity may be lost due to a researcher perception changing (Becker 1958), and recoding such observations are further subjected to the researchers views and position. In order to address this issue and minimise the chances of subjective outcomes, a transparent and systematic recoding of the processes were used to collect, describe, and analyse the data within this study.

In addition to these concerns, the particularity of the study also required consideration, especially with regards to the framing that the methods and the analysis produced. A number of points need to be raised here, first, with regards to the type of data being captured and used within the analysis, and how this has shaped the analysis performed. Despite using multiple data sources to analyse the growth of the UK OGD Web activity, this thesis recognises that some data sources may remain hidden, and as a consequence may provide an account richer in detail in specific places. Furthermore, the thesis recongises that the methods used to select interviewees may have risked providing a limited perspective on the OGD Web activity, yet these risks were reduced by using multiple data sources to obtain verify the statements being made.

The second point relates to the type of 'history' that is produced by studying the emergence of the UK OGD community. This is related to the first point with regards to how the types of data used will provide a specific type of account to understand the growth of UK OGD. Given that the study aims to understand the emergence and growth of a Web community, then the analysis did not intend to focus specifically on a single event or set of activities, but instead examine the outcomes of these activities and their wider effect. As a consequence of this, some activities, events and interactions may appear less significant than others, or may not feature directly in the study.

5.6 From Methods to Investigation

This chapter described how the single case mixed methods approach was used to collect and analyse data about the UK Open Government Data community. The next chapter will now provide the first level of analysis of this study, drawing upon the interview and social networking data in order to examine the associations between actors within the UK OGD community.

Chapter 6

The Network of UK Open Government Data Web Activity

6.1 Introduction

This chapter provides the first level of analysis into understanding the growth and structure of the UK Open Government Data community. By drawing upon the interview and social network data sources that were introduced in the previous chapter, a range of analytical techniques are applied to the data in order to examine what insights that they may contain to help explain the translation of the Web activity. Using the data collected, this chapter will demonstrate the growth of the online associations and communications during the emergence and development of UK OGD towards a coherent and connected network.

The UK OGD community is a complex Web activity comprising of different actors who are diverse in geography, and technologies which are diverse in application. In order to understand this complex network of activity, and in-line with the mixed methods approach developed, the quantitative and qualitative data collected will be used to examine the network of associations identified within the interview data and then Twitter social networking data.

6.2 From Interview Data to Online Communications

The section draws upon the interviewee data which represents a collection of actors within UK OGD. Figure 6.1 represents an association network graph of the dialogue within the interviews; the nodes represent different actors, and the edges represent the associations between two actors. As Figure 6.1 shows, by following the physical associations between actors, what emerges is a network of their activities and interactions. The red nodes in Figure 6.1(a) represent the selected interviewees and additional actors that were mentioned in terms of their activities or involvement with the OGD community. In addition to this, Figure 6.1(b) represents a filtered network to show the associations that exist directly between interviewees. Figure 6.1 shows that certain actors are 'authorities' within the network; and as Figure 6.1(b) shows with more clarity, these are highly mentioned actors within the network of associations. These findings are similar to Ball & Newman (2013) study of friendship network, indicating that the formation of offline friendships (or associations) exhibit characteristics of an uneven distribution of associations, similar to the formation of associations in online social networks.

However, although certain individuals have a proportionally larger number of associations than others, without examining the context behind these associations, assuming their importance would be speculation, at best. As Figure 6.1(a) illustrates, despite these actors having a higher in-degree (actors mentioning them), the number of actors that they are associated with, or simply their out-degree (number of mentions to other actors) does not reflect their in-degree value; the additional actors within the association network are the result of the associations made by the weakly connected actors within Figure 6.1(b); these are the actors that are connected to a wider network of non-interview actors.

As a method to link the network of interviewees from the offline associations to those online, the analysis draws upon the actor's associations made using Web-based technologies. Twitter played an important role in enabling the nascent OGD community to discuss and share information and news, thus this platform provides a way to expose the online associations between the interviewees, and may offer some further insight into the role of the actors within the UK OGD community.

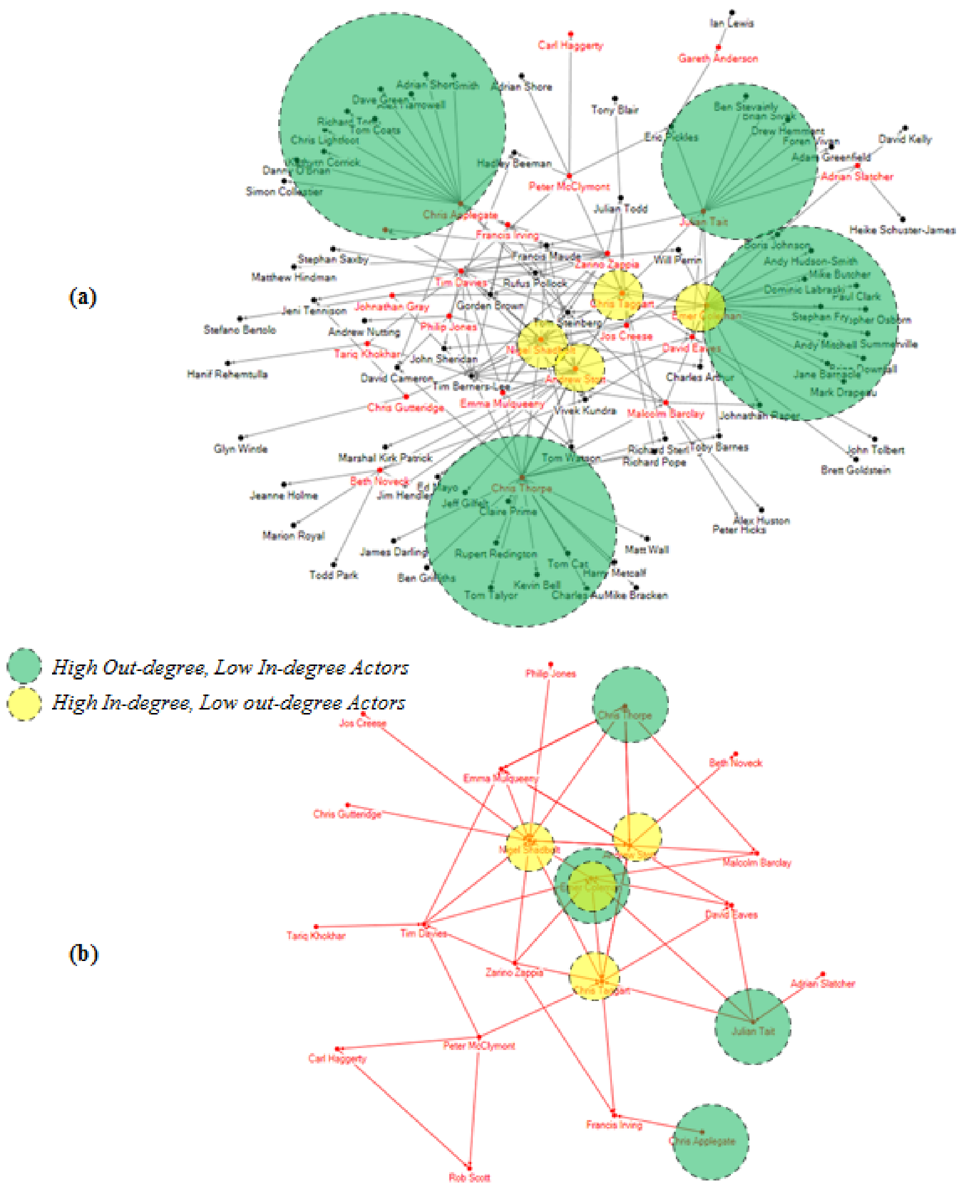


FIGURE 6.1: The network of associations between the 24 interviewees based upon the transcribed dialog. Each Node represents an actor, and the edge (directed) represents the mention or association between the two actors. (a) represents the entire association network, the red nodes represent the interviewees (b) represents the associations restricted to the interviewees. Overlaid are indicators for actors with an either high or low in- or out-degree

Using Twitter profiles of the 24 interviewees, a network was constructed using data of the interviewee's Twitter 'friends', which are the other Twitter users that a particular individual follows. As Figure 6.2(a) illustrates, this network contains 4271 users and 9038 'friend edges'. Similar to before, there are actors (including those identified in Figure 6.1) that exhibit properties of high in-degree (they have a lot of people who are friends with them), but a low out-degree (they do not necessarily form friends with others), and as Figure 6.2(a) illustrates, these actors are within the core of the network, strongly connected, but not forming centres of connectivity; these actors that have been labelled with by the High Out-Degree, Low In-Degree overlay.

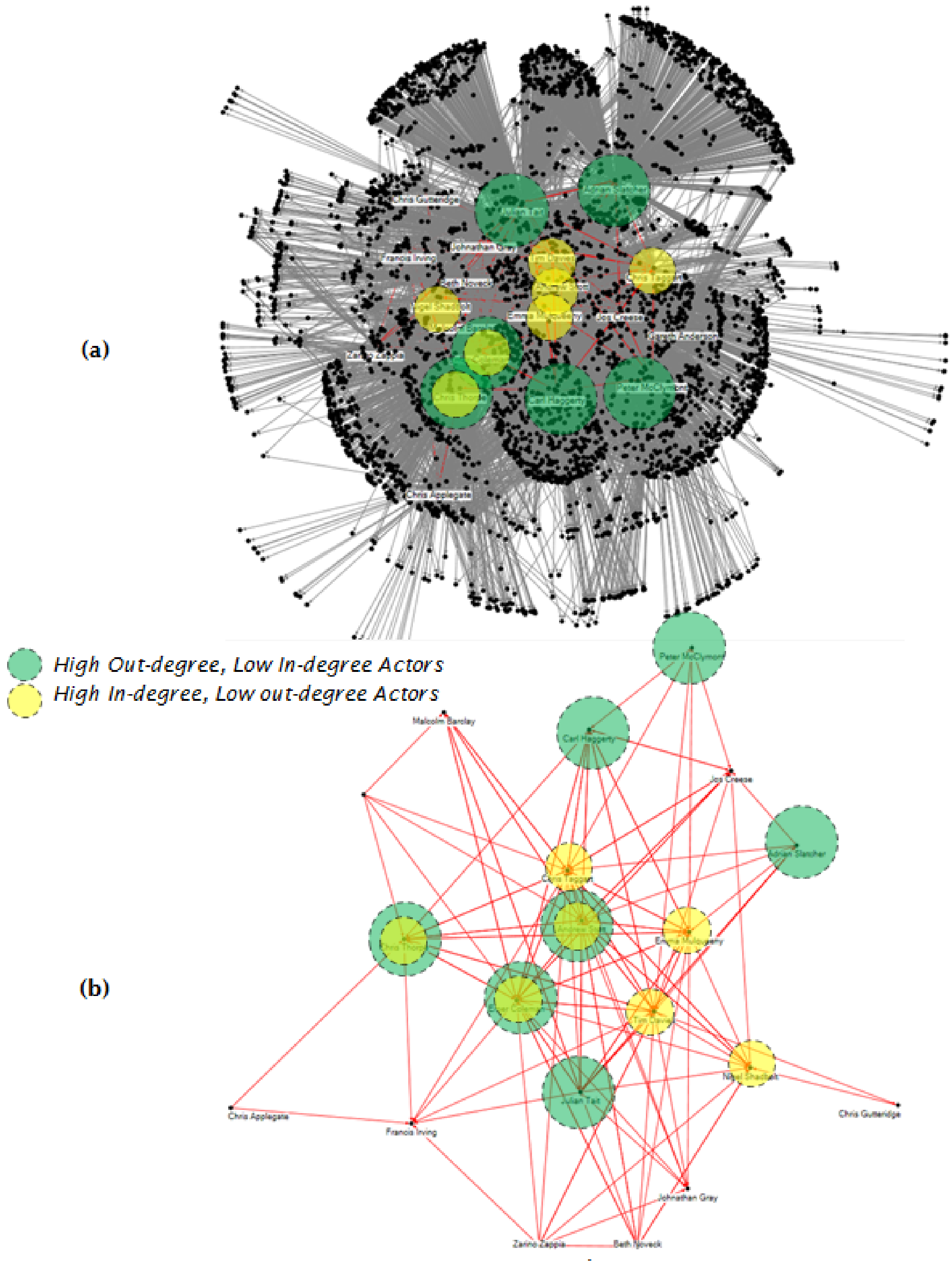


FIGURE 6.2: An illustration of the Twitter Friends network based upon data collected using the Twitter usernames of the 24 interviewees. Each node represents an actor, and the edge (directed) represents their Twitter friendship with another Twitter user. The total dataset contains 4271 nodes and 9038 edges. (a) Represents the entire Friendship network, the red edges represent the friendships between interviewees, (b) represents a filtered network, exposing just the Twitter friendships between interviewees. Overlaid are indicators for actors with an either high or low in- or out-degree

Statistically, the average in-degree is 2.3 and a max in-degree of 16, with a similar out-degree of 2.2 but much higher maximum out-degree of 1549. Putting this into context, this means that on average, each Twitter user will be friends with and are friends of (followed by) at least 2 of the members within the network, however, some users may have a slightly higher than average number of followers, and some may be friends with a large number of other users. Comparing this to the network structure in Figure 6.1, although there still are some of the interviewees that exhibit a high in-degree and low out-degree across the online and offline associations, some interviewees have become both highly-connected and associated with a larger set of actors. This change in role may be an indication that specific users, especially those with a large out-degree (connections to other users) may be actors who are central to the operation of the UK OGD community.

Whilst not deriving context from just examining the structure, the data and associated metrics provide an insight of the actor's online associations and the number of other actors involved, and also provides a means to compare how the online network structure differ from that of the physical associations extracted from their interview dialog; revealing similarities of highly connected actors which are central or focal within the community.

By examining the friendship network between interviewees both online and offline, a static, 'end-state' snapshot of the current state of the observed OGD community is shown. However, this hides the processes and different stages that the network went through in order to reach this current state. To address this, this analysis now turns to another set of collected data specifically a dataset which contains the Twitter communications surrounding a keyword (or hashtag) that was identified as a source of discussion for the UK OGD community. The following section will explore this dataset in order to provide another method to follow the digital traces of actors, and provide insight into their online communications.

6.3 Online Communications: "The community lives on Twitter"

Building upon the initial analysis of the networks that exist within the online and offline associations of the interviewees, this section will draw upon the Twitter social network dataset which contain the online communications made during the formation and evolution of the UK OGD community. Interviewees claimed that Twitter was a useful tool and observational window for the online activities of the UK OGD community, and some interviewees suggested that not only is it an incredibly important tool for the growth OGD, but without it, certain activities would have not been possible.

A specific stream of Twitter data which was a channel for UK OGD communications and activities was identified by the #opendata Twitter hashtag [Andrew Stott: 343], this communications channel includes a large number of active participants (in role and location), and is frequently used [Tariq Khokhar: 203]. Taking this into consideration, this section will provide an overview and initial analysis of the Twitter communication data collected, which contains tweets (messages) using the #opendata hashtag.

6.3.1 Static and Dynamic Analysis

The analysis will consist of two parts; the first section will examine the network structure of the communications between Twitter users communicating using the #opendata hashtag. Social network analysis (SNA) techniques have been applied to the retweet and mention communication networks in order measure the centrality and assortativity of individuals who participated within the communications. Metrics will also be used to analyse the characteristics of individuals and their role within the network (Tinati, Carr & Hall 2012), in order to expose those who were potentially influential within online network of communications and associations. The second section exploits the dynamic properties of the data and exposes the evolution of the communication networks that led to the current network state. Effectively, section two provides a way to follow the digital traces of the actors (Latour et al. 2011) by exposing this data across the longitudinal axis. Consequently, context will be added to the initial static network analysis, which has been used to help explain the processes that led to the structure of the network and actors communicating about UK OGD.

Dynamic network analysis is a relatively uncovered area of work, not only in the social sciences and sociology, but also in network science; only recently has it become an area of interest and study (Liu et al. 2011) (Bakshy et al. 2011). The aim of second section is to take the initial analysis of the snapshot of retweet and mention communications provided by section one, and explore this in terms of the longitudinal activities of the Twitter

Table 6.1: Initial Metrics from #opendata Twitter Dataset

Metric Name	Metric Value
Number of Tweets	623,061
Number of Unique Twitter Users	108,513
Number of Retweets	357,367
Number of Mentions	128,986

users, considering the varying network structures across the timespan of tweets made and collected. As described in the methodology, it is possible to explore the dynamics of Twitter conversations using computational software approaches (Tinati, Carr & Hall 2012), which involves techniques which form an indiscrete time series mapping of the #opendata communications. Based on this, snapshots of the network at chosen time slices can be examined, providing both visual and contextual evidence for analysis. Given a specific date (granular to the minute), or range of dates, communications can be examined, including the content that led to the communications.

6.3.2 #opendata Twitter Dataset: an Introduction to the Data

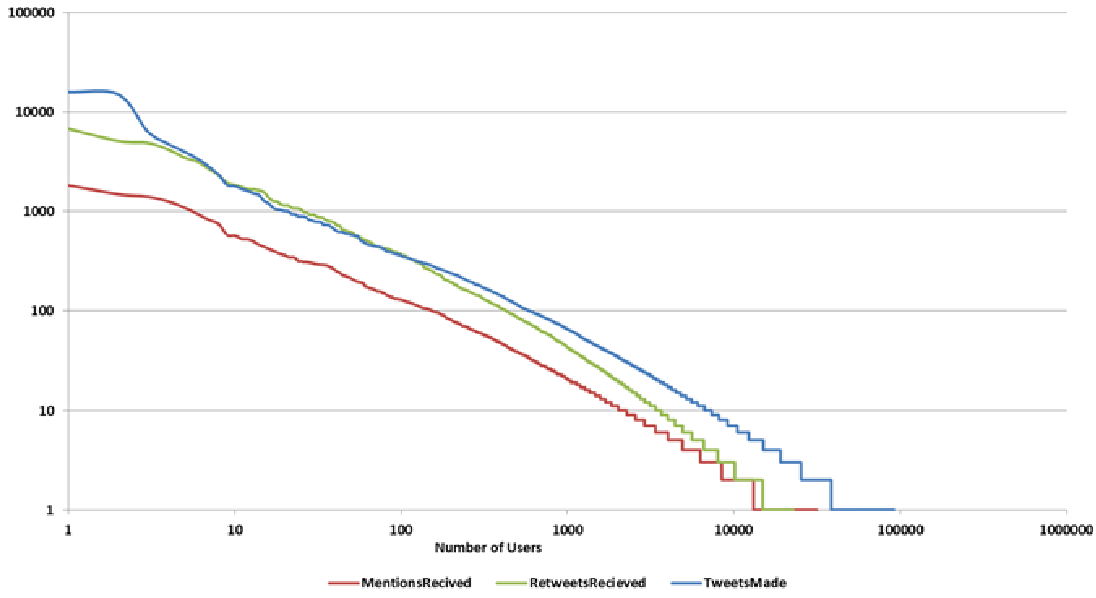
The Twitter #opendata dataset collected consists of a chronological list of tweets (Twitter messages), containing 623,061 messages (or tweets) that use the hashtag #opendata, spanning from 1st January 2009 to 31st December 2012. A smaller collection of Tweets between January 2006 and December 2008 which contained 1925 tweets were also harvested, however these have been excluded from the analysis in this chapter as the data was not consistent due the architectural changes of the Twitter platform and API during this time period. The method used to collect this data was a mixture of pre-archived tweets collected from an external source and a custom designed Twitter data harvester. The harvest collected as much data as possible regarding each individual tweet, which included the unique identifier of the tweet, the tweet message, the timestamp, and the unique screenname of the user that authored the Tweet. A secondary process of cleaning the data was performed which made sure all collected tweets were formatted correctly and contained the required fields to enable the analytical toolkit to process the data.

Table 6.1 shows a number of dataset metrics elicited from the data cleaning and sanitising process. These metrics, which are a simple measure on the size and content of the collected tweets, provide an initial insight of the magnitude of the data collected, and the proportion between communications methods (retweets, mentions and plan tweets). Extracted from the communications were a set of 108,513 unique Twitter users, contributing to both the mention and retweet communications, which represent 128,986 and 357,367 tweets, respectively.

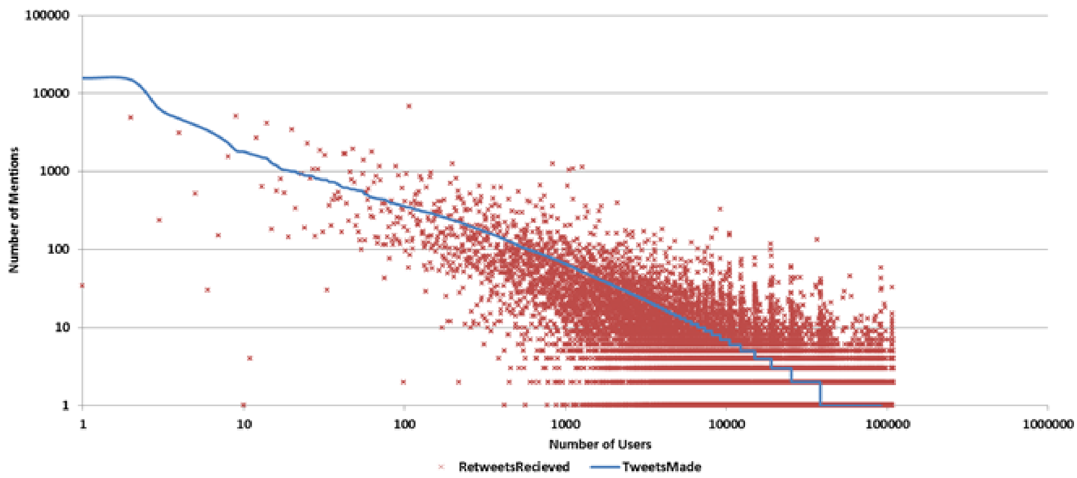
During the process of sanitising the data, it was also possible to examine - at a high level - the data in terms of the most active Twitter users in terms of the number tweets, retweets and mentions, and also users that had received the highest number of retweets and mentions. The analysis also examined the content embedded within the collection of tweets, which included examining the most widely used hashtags (excluding #opendata), of which #opengovuk, #localgov, and #datagovuk were within the top 10 used keywords or hashtags. The content of the tweets were also examined for embedded URIs (e.g. hyperlinks such as http:// or https://) within the tweets, of which there were 451,021 (some tweets contained more than one URI).

As Figure 6.3(a) illustrates the distribution of tweets made, retweets, and mentions exhibits properties of preferential attachment often associated with large-scale networks, including online social networks (Capocci et al. 2006), and the distribution of links between documents within the Web graph (Barabasi & Albert 1999).

Interpreting the distribution shown in Figure 6.3(a) and Figure 6.3.b in terms of what this represents within the #opendata Twitter dataset, it suggests that only a small fraction of the users are either publishing a large volume of tweets, and similarly, only a small fraction of users are receiving a large volume of retweets or mentions. Instead, the majority of users have only contributed a small number of tweets, or received a small number of retweets or mentions. By comparing the distribution of user tweets with the number of retweets received (Figure 6.3.b), applying a linear regression to the number of retweets a user makes and the number of tweets made a r^2 value of 0.66 was found (based upon a subset of the identified users, filtered by users who were frequently participating in communications). To some degree, it is possible to conclude that the more active users in terms of tweets made, the more likely they would be retweeted. Findings were also similar between mention and tweets, with an r^2 value of 0.51, suggesting that those that tweet more tend to receive more mentions. This reflects the earlier findings that certain users have a greater number of offline and online associations (friends) within



(a)



(b)

Figure 6.3: Graphically illustrating the (a) log-log distribution of Tweets, Retweets (received), and Mentions (received) between the different Twitter users in the #opendata dataset. For each unique Twitter user, their number of Tweets, retweets received and mentions received have been counted, and then in descending numerical order. (b) log-log distribution of Tweets made by users in comparison to number of retweets received

the OGD community, and is an indicator that there are similar structures in terms of the roles of actors within the online communication network. Some actors are more popular than others, and those that are tend to be able to share their message to many individuals.

In order to examine this further, the next two section will construct an end-state snapshot of the mention and retweet networks, representing the entirely of the communications made during the time period of data collected.

6.3.2.1 #opendata: Mention Network

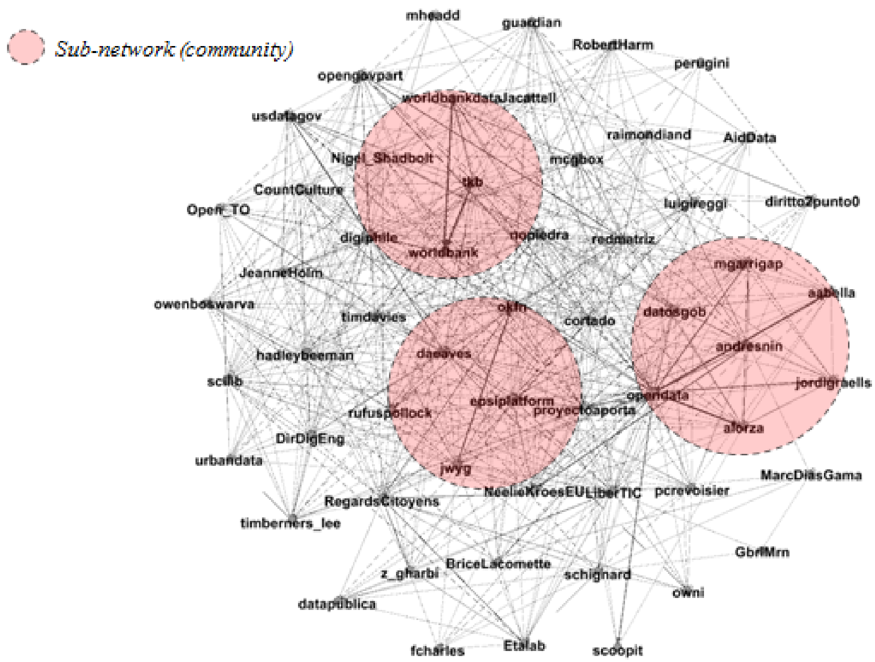


FIGURE 6.4: An illustration of the Mention network graph extracted from the #opendata dataset of communications between January 2009 and December 2012. For clarity, the graph has been restricted to show only nodes with an in-degree of over 400, and the edges are weighted based upon number of mentions between particular Twitter users. The thicker the edge line, the more communications made between the two. The red overlay provides an illustrative example of the sub-networks that have formed

As shown in Table 6.1, during the time period of data collected, there were 128,986 tweets made which contained one or more mentions within them, illustrated in Figure 6.4. As an example, below is a tweet from the dataset which contains a mention of another Twitter user; extracting this information along with the author of the tweet makes it possible to construct the mention communication:

Well done RT @eudata Enschede first Dutch city with a #opendata publishing platform launched today in beta at #od053 [Twitter data stream, 'watty62', 2nd November 2012]

Table 6.2: Mention Network Statistics of #opendata Twitter Dataset

Metric Name	**Metric Value**
Number of Nodes (Twitter Users)	42,745
Number of Unique Edges (Retweets)	89,934
Average Degree	2.1
Average Weighted Degree	3.4
Maximum In-Degree	639
Maximum Out-Degree	191
Weakly Connected Components	1915
Strongly Connected Components	36430
Modularity (Randomised)	0.72
Communities	2034

Figure 6.4 provides a visual representation of the mention network filtered by users with have an in-degree of 400 or greater (this value is chosen based upon the top percentile of users and mentions), it provides an illustration of the structure that exists within the extracted network. Similar to the association networks there are users that have a higher proportion of mentions. These form part of their own sub-networks of mention communications with other Twitter users, which can be identified by the thicker weighted edges between users.

Examining this further, based upon the statistics derived from the #opendata mention network, Table 6.2 provides additional metrics of the networks structure and properties. Comparing the total number of edges (128,986) with the number of unique edges (89,934), it shows that between particular users, multiple mentions are being made (thicker edge lines), which may be an indication of closely related users, or those that are part of their own sub-network. Indeed, by examining the number of strongly connected components (users who form a closed network, where all users within the network communicate with each other), a number of sub-networks were exposed.

Furthermore the number weakly connected components which represent the users who are not part of a reciprocal network of communication between each other (i.e. user A mentions user B and C, but user A is only mentioned by user B, not C), is significantly lower than the number of strongly connected components and is an indication of dialogs between users, which can be either single messages or a long interchange of messages. In addition to this, the relatively high modularity metric indicates that there are a number of tightly connected clusters (or communities), which are connected via intermediaries, something similar to what was illustrated by the analysis of the Interview association network.

Interrogating the data further, Figure 6.5 illustrates an extracted mention network from the #opendata dataset, using the interviewees Twitter username as a seed. The extracted network has similarities with the friendship network created in the previous section. By extracting the subset of interviewee mention network, it was possible to obtain a network of 1190 unique Twitter users, who made 3380 mentions (1191 unique) messages.

Similar to the structure of the friendship interviewee network, the interviewees represent authorities for different groups of Twitter users; reflected in the lower average degree value of 1.5. However filtering (restricting) the network by users who were the interviewees a higher in-degree and out-degree of 10.1 and 8.5 was found, which suggests that they were communicating and conversing much more than the 'average' Twitter user. Despite not examining the details of these communications, the metrics provide an indicator that these strongly connected sub-networks of mention communications may relate to the interviewees activities.

6.3.2.2 #opendata: Retweet Network

The retweet network extracted from the Twitter #opendata dataset provides another method to interrogate the communications of the actors involved with the OGD community, offering additional insight into the structure of the network and the roles that particular actors play. Retweets are one of the core Twitter features; it enables messages written by a particular user to be shared across another network of users. For instance, if user A publishes a tweet, it is possible for another user (B) to retweet it; if user B has a set of followers that are different to user A, then the message (along with the name of the original user) is shared to a larger network of twitter users. Similar to a mention tweet, a retweet contains the '@username' string, however, it is distinguishable from a mention as it is preceded by a 'RT' or 'Via' string:

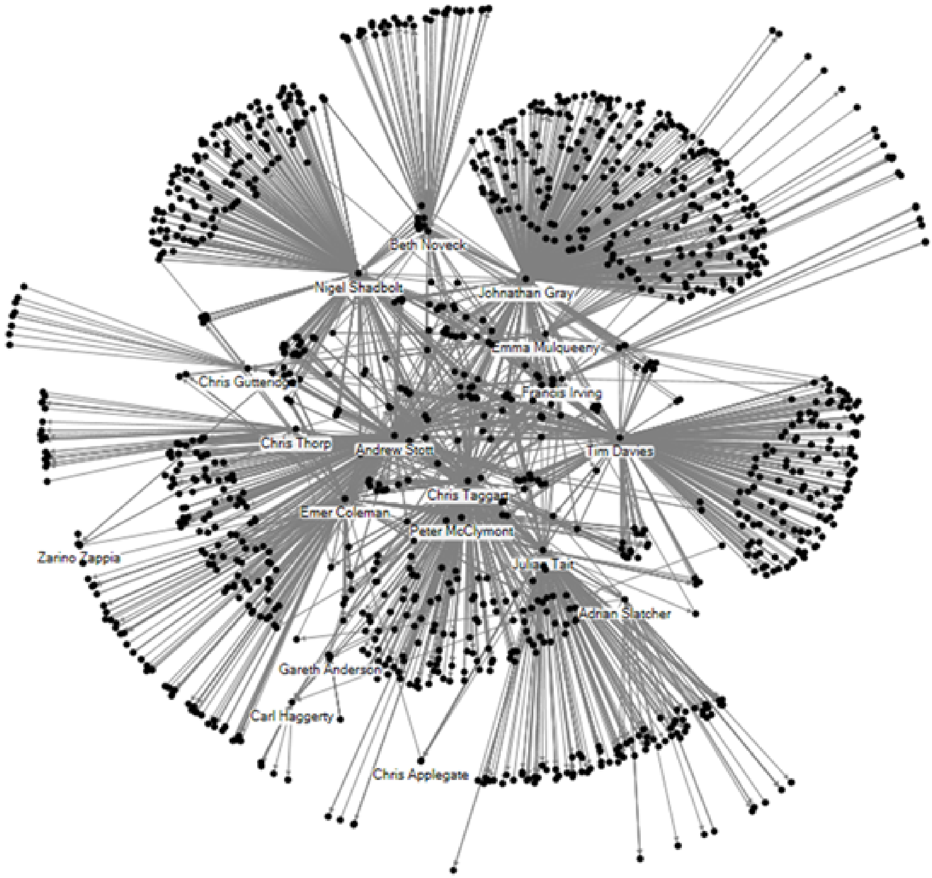


Figure 6.5: An illustration of the Interviewee Mention network graph extracted from the #opendata dataset of communications between January 2009 and December 2012. A subset of mentions (3380) that contained any of the Interviewees was extracted from #opendata dataset in order to achieve this. Nodes that correspond to interviewees have been labelled

RT @hadleybeeman: Met Office to publish more #opendata under the OGL (from Transparency Board mtg minutes July) http://t.co/Lt8puAoW [Twitter Data Stream, 'Bolster', 6th October 2011]

Using the notation of a retweet, a retweet network was extracted from the #opendata communication dataset, and as Figure 6.6 shows, it exhibits structural similarities with the mention network. The retweet network shown in Figure 6.6 is an illustration of the top percentile of retweeted users, and as before there exists certain sub-networks between these Twitter users.

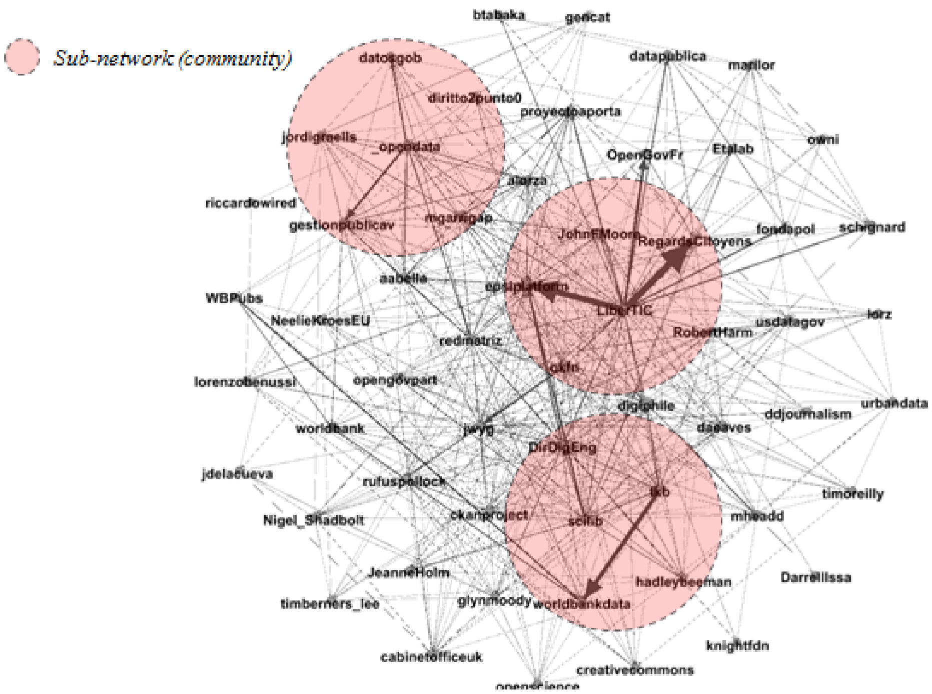


FIGURE 6.6: An illustration of the Retweet network graph extracted from the #opendata dataset of communications between January 2009 and December 2012. For clarity, the graph has been restricted to show only nodes with an in-degree of over 400, and the edges are weighted based upon number of mentions between particular Twitter users. The thicker the edge line, the more communications made between the two. The red overlay provides an illustrative example of the sub-networks that have formed

The 'mention' mechanism facilitates discussion between users (there may be more than one mention in a tweet), whereas the retweet mechanism enables users to share information, which may be an indication of valuable content. Taking this into consideration, the sub-networks visible in Figure 6.6 are potentially indicators for important activity within the UK OGD community, and also may provide some insight into based on two factors, the number of retweets that they have obtained, and the sub-network of Twitter users that they associate with.

Examining the metrics in Table 6.2, the number of extracted retweets (357,367) in comparison to the number of unique edges (181,167) is an indication that Twitter users tend to gain one or more retweets from the same user. Furthermore, comparing the proportion of unique retweet and mention edges to the total number of retweet and mention edges, 51% and 70%, respectively, it appears that retweets promote a network of homogeneous activity. Users tend to retweet the same user multiple times, which is

Table 6.3: Retweet Network Statistics of #opendata Twitter Dataset

Metric Name	**Metric Value**
Number of Nodes (Twitter Users)	80,143
Number of Unique Edges (Retweets)	181,167
Average Degree	2.2
Average Weighted Degree	3.6
Maximum In-Degree	3905
Maximum Out-Degree	260
Weakly Connected Components	1457
Strongly Connected Components	69,257
Modularity (Randomised)	0.65
Communities	1522

not only reflected in the higher average degree and weighted degree in comparison to the mention network, but the number of strongly connected components.

However, comparing the number of communities identified in the retweet network to the mention network and taking into consideration that there are significantly more retweets than mentions, then it would appear that as the modularity of the networks are similar (0.65 compared to 0.71), the lower number of communities identified in the retweet network (1522) compared to the mention network (2034) suggest that the communities formed are much larger in size. This may be a result of the simplicity of a retweet, compared to that of a mention, which is requires the user to commit to direct communications with other users.

Supporting the earlier findings Figure 6.3, the range between the maximum in-degree and maximum out-degree in comparison to the average degree is significantly different; putting this into context, the in-degree, or the number of retweets a Twitter user obtains may be an indicator of an influential user, or potentially influential content. The same applies for the maximum out-degree, which is an indicator of a user that are sharing multiple streams of information act as aggregator and curator of content within the #opendata twitter communication network, effectively becoming a 'hub' for #opendata communications.

By extracting the sub-set of retweets by using the Twitter usernames of the interviewees as seeds, a network is exposed containing 2807 Twitter users, 7847 retweets, of which 2562 were unique. As Figure 6.7 shows, the retweet networks structure contains different hubs or sub-communities similar to the mention network extracted. This structure may reflect the offline associations and activities with a particular set of actors within the OGD community. In addition to this, comparing the average in-degree (11.2) and out-degree (3.4) of the interviewees to the complete #opendata communications retweet network, the actors within the extracted network are more closely connected, yet there is still a distribution of retweets, with a maximum in-degree of 1063, and a maximum out-degree of 324. This suggests that even between the interviewees, there are particular individuals who receive a higher volume of retweets than others, possibly an indicator of their presence and role online as well as offline.

From this initial analysis of the snapshots of communications it has been possible to construct a representation of how the Twitter communications of the OGD community, and hidden within the static network lies a complex network of sub-networks where specific actors have different communication characteristics. For instance, actors who obtain a high proportion of retweets are potentially more 'visible' within the network of communications. By bridging the gap between the online communications and the associations mentioned by interviewees, the analysis demonstrated a community structure across both the mention and retweet network, and how this shares similarities between the identified offline network of associations. However, whilst examining a static snapshot of the network is useful for an overall understanding of the community's structure at a single end-state, to examine how the network is growing, a dynamic analysis is required.

6.3.3 #opendata: Dynamic Analysis

This section will provide a dynamic analysis of the Twitter #opendata dataset, examining both the retweet and mention streams of communications. Whilst the previous sections provided an overview of the end 'snapshot' of the network structure that formed from the communications in the #opendata dataset, the dynamic analysis in this section will explore how these structures have evolved overtime, in effect, providing multiple

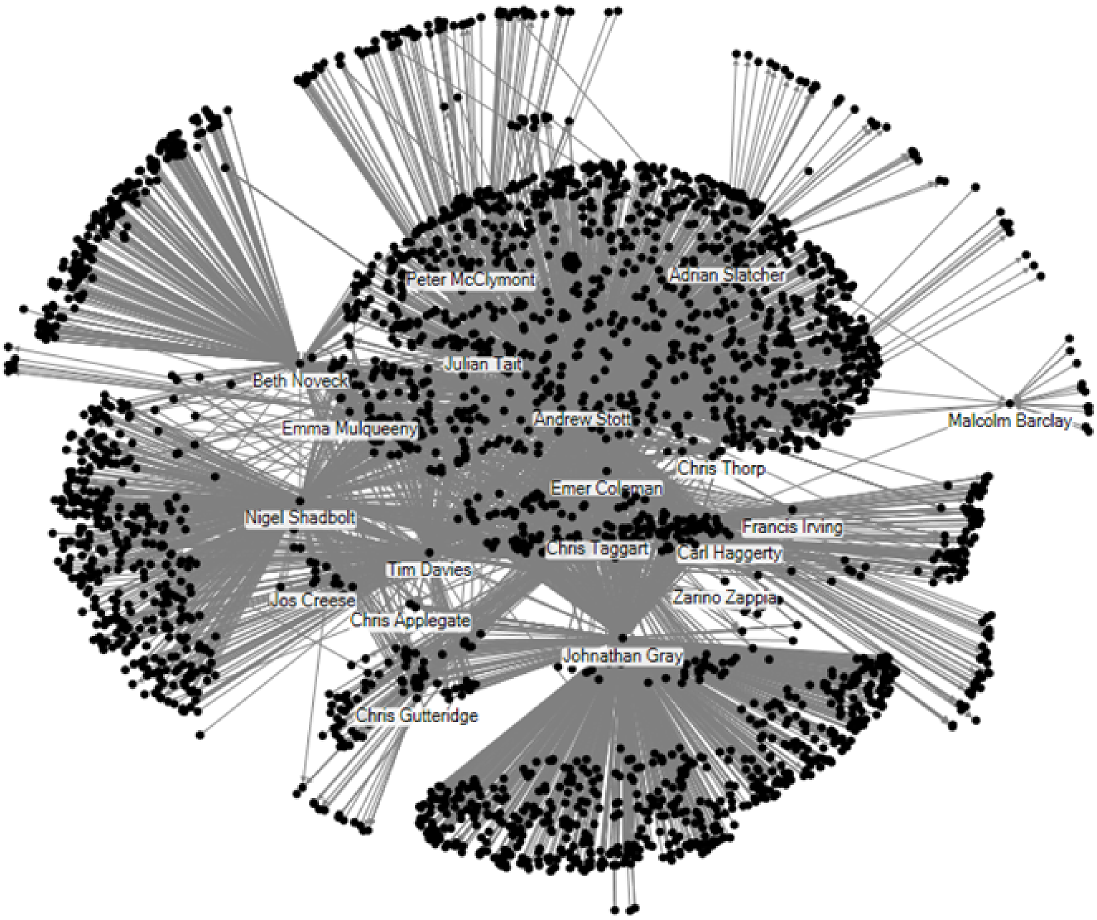


FIGURE 6.7: An illustration of the Interviewee Retweet network graph extracted from the #opendata dataset of communications between January 2009 and December 2012. A subset of retweets (7847) that contained any of the Interviewees was extracted from #opendata dataset in order to achieve this. Nodes that correspond to interviewees have been labelled

snapshots in order to reveal particular events or anomalies in activity that may be important to investigate (or simply observe).

As an initial way into exploring the dynamics of the communications within the #opendata dataset, a time series will be constructed from the total tweets made and the sub-set of retweets and mentions that contribute towards this. As Figure 6.8 shows, the volume and frequency of tweets between January 2009 and December 2012 increased steadily, and the first glance of the data indicates that the volume of daily retweets in comparison to the mentions is always significantly higher, and at no point within the 4 year timeframe does the volume of mentions exceed that of retweets. Furthermore, and as Figure 6.8(a) shows, during 2009, the volume of activity was significantly lower than that 2010 to 2012, and although no in-depth analysis of the content of tweets and communications has been performed, the increase of activity from 2010 onwards might be an indication of the initial formation of the UK OGD community, or more specifically their online activities.

By examining Figure 6.8, it also is clear that activity within the network is not constant, and that during certain time periods, more activity (tweets, retweets, and mentions) occur. This may provide an indicator that at these specific points in time that certain events, milestones or developments in the UK OGD community occurred, offering another way to interrogate and triangulate different data sources in order to construct an understanding of processes that enabled the community to translate.

Building upon this initial insight into the patterns of communications, the mention and retweet network will be examined in further detail, exploring at a more granular level the networks of communications that reflect the metrics values identified. An approach to looking at the dynamic properties of the communication is to apply metrics that are used in traditional end-state network analysis, and combine it with a windowing technique; that being, taking multiple snapshots of the network over the timeline of communications and 'stitching' them together to form a dynamic time series of the communications. This provides a way to examine the network from another perspective, whilst the static analysis provided an understanding of the end-state of the communications network; the dynamic analysis provides its structural evolution over time.

The dynamic analysis will use a 24-hour windowing and tick sample over the entire dataset timeline, which simply means that the dynamic analysis will stitch together the day-to-day snapshots of the communications and metrics.

6.3.3.1 Dynamic Analysis: Mention Communications

By applying time-based sampling techniques (Ahmed & Berchmans 2010) to the extracted set of mentions from the #opendata Twitter dataset, it is possible to use a set of metrics to transform the static network of mentions shown in Figure 6.4 and Figure 6.5, into a time series communication pattern. Whilst Figure 6.8 provided a basis for the frequency of mentions between 2009 and 2012, it is possible to extend this and examine the network in terms of additional metrics, including the dynamic number of

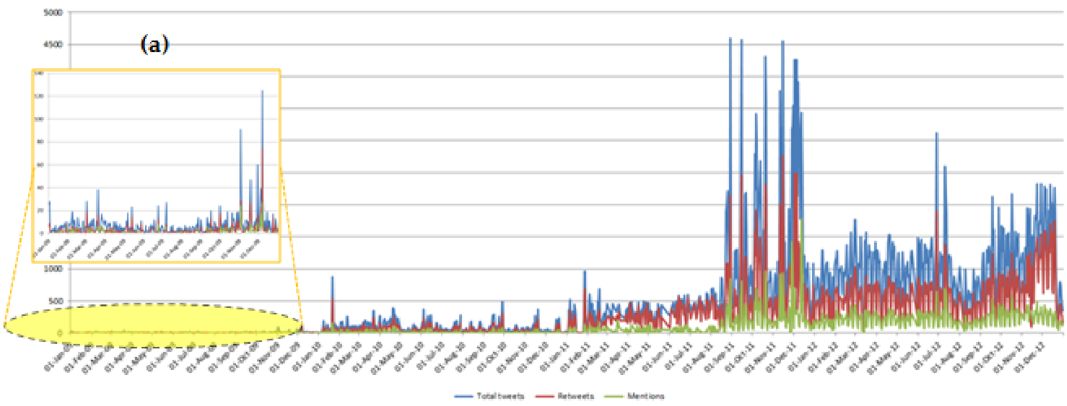


Figure 6.8: The timeseries of tweets (blue), retweets (red) and mentions (green) extracted from the #opendata Twitter dataset, between January 2009 and December 2012. Values have been aggregated based on a daily window sampling size. The volume of daily total tweets may be larger than the combined total of retweets and mentions as not all tweets extracted were part of a form of communications. (a) has been provided to clarify for the volume of tweets, retweets and mentions during January and December 2009

nodes (Twitter users), edges, average degree, and also the average clustering coefficient; these have been calculated and shown in Figure 6.9.

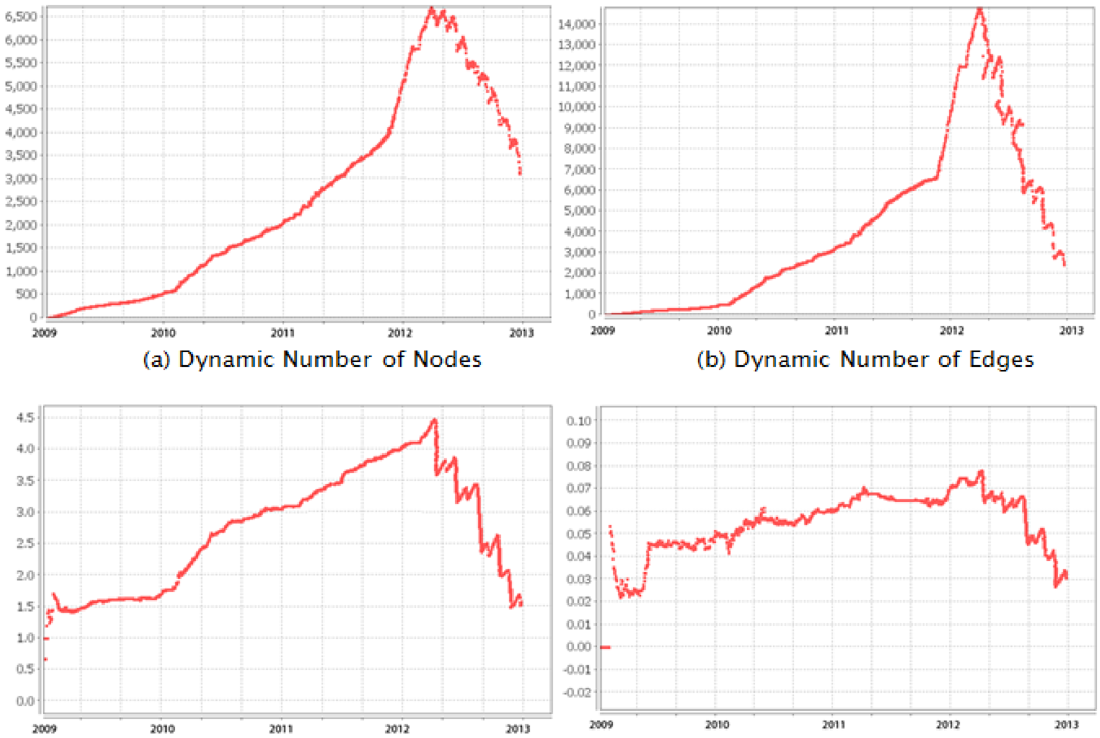
As Figure 6.9(a) shows, the numbers of users have grown considerably since the beginning of the dataset in January 2009; and although the volume or growth before this date is unknown, using the activity show in Figure 6.8 as an indication of activity, then it seems that 2009 may have been the early stages of engaging with Twitter as a communications tool. At the start of 2010 (January), there appears to be a sudden increase in the number of users, which increased at a steady rate until the start of 2012. The number of users contributing during the first half of 2012 increased dramatically, reaching its peak around March 2012, and then has declined at the same rate as it increased by.

Bearing this in mind, considering the dynamic growth of edges shown in Figure 6.9(b), similar to the growth in the number of nodes, the growth in the number of mentions between 2009 and 2010 was not as rapid as between 2010 and 2012, which increased at a much faster pace; at first glance an indication that at the start of 2010 is when the activity and translation within the OGD community began to occur. It appears that although the first 3 months of 2012 witnessed a dramatic rise in the number of mention communications (along with an increase in the number of Twitter users participating in it), this was only sustained for a relatively short period of time, and subsequently dropped at an equally dramatic rate; perhaps the sudden increase of new users was responsible for both the increase and decline in activity during 2012.

Keeping this growth (both nodes and edges) in mind, examining Figure 6.9(c) reveals some interesting properties of the communication and user network structure. The dynamic average degree is a representation of how well 'connected' the users are within the network, i.e. how many communications is a single user making with other users. During 2010 until the start of 2011 the degree rises from 1.5 to just below 3.0. However, between August 2011 and February 2012 shows a significant rise in the networks overall degree, which is an indicator that the network became much more connected in terms of associations between users; users were talking to many other users, and because the number of users also increased during this time, it may indicate that these communications were including new members of the UK OGD community. This is also reflected in Figure 6.9(d), which a dynamic representation of the clustering coefficient, a metric which measures the degree to which nodes in a network tend to cluster together, and in the context of Twitter communications, would represent different clusters of users communicating with each other, connected together by a weak tie, or user that know users in multiple clusters of users. This is similar to the Communities and Modularity metrics, which represent the sub-communities that formed within the conversation networks and the use of intermediaries between sub-networks (i.e. the users who join the networks together). As Figure 6.9(d) illustrates, the clustering coefficient between mid-2009 and March 2011 steadily increased from 0.045 to 0.7, and peaked at just below 0.8 before decreasing during the later months of 2012. Despite this, the average clustering coefficient shown supports the value of the communities metric shown in Table 6.1, and also provides another perspective on the rise in average degree shown in Figure 6.9(c); although the 'connectivity' between users have increased and more users are talking to other users, this is potentially happening in clusters or hubs of individuals, interconnected together by the 'weak links'.

6.3.3.2 Dynamic Analysis: Retweet Communications

Using the same techniques as the previous section, the four dynamic metrics will now be considered in terms of the communications based on retweets network extracted from



(a) Dynamic Number of Nodes

(b) Dynamic Number of Edges

(c) Dynamic Average Degree

(d) Dynamic Average Clustering Coefficient

FIGURE 6.9: Time-based dynamic metrics relating to the #opendata Twitter mention communications between January 2009 and December 2012. The metrics are calculated using a Using a 24 Hour tick and window sampling rate

the #opendata dataset. As Figure 6.10 shows, the profile of the growth and decay for the different metrics for the retweet network are similar to that of the mention network.

The similar increase in the number of nodes and edges between the mention and retweet network are not surprising, as this reflects the overall activity of the network as shown in Figure 6.8, when more users are sharing ideas and creating tweets, it is expected that users will communicate between each other to discuss this. However, although the profiles of the metrics are similar what is apparent is the volume of retweet nodes and edges is an order of magnitude larger than that of the mention network, reaching a maximum value during 2012 of 11,000 users and 30,000 tweets in comparison with the mention network which peaked at a maximum of 7,000 users and 15,000 tweets. An explanation of this could be related to the simplicity of the retweet function; it requires very little effort to retweet someone in comparison to mentioning someone, which requires the user to write a message with some content. Despite this, the number of individuals participating is an indicator that they have some interest in the community and also are

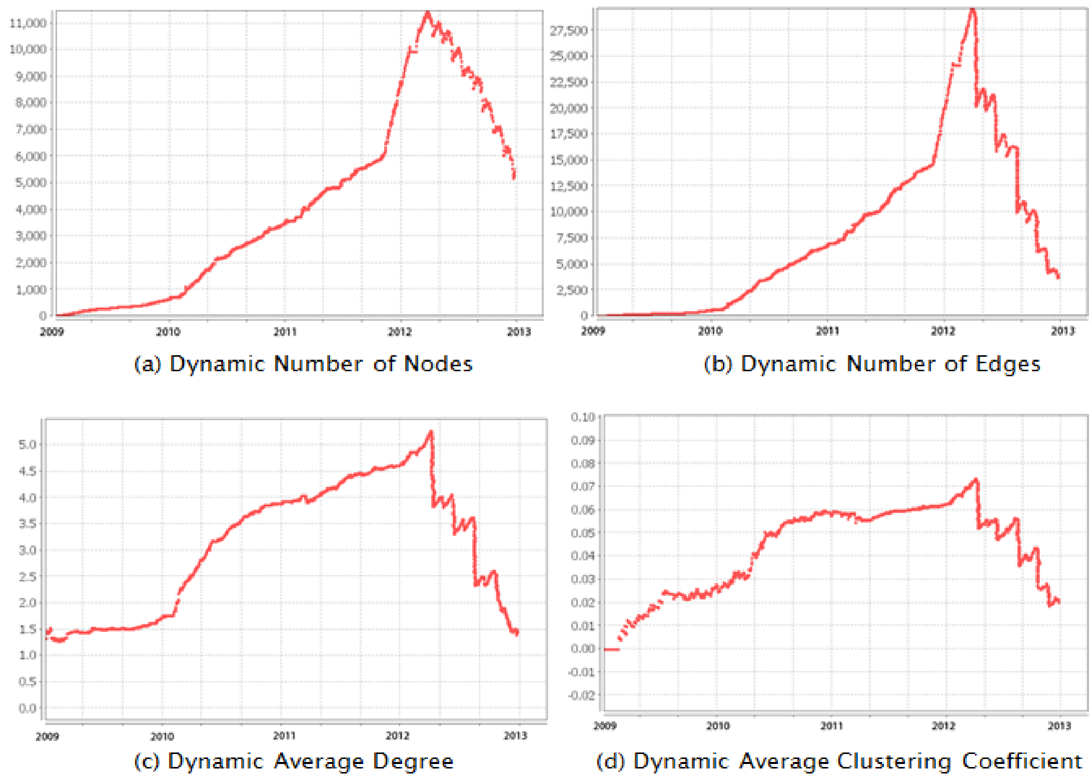


FIGURE 6.10: Time-based Dynamic metrics relating to the #opendata Twitter retweet communications between January 2009 and December 2012. The metrics are calculated using a Using a 24 Hour tick and window sampling rate

keen to share information and content with their audience.

Although the dynamic average degree between 2009 and 2011 of the mention and retweet network were similar in profile and values, the peak average degree reached in 2012 is much higher within the retweet network, and may be an indicator of the sharing of content with a wider audience, especially as the number of individuals participating increased during this frame. Examining Figure 6.10.d, the average clustering coefficient of retweet network, it was not until January 2010 that it reached the same value as the mention network, which may reflect the lack of valuable content or communications being published, or an indication of the way Twitter was engaged with before January 2010 (retweeting was not officially implemented by Twitter until August 2009).

In reflection to the findings, whilst retweets are not an indicator that individuals are communicating with each other, the volume of retweets can be a measure of valuable content or information, which can include external URLs or some other media (video, image). What is interesting is the social importance of retweeting; although an area of current debate, literature (Boyd et al. 2010) (Macskassy & Michelson 2011) suggests that the act of 'retweeting', which is nothing more than sharing someone else's content' is an indicator of influence within a social network, i.e. the more retweets an individual gains, the more important they are within a network. Taking this into consideration, it may be that the retweet network did not reach the same average clustering coefficient value as the mention network until mid-2010 as the community using the #opendata communications channel had not established their presence within the OGD community (online or offline). However, by the start of 2011, there had been sufficient activity to enable communities to form and for individuals who had established their offline and online profile to become authorities of valuable information, which effectively causes the clustering coefficient to increase.

6.4 A First Insight into a Growing Web Activity

The analysis of the association's network between interviewees and the wider network of actors involved in UK OGD provided a starting point to understand the possible emergence and structure of the UK OGD community based upon the roles and communications of different actors. Using this association network as a reference point, the online associations of the interviewees were explored which exposed their associations with a larger online network of individuals. A network similar in structure to the offline interviewee association network was discovered. The analysis also exposed how the structural positions of particular interviewees were different in the association and communication networks, and these positions could be associated with different roles in facilitating the communications.

Building upon this, examining the #opendata Twitter dataset helped bridge the gap between the offline associations and online communications and exposed network of dialog between individuals through the use of the mention and retweet Twitter mechanism. By examining the online communications between the interviewees illustrated the similarities between their role within the offline association network and their online communications mention and retweet network. To some extent, shifting between the offline and online networks demonstrated the blurred boundaries between the physical and digital traces of actors (Latour et al. 2011). Ultimately, this is made possible by being able to explore a mix of data sources and analytical methods that can take advantage of the external linkability within and between datasets.

This initial inquiry has used computer science and network science techniques to analyse the data, and whilst this analysis has provided an initial overview and insight into the online and offline associations and the growth and structure of the online communications, the analysis has yet to expose the context and underlying social and technical processes that enabled a complex, heterogeneous and multi-faceted Web activity to emerge. In order to unpack this, the analysis will now shift focus, examining the translation process of the UK OGD Web activity more closely. Framed by the theoretical position of ANT and the process of translation which is a concept to help explain the formation, configuration and functioning of a network (Callon 1986*a*), the individual activities and interactions between UK OGD actors will be examined, and by combining this with the initial findings, the analysis will expose the socio-technical processes that facilitated the translation of the Web activity.

6.5 Towards The Process of Translation

Framed by the process of translation, the next chapter will work with the longitudinal properties of the case data to expose the series of socio-technical processes that enabled the UK OGD Web activity to emerge and translate. The initial findings of growth and structural changes within this chapter are an indication of the dynamic and fluid growth of UK OGD. Based upon these initial findings, the next chapter will consider the underlying processes that result in the observed fluctuations.

Chapter 7

The Emergence of UK Open Government Data

7.1 Analysing the Translation of a Web Activity

This chapter will provide an analysis of the socio-technical processes that contributed to the emergence of the UK Open Government Data Web activity. The previous chapter provided an introduction to the UK OGD Community, along with an initial insight into the collected data sources. By using quantitative techniques, the data was unpacked, exposing the actors, their associations, and online communications. The initial insight found that the communication and interactions between actors within the UK OGD community form a messy, multi-network of actors and communications. The emergence of the community was found to be non-linear in terms of rate of activity, and also exposed the number of communities connected together through weak ties (in the online communications). By considering this through the lens of ANT, what becomes apparent is how the emergence of OGD can be seen as part of a complex process of social engagement and technological development; made possible by the interactions of humans, and shaped by specific technologies.

In order to expose the underlying social and technical processes that have been responsible for the growth and structures observed, the investigation turns to asking how this complex, and to some extent, 'messy' network of actors and activities have emerged to become a coherent and functioning Web activity. The analysis will draw upon the findings in the previous chapter, alongside the coded interview data and observations of the field, and a timeline of events constructed in Appendix C. By 'following the actors' (Latour 2005) and tracing their activities and associations with other actors, the translation of the UK OGD community will be examined.

7.2 The Process of Problematising Actor-Networks

In order for networks to assemble and begin the translation process, a stage of problematisation will first occur; a stage where actors identify an issue or problem, and try to elicit a suitable set of additional actors and actants to achieve a goal (Latour 2005). The analysis of UK OGD translation begins by examining how problematisation was reached and what socio-technical processes and network objects were part of this.

7.2.1 The Problematisation of Public Sector Information

Drawing upon the synthesis of the interviews and documents collected, the emergence of the OGD network can be characterised by the interplay of three actor-networks. The first actor-network consists of the academic researchers involved in exploring knowledge representation and ways to apply it to the Public Sector Information (PSI) domain. In turn the PSI actor-network draws in Government, and the policy agenda promoting publication of public sector data. The network of actors that represented the academic community during this stage was working towards achieving the goals aligned with the ideas and interests with the PSI community. Finally, the Open Knowledge Foundation (OKFN) actor-network draws upon interested actors involved in lobbying and providing Openness across numerous landscapes. To begin, the analysis will first focus on the activities of the academics and PSI network.

Prior to the problematisation of Open Government, the Academics were involved with an agenda concerned with knowledge representation and using alternative data formatting standards and technologies, notably the Semantic Web and Linked Data technologies. This agenda required using and developing technologies, and the successful outputs of their research worked as an incentive to entice actors within the PSI network to become interested in the potential of Linked Data technologies:

> *"They were looking at whether or not he [PSI Actor] had heard of the Semantic Web it offered a way to integrate different information sets. [...] We showed that you could use these standards and that it was really a good level of abstraction across lots of different datasets and understanding how you could use some common tools to link all this data together. That was reported to parliament [...] that was great, it got some interest."* [Nigel Shadbolt: 17]

The development of the knowledge representation and Linked Data technologies became a driver for the academics to gain interests of the PSI actor-network; and the chances of gaining their support was helped by the common actors that had interests and roles in both networks. The academic actors involved in previous activities of 'knowledge representation' had previous relationships with those within the Government, thus making contact and communications simpler [Nigel Shadbolt: 1].

7.2.1.1 Using Immutable Mobiles for Problematisation

'Immutable mobiles' play an important role in the formation of a heterogeneous network, they are entities or objects within a network which move around but hold their shape, both physically and geographically, but also in a functional sense (Latour 2005). These immutable mobiles can become a device to shape the outcomes and relations of actors that they become tangled in (Law & Singleton 2005). Immutable mobiles are not only objects within a network that contribute towards the network goals, but they can be converted into inscriptions or devices of interest(Tatnall 1999). By adopting this concept, it is possible to examine how entities from other networks (which may be out of scope from the analysis) can carry through and situate themselves within a new translating network

The existence of network artefacts that were carried over from previous activities between the Academic and Government actor-networks helped shape their forming relationship. These existing artefacts become part of the goals and interests of the actors in their previous networks, which are carried through to the newly emerging networks between the UK government (PSI) and academics.

Government Acts and written laws function as immutable mobiles. Dating back to the 1980's, Acts such as the 1985 Government Act (HM Government 1985) defined the public's access to local authority meetings, reports, documents, and gave duties to local authorities to publish certain information. This was used as a foundation for later research to adopt and apply to Public Sector Information (Office of Public Sector Information 2007). These were mandatory requirements from central government, consequently the activities of the PSI network had to conform to them. This aligned with the interests of the academics as it provided a means for their knowledge representation technologies to be used.

> *"At that time, the office of public sector information's main responsibility was looking after a thing called the information assets register and managing crown copyright. So we got together a bunch of local councils, authorities, people like ONS, and this is written up in a paper an ISWC paper that we wrote on this experience."* [Nigel Shadbolt: 12]

The PSI network had already established an interest towards the public access of public sector data, requiring actors to align to the activities of managing and curating information, inscribed by immutable mobiles from previous networks of activity. Effectively, the PSI network had already undergone translation, previously problematised and driven by government acts and legislation. Although this is out of the scope of the investigation, it is important to recognise that the activities of the PSI network were occurring prior to the problematisation of OGD, which has been part of a much larger set of incentives and drivers:

> *"money and markets have always been a big motivating factor. The economic value of public sector information. If you look at the background to the European Public Sector Information Directive, there was a lot of interest in economic activity around public data [...] if you look at speeches and official communications from the US, the EU, or from numerous other countries, they often refer to the economy and potential job creation, as well as transparency, accountability, new digital services for citizens and so on."* [Jonathan Gray: 126:130]

The government acts and legislation previously described enabled the academic and PSI network to work towards a common task. They acted as a driver to develop their relationship which in turn produced new social and technical solutions for publishing public sector information.

The common interests and activities between the PSI and academics to find a suitable approach to data publication produced a scholarly research paper (Alani et al. 2008) that acted as an artefact embodying the commitment and activities between government and academic researchers, and also as a tool to gain the interests and attention to other potentially valuable actors.

The scholarly paper documented the success of specific collaborative projects, highlighting the capabilities of pre-existing technologies such as the Semantic Web (Alani et al. 2008). The paper has a dual purpose, to illustrate the activities between the PSI and academic network and also demonstrate the social and technical merits of adopting Semantic Web technologies:

> *"The work reported here was intended to show that SW technology could be valuable in the governmental context. An important outcome of the project is the level of awareness that has been built up in government about the potential of SW technology. Having seen what the SW technology is capable of, and the success of the pilot study of AKTivePSI, OPSI is now funding a second project"* (Alani et al. 2008)

This paper had been inscribed with the interests of academics, and by doing so, made them an indispensable actor in the emerging PSI and academic actor-network. Simply put, the academics were required for social or technical development to proceed. The academic actors became a focal actor, they were drawing upon previous research interests and technologies, and setting the obligatory passage points for which all the actors involved were required to pass through in order for the network to translation.

Whilst scholarly publications were important artefacts within the alignment of the academics and government (PSI) network; their production impacted other actor-networks. These artefacts (Alani et al. 2008) were presented at the International Semantic Web Conference (ISWC2008), an international event containing actors with an interests towards the social and technical development and impact of Semantic Web and Linked

Data. By presenting the activities to the Semantic Web community illustrated the academics commitment to previous networks, whilst also showing that they were operating with a new agenda, and as part of this, drawing upon the Semantic Web and Linked Data in order to achieve this:

> *"The UK government is committed to leveraging and producing open standards, and the GovTalk programme has key documents that describe inter-operability frameworks and metadata standards. With this in place, the scene is ideally set for SW technologies now to take centre stage."*
> (Alani et al. 2008)

Examining the above quote by Alani et al. in more detail, embedded in this text is not only the announcement a forming relationship between the UK government and academic researchers, but also the academic's inscription of a technology they were previously aligned with. Semantic Web (SW) technologies were problematised as a tool to help advance the UK governments' open standards. This provided a way for the Semantic Web technology to become part of the network, and also positioned it as a catalyst for other actor-networks to become interested.

The development of common interests and goals between the academic and PSI network was helped by the activities related to the Semantic Web and the social processes of publishing Public Sector information. The problematisation of actors had been achieved, a process aided by the use of immutable mobiles and network artefacts (papers, projects, reports).

The analysis now extends to examine the immutable mobiles used by the third network in the emergence of the UK OGD community. The OKFN actor-network represents a collection of actors who share an existing interest into the social and technical affordances of openness on the Web. This alignment of actors within this network emerged from a shared interest and rhetoric of providing various forms of digital openness, which included data, information, knowledge, and access. This network contains a diverse mix of committed actors, as is most suitably labelled as the Open Knowledge Foundation (OKFN) actor-network.

OKFN's interests towards opening up government data can be associated with a variety of different network artefacts and methods of engagement. Public reports were one example of a pre-existing study which helped the OKFN establish itself as a network interested in opening up data (Newbery et al. 2008). The report highlighted the benefit of publishing public sector information, and by doing so raised the profile of the OKFN network, demonstrating their interests towards opening up government data; as describing by an interviewee:

> *"I think it was important, in that those Figures of the economics of Rufus's work were quoted heavily in campaigning and then also made it into the Tory party manifesto or their IT strategy [...] Rufus and others got the attention of Francis Maude and others in the government"* [Tim Davies: 152]

Rather than approaching this from a technical perspective, they discussed the social processes required to publish public data in reusable formats, and highlighted the economic benefit by doing so. This helped raise awareness of the OKFN's activities to other networks already involved within examining the publication of government data, including the actors in the academic and PSI network.

The use of online communications platforms offered OKFN a way to share their findings through various forms of digital mediums: blog posts, conferences, and media broadcasts. By using these technologies, they were able to demonstrate their activities and interests of opening up government data:

> *"We want the data raw, and we want the data now"* ['OKFN Blog', 'Give us the data raw, and give it to us now', November 7th 2007]

This statement, as interviewees suggest, was one of the milestones for the OKFN to begin its alignment with other actor-networks involved with openness and Open Data. This blog problematised the technical issues of public sector information, such as the balance between the focus on data, and the design of "shiny front-ends". Consequently, the statement of 'data is primary, the interface secondary' helped gain the attention of the academic and PSI actor-networks as an interviewee suggested [Johnathan Gray: 181].

Online Communications have more than one use, on the one hand, they are a tool for self-promotion and reinforcement of a network's agenda and role [Tariq Khokhar: 209], on the other hand they provide a way to find and share information [Chris Taggart: 109]. Network artefacts can transcend online communication technologies, potentially enabling a larger audience to view and comment. Take for example the concerns raised in blog posts by the OKFN, responses to this were discussed and addressed in communication platforms including the OKFN mailing lists:

> *"The work must be provided in such a form that there are no technological obstacles to the performance of the above activities. This can be achieved by the provision of the work in an Open Data format, i.e. one whose specification is publicly and freely available and which places no restrictions monetary or otherwise upon its use."* ['okfn-discuss', August 07]

By examining other posts within the mailing list demonstrated the translating agenda, as discussions regarding Open Data became focused towards 'Government', which eventually led to the introduction of the 'Open Government' concept.

> *"Looking specifically at the principles it seems that they [Government] are looking at something slightly wider – what one might term *open government* as opposed to simply *Open Data*."* ['okfn-discuss', December 2007]

Despite being a simple statement, the introduction of 'Open Government', which drew parallels with 'Open Data', acted as a re-problematisation of the current technological artefacts and social practices inscribed in the Open Data activities: Although actors within the OKFN network were already problematising ways to examine government data; during this timeframe, the activities of Open Government Data were not yet established. However, this demonstrated how an agenda is built upon multiple inscriptions from different networks and from previous activities of actors.

Other technologies were also used during OKFN's problematisation. As part of the heterogeneous network of activity, OKFN offered technological artefacts, which were inscribed with their research interests of providing an open source, Web-enabled data storage and retrieval platform. CKAN (Comprehensive Knowledge Archive Network) represented an actant (and actor-network) within the OKFN and also an interessement device for other networks working towards the Open Data agenda. As part of the technical architecture for storing and accessing Open Data, it provided support for Linked Data; a data format specifically engineered the Semantic Web. CKAN played an important interessement device between the OKFN and the academic actor-network; it responded to the initial problematisation of using Semantic Web and Linked Data technologies to support the publication of public sector information Alani et al. (2008).

7.2.2 The Interessement of Networks

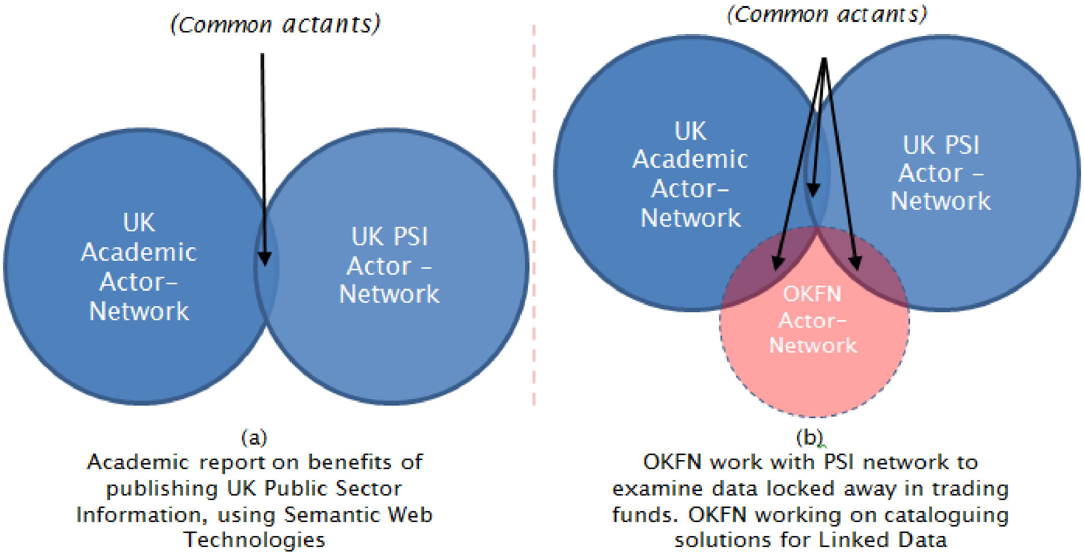


FIGURE 7.1: Illustrating the simultaneous translation of the Academic, Government, and OKFN networks, and how common actants enable these networks to become aligned to a common goal. (a) The alignment between the Academic and PSI network. (b) The alignment of the OKFN network with the alignment of the newly formed alliances of the Academic and Government networks

Figure 7.1 illustrates the alignment of the academic, PSI and OKFN network. Through shared common actants, interessement devices and different forms of immutable mobiles, interactions and relationships were forming. However, as Figure 7.1.b illustrates, although there are common actants beginning to emerge between the Academics, Government and OKFN, common actants that overlap all three networks are still yet to emerge.

Interrogating the online communication data introduced in the previous chapter, it is possible to 'zoom' into the data to explore the online communication patterns during the time of these activities described. As Figure 7.2 illustrates, the communications

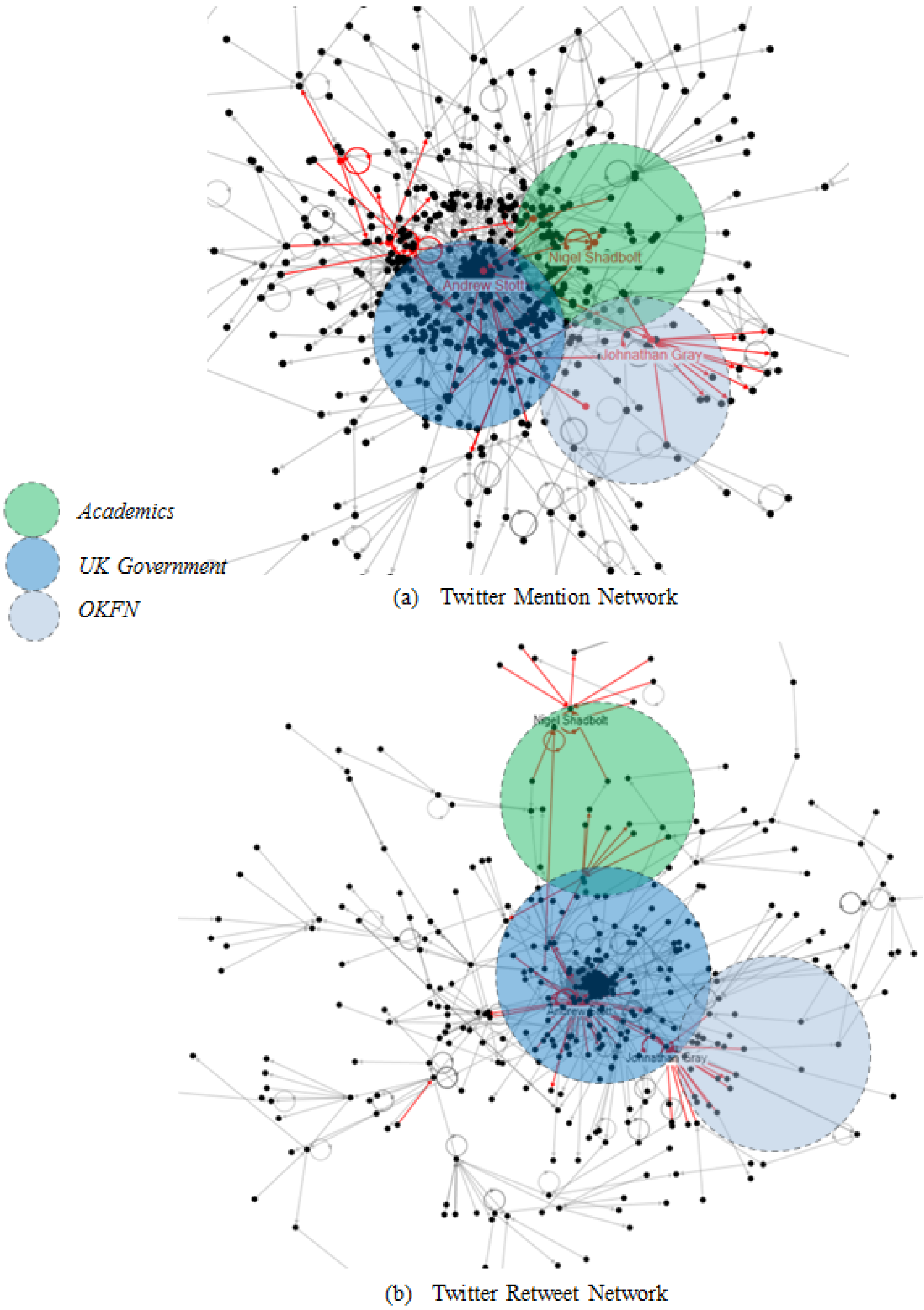


FIGURE 7.2: Visualisation of the Twitter (a)mention and (b)retweet network during the time period related to the events discussed between Sept 07 May 09. The highlighted edges and names represent the different actor-networks discussed Academics, OKFN, UK Government

between these networks are achieved by a number of individuals acting as bridges, or structurally, the 'weak ties' between networks. Specific actors act as bridges to enable different actor-networks to communicate and interact with each other, a structure that occurs in both the retweet and mention Twitter communication networks, demonstrating how information is spread and discussed between the different actor-networks.

By using techniques to filter the communications (Tinati, Carr & Hall 2012) it is possible to identify the diffusion of content which was popular and spread between the identified networks, including information such as:

> *"RT @danslee: UK gov pledge for #opendata. There will be pain. But it'll be worth it. (@govwiki @Nigel_Shadbolt) http://bit.ly/dcBhTq #gov20"* [Twitter Data Stream, 'sarahlay', 17th July 2008]

> *"How #data transforms democracy. Big report by @DeloitteGov http://bit.ly/aQneS0 #gov20 #opendata"* [Twitter Data Stream, 'danslee', 12th October 2008]

Tweets such as these demonstrate the government's interests towards engaging with Open Data as well as engaging with actors who were part of the academic network (the tweet provides a direct connection between both UK government and the academics). Building upon the initial analysis, it was noted that during this timeframe, the degree of connectivity between the three networks is greater within the mention network, whereas in the retweet network, actors within the government network were (structurally) the bridges between the academic and the OKFN network, even though the OKFN were actively pursuing the alignment of the academic network with CKAN's Linked Data capabilities.

In addition to conversations, the diffusion of information provided a method for actors to share network artefacts to a large number of actors. Take for instance the 'raw data' statement made by actors in the OKFN network; this was discussed and shared between many actors using online communications, contained in tweets such as:

> *"I read Impossible possibilities A wishlist for linked data by @pezholio #opendata #rawdatanow #egov http://bit.ly/2nqZNR"* [Twitter Data Stream, 'philipjohn', 10th October 2008]

> *"The Open Data cook book. Sourcing and scraping raw data. http://ldata.in/evJ09k #opendata"* [Twitter Data Stream, 'dHolowack', 29th January 2009]

As part of the diffusion of the 'raw data' message, subsequent tweets followed which problematised the use of Semantic Web and Linked Data to format, store and publish Open Data. Drawing upon analysis in the previous chapter, these tweets were shared by actors identified as the 'weak ties' that were connecting different networks of actors together, which as Figure 7.2 illustrates represented the academic, OKFN and government actor-networks.

Online communications were not the only way common actors were helping shape the formation of the UK OGD community; the offline activities and associations of actors-networks were also responsible for this. In order to illustrate this, the analysis turns to looking at an international actor-network that was driven by a similar agenda of producing social and technical solutions for publishing Public Sector information. Like the PSI and academics, The US OGD network were actively seeking ways to expose public data, and had made use of Semantic Web technologies as a solution to publishing government data.

However, shared between the US and UK academic-network was common actors that had previously been associated through academic ventures [Nigel Shadbolt: 23]. These common actors as shown in Figure 7.3, operated as a bridge to enable these two actor-networks to communicate and interact with each other ['The Guardian', 10th June 2009]. They functioned in a similar way to that of the common actors identified between the UK academic and PSI networks, where their shared associations in both networks enabled them to mediate information and activities between the networks. Essentially, their activities were identified as the "pivotal points" [Chris Taggart: 96] to enable the different actor-networks to interact.

Reflecting on the observations of networks of similar agendas, often, networks that have a set of similar agendas, goals and outcomes are identified as networks in competition with each other; typically, competition is considered as a threat to a network's translation (Latour, 2005). However, despite both the US and UK networks striving to achieve similar outcomes by using Semantic Web and Linked Data technologies to publish government data; the activities of the US network were not considered as a threat to the UK network's translation; instead, they offered support and guidance:

> *"The government is known to be working on the creation of a central data source from which all sorts of government data could be accessed, as has*

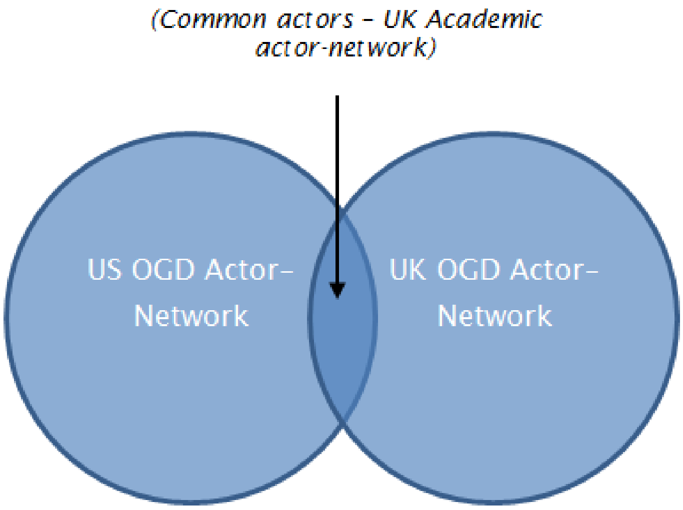


FIGURE 7.3: Illustrating the shared actors between the UK and US OGD network, the US Actor-Network has been punctualised as has the UK OGD Actor-Network. The common actors represent the UK Academic Actor-Network who share an interest within both networks

> *been introduced by the Obama administration in the US"* ['The Guardian', 'Web Inventor to help Downing Street open up government data', 10th June 2009]

The relationship between the US and UK was considered as one that formed through 'healthy' competition [Johnathan Gray: 351], and their individual and joint interactions and activities, whether through communication, support, or producing network artefacts has been essential to the translation of the UK network:

> *"There have been lots of little things which have added up to shape the current state of official Open Data initiative [...] there isn't a recipe and it's hard to abstract from this what the most important factors were. One has to talk to lots of people, consider things from lots of different angles, pay close attention to developments, and make the most of opportunities as they arise."* [Johnathan Gray: 182]

7.2.3 Finding Common Interests and Strengthening Ties

The technological developments and social processes inscribed by the OKFN had provided sufficient activities to gain the interest of the academic and government actor-network. The ability to entice actors using technological interessement devices led to the formation of associations between networks, which effectively enabled shared actors between networks to become established. This provided an opportunity for the OKFN to strengthen their relationship with the government network, and was helped by the alignment of focal actors within the academic network who took on OKFN's agenda of 'raw data':

> *"Sir Tim Berners-Lee taking up Rufus Pollock's call for 'Raw Data Now' in his TED talk"* [Jonathan Gray: 180]

> *"Ministers wanted to meet Tim, rather than Tim having to find a slot. Anyone else wanting to meet with a minister has to go through layers and layers of diaries"* [Tim Davies: 164]

This shows how specific actors work as intermediaries between networks. In order for these actors to bridge between networks, they need to to have a common interest within both. However, as the quotes above suggest, an actors role and presence within the network the, which could be their connectivity, or simply their 'star power' can be useful for drawing different interests together.

7.2.3.1 Formation of an Actor-Network

A complex network of networks is beginning to emerge. Each network is simultaneously translating both towards their own set of goals that have been driving them previously, and also developing common interests between networks. Consequently, a new actor-network is beginning to emerge. This is driven not by a single set of actors with a pre-defined set of interests, but by a dynamic translation process consisting of various actors; from high level government networks, to academics and not-for-profit organisations.

The unfolding social and technical activities are driving a co-constructive process whereby humans, technologies and network objects push towards translation. The agency of the actants changes as a result of their entry into new networks. In as much as there is any stability, this is temporarily based on the stability of the existing networks, committed to their original goals, or at least, not overridden entirely due to the realignment with new goals and networks.

The relationship forming between the academics, PSI and the OKFN actor-networks raises some questions about the possible triangle of interessement (Callon 1986*a*) that might occur between these networks. Simply put, the networks that have the strongest ties (common actors, activities, goals) tend to threaten the relationship between the weakest networks. However, as the engagement through the online communications and interest towards their reports (Newbery et al. 2008) illustrated, this does not appear to be the case. OKFN have gained the interessement and alliances of both the academic and government network.

The interactions and strengthening of ties between the government, academic, and OKFN networks has enabled a translation process. These networks have individually identified common interests of publishing government data, and how their own activities can contribute to this. In addition to this, the influence of other international networks with similar goals have been beneficial to the initial translation process within the UK, a finding which contradicts the belief that competitive activities from other similar networks threaten and negatively disrupt the process of translation (Latour 2005).

In summary, discovered in this section:

- The academic, PSI, and OKFN actor-networks have been identified, their previous activities and their interests towards the agenda of publishing government data have been examined.
- These networks which were translating separately have identified common interests using network artefacts and started to re-problematise their own agenda towards a new emerging actor-network
- Focal actors within the network provide online and offline bridges for the actor-networks to enable activities, communications, and artefacts to be shared.

Whilst applications of the concept of translation within previous ANT studies tend to focus on the processes surrounding a single and coherent network of connected actors (Callon 1986*a*), this section will use the concept to demonstrate how multiple disconnected actor-networks became part UK OGD, rather than one coherent network configured by a single agenda.

7.3 Another Set of Problematised Actor-Networks

Whilst (central) government, academics and OKFN were involved with the drive towards an Open Government: creating technologies, policies, passing legislation and examining the benefits that providing access to this information would create; local government networks were also problematising a similar agenda. Actors in the Greater London Authorities (GLA) were exploring the potential of opening up London council and local government data. Actors in the 'GLA actor-network' existed previous to the problematisation of public sector information or Open Data, both at the national and local level.

7.3.1 The Problematisation of London Government Data

The context for the emergence of the London actor-network surrounds the election of a new mayor of London, whose instantiation along with a new administration introduced a new agenda of offering access to London Government data which was previously unpublished and not available for public consumption:

> *"The Innovation Exchange Programme will focus on designing practical, efficient, and effective policies, overcoming implementation challenges, and evaluating success in making government more transparent, accountable, accessible and efficient"* [Mayor Report to the Assembly, June 2008]

In addition to setting the goal of providing data access, the report drew upon the mayor's previous associations with other international networks (New York City's government) and described how 'transparency and accountability' would be at the forefront of London's agenda. The agenda set out in the first report was soon followed up a second publication reconfirming the agenda openness and transparency. The inscription within this report was stronger, and reconfirmed the commitment towards a more specific set of goals regarding the social processes that the London network aim to achieve:

> *"greater transparency through the publication of adviser's job descriptions, register of interests and salaries online"* [Mayor of London Report, September 2008]

The initial problematisation of the London network was led by the production of the statements embedded within the reports and welcoming speech of the London Mayor. These gained the attention of actors interested in accessing government data and working on data projects. However, reports were not the only tools available for the London network to gain the attention of other actors to help the translation of their agenda.

The use of data was also critical for the problematisation of the London network. Driven by the new agenda of publishing data and information, incentives in the form of datasets and a new website were made by the digital services department within the London government. This had a two-fold effect: it strengthened the ties of actors already established within the London network, and also demonstrated to other actors the commitment (including financially) towards achieving their agenda:

> *"I have approved three elements of work related to the development of the London.gov.uk website. [. . .] I have approved the variation of the current OJEU contract with Amaze to carry out additional data migration services for the purposes of migrating content and data from the existing London.gov.uk website and associated applications to the new content management system based website. [. . .] The cost of this work will be £32,425."* [Mayor Report to the Assembly, January 2009]

The development of the London.gov.uk website and the migration of the data to a new content management systems became part of the London network's obligatory passage points; the goals that need to be achieved in order for it to occur. By declaring a budget demonstrated a commitment towards achieving the goals, setting the new website as a passage point.

Whilst the new website was being promoted as an 'information Laboratory or Observatory' [Emer Coleman: 17], which helped strengthen ties within the existing GLA network, it was also positioned in a way to interest other actors that might find the publication of London government data useful. These actors, the 'users', were seen as important in this process since successful use of the data ensures translation:

> *"let's just try and connect these people who are the primary users of the data, which in the first instance will be developers, and then we used an open call on Eventbright, and said we would like to open up this data, we need you to come and help us."* [Emer Coleman: 20]

> *"How do you make open government data so no one can turn it off? Well that does mean people have to be using the data, and preferably, building solid business on that, who will scream if the data ceases to flow. If you build a business on that [Open Data], and the data stops suddenly because the government is not giving out the data that it holds."* [Andrew Stott: 244]

Gaining the interests of users was achieved by processes at the technical layer as well as the social. Their alignment with the London network was the result of a mix of passage points: firstly developing a Web platform which stepped away from being "proprietary and closed" [Emer Coleman: 19], and secondly, recognising that the engagement of data users was necessary for future stability. As a result of these activities, users were able to "play" with the data, which in turn, produced additional artefacts to spur on further discussions and activities:

> *"RT @hadleybeeman: New dataset Greater London Auth. expenditure incl all payments over £500. #opendata #transparency http://bit.ly/alhEfh"* [Twitter Data Stream, 'jkonga', 2nd March 2009]

> *"RT @OpenDataLondon: City of #ldnont Tree Inventory Data (Trees on Public Property) is now available http:ht.ly46yM6 #opendata via @rfldn"* [Twitter Data Stream, 'rfldn', 24th July 2009]

However, the data users that the London network gained the attention of, were already a network of actors working towards similar activities of using Web data (of various formats and types) to achieve a technological development or service:

> *"The developers were always there, wherever you look at an initiative like this, there are precursors, so we were lucky that we got top down backing at that point [...] there had been a group of developers agitating from the bottom-up, asking why we can't get our postcode data, what's this about not being able to get hold of other types of data. Lots of requests for data, they [developers] needed it to get stuff to happen."* [Nigel Shadbolt: 71]

> *"There were people who were doing stuff around Open Data far before then, [...] they had quite a big network of developers that this is simply around the idea of Open Data in all its forms, not just for the kind of democracy and transparency"* [Julian Tait: 103]

These actors had previous interests in the development of technologies and share an interest in Open Data and were active in requesting and using data in order to develop tools and software, both for their own private interests and for business purposes. Given this, the analysis will turn to examine how these actors were also undergoing their own problematisation and became interested with UK OGD.

7.3.2 The Problematisation of Developers

The problematisation of the developer actor-network to the OGD agenda is somewhat different to that of the other networks. Initially this network was less coherent in structure than the actor-networks already introduced. Rather than offline associations, actors were associated through various forms of technological mediation such as online communications and mailing groups:

> *"I don't consider myself as formally part of a community. Perhaps at the edges of it, I've known some people in it, [...] "we're not in this community that is very close knit, but we know each other's names, and we know what each other is doing and know each other's work. I think that's how I kind of got onto the stuff"* [Chris Applegate: 5:35]

Other developers reflected this view of the developer 'community' as a distributed and loosely connected network of actors, whereby the majority of their activities and associations were facilitated by the use of Web communication platforms such as Twitter, or

Blogs. This reflected the findings of the previous chapter, where the network could be identified as multiple sub-networks of actors.

The formation of the developer network can be attributed to activities of the actor-networks already discussed: the academic, government, OKFN, and also the London network. These networks provided a way for the actors only associated by online links to meet offline, in face-to-face meetings, events and workshops:

> *"So I really noticed it starting to change was when Emer contacted me and when developers started to realised that they had to engage with government and policy makers on this."* [Malcolm Barclay: 43]

> *"we got about 60 developers and people interested in Open Data turned up to that workshop and they kind of set the agenda for me, they said don't worry about formats, just get it out there, and if there is something wrong with it, as long as it's not PDFs, we'll manage it."* [Emer Coleman: 22]

Developers were now offered a new way to interact and connect with others that were interested in similar activities. However, this was not entirely a social process, as it was supported by the use of data as an incentive for developers to participant. Nor was it only beneficial for the developers. By allowing the developers to play with the data, in turn it demonstrated the capabilities of publishing government data. Effectively, the data, the social activities, and the outcomes had provided a chance to establish common goals and further develop the support and interest of the actor-networks involved.

Figure 7.4 illustrates, as a consequence of the interessement device, offline interactions and events such as the workshop described above led to actors associated with the GLA network to become a common actor across both networks, becoming indispensable within the network, thus focal in the translation process. Prior to the GLA network's engagement with the developer actor-network, actors were weakly connected and did not share a common obligatory passage point. However the role of the GLA helped strengthen the ties between the actors, which helped form associations; providing an agenda and goal for the translation of the developer network:

> *"So there are people that get a name for themselves that get involved in these things, but I don't know whether there is a hierarchy, they probably set*

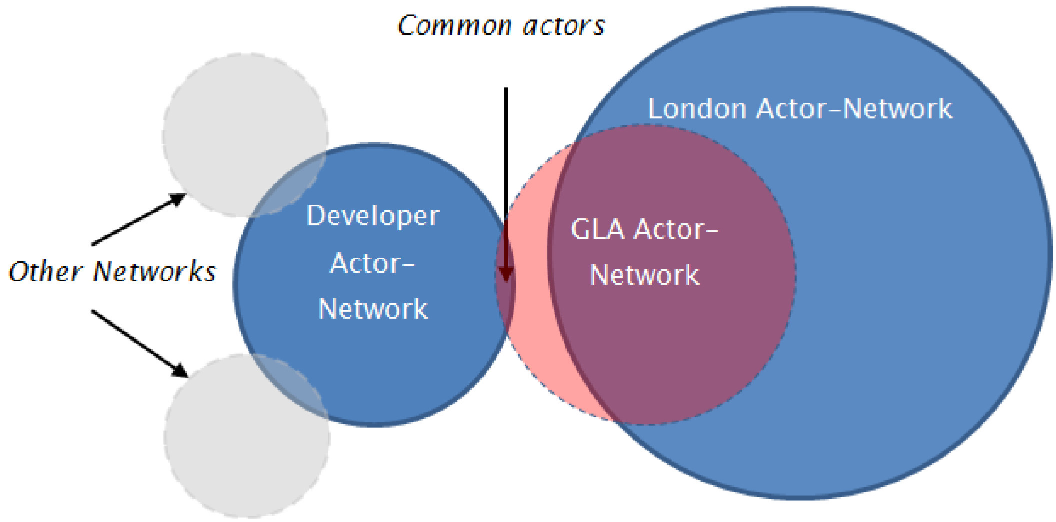


FIGURE 7.4: Illustrating the relationship between the London Actor-Network (and the GLA Actor-Network punctualised within) and the Developer Actor-Network through the use of Interessement Devices. The diagram also demonstrates a number of other networks attempting to gain the interests of the Developer Network (Gray 'Other Networks')

> *the agenda to an extent [...] Emer Coleman sets the agenda that is interested in London data, as that is what she in-charge of. Not that she is actually forcing people to study this stuff, but she shapes the Datastore, and how data gets released."* [Zarino Zappia: 72]

However, gaining the interests of the developers was more complex than this, and to demonstrate how, the analysis will focus on the problematisation of OGD from the developers network. Previous to the associations with the OGD agenda, developers had been involved in creating technologies using Web data, but this was laced with technical, financial and legal issues.

> *"I know developers who have been on this for years, and we don't give up! They just keep on harping, and with the internet at our fingertips and the detailed understanding of how it works, it becomes an incredibly powerful tool, transparency for us is instant, especially for us. We know we are up against a brick wall, especially when we are confronted with an SSL logger and contractual lists that are five pages long that sign away your life."*
> [Malcolm Barclay: 93]

Developers were active campaigners for the publication and re-use of public data sources to create websites and tools, specially targeting data feeds from public services and reformatting for their own use. As actors within the developer network explained, obtaining public feeds such as the Transport for London (TfL) data stream was not possible and were often faced with social, technical and even legal barriers.

> *"I released TubeStatus in 2008 and the only way I could get that status data out to an iPhone App was to actually scraps the TFL website. Even though they had their own XML feeds, they were not known and not published and nor was it encourage if you used it either. So I was in a bit of a conundrum as to the many issues that I was doing. So that's how it started out"* [Malcolm Barclay: 7]

Given the social and technical barriers between themselves and the access of data, developers required alternative ways to obtain the kinds of data they were interested in. The challenges faced were not possible to overcome without the help of other networks.

> *"Developers realised that we can scrape and hack away as much as we like, but to actually get something, we have to engage with policy and decision makers, because otherwise you will just be renegaded"* [Malcolm Barclay: 39]

However, when presented with an opportunity to obtain free, reusable forms of data, rather than having to tackle the technical issues of scraping Web pages, risking legal issues, or working with inconsistent or unreliable data, Open Data, specifically government data became a highly prized resource.

> *"They [government] were actually inviting us to be included and consult in the discussion of policy forming around Open Data."* [Malcolm Barclay: 44]

From one analytical perspective, the developers could be depicted as an actor-network being problematised by other actor-networks involved in the translation of OGD. However, examining this relationship from the perspective of the developers illustrates how they were also problematising other networks. This dual sided commitment is an example of how actors (and actor-networks) become aligned and enrolled during translation, and also the formation of a co-constructive relationship which is reliant on the cooperation of humans and technologies.

In short, multiple networks have contributed to the initial stages of translation of UK OGD. There is a non-linear process of activities and complex network of actor-networks, each inscribed with their own goals and network activity, yet brought together by common interests and shared passage points. This reasserts the point raised earlier regarding translation not being a process of a coherent and connected set of actors or networks, but rather a distributed process, requiring multiple observational windows and sources of data.

7.3.3 Bridging Disconnected Actor-Networks

Whilst attention has been paid to the interactions and relationships of the different actor-networks interested in the OGD agenda, observations have shown that their activities are not isolated, and various network entities are available to bridge the networks together. The analysis will now explore how these entities: humans, technologies, data, and online and offline communications formed these bridges.

In the previous section, the emerging set of associations between the OKFN, academic, and government actor-network was helped by common actors working as bridges. These common actors shared interests in more than one network, and through their associations, artefacts could be exchanged, and relationships could be formed. Observations showed how actors within the OKFN and academic network provided offline and online connections and technological artefacts (e.g. the call for raw data, the development of CKAN) to strengthened the relations between all three networks. These types of actors were also essential for working with developers and those interested in the use of Open Data:

> *"I could leverage in the help from the developers, in a way that central government was much was cautions about."* [Emer Coleman: 61]

> *"My role is to work with those stakeholders interested in Open Data and alongside public bodies who publish government data. You could broadly break this down into working with the data supply side on the one hand (e.g. public sector bodies looking to publish data), and the data demand side on the other (e.g. developers, journalists, NGOs and others who wish to use the data to do something useful or interesting)."* [Johnathan Gray: 33]

The quotes above taken from bridging actors in the London and OKFN network describe their roles in connecting different actor-networks together. This included interesting new actors, especially those that can make use of the data for development purposes. By taking advantage of the common interests towards Open Data, public sector information, and transparency, the OKFN actor-network was able to leverage their position to help form a set of associations between the developers, PSI, academic, government, and themselves. By providing face-to-face events on multiple occasions, these actor-networks were given the chance to become aware of each other's activities, providing the conditions for common goals and ideologies between networks to be established and flourish:

> *"The Open Knowledge Foundation ran a workshop with civic servants, developers, journalists, researchers and other people, and we spoke about the value of having a central point of reference for Open Data which was available across the UK. We mapped common datasets that people wanted and created records for them in CKAN"* [Johnathan Gray: 163]

The OKFN's ability to take advantage of their bridges with other actor-networks helped produce alliances between those networks previously unconnected, and helped strengthen ties with those already establish. It also provided an opportunity for actants CKAN to establish itself as a valuable technology for managing and publishing government data in various formats, including Linked Data (which catered for the academics interests).

> *"CKAN was useful, in that it removed a barrier [...] they needed a way to publish the data, and CKAN was that technology"* [Chris Taggart: 198]

CKAN contained the inscriptions of actors within the OKFN; it was developed with their ideas and interests, and its capabilities of publishing data using open and reusable formats reflected this. By CKAN becoming the technology of choice, it became an Obligatory Passage Point, becoming indispensable for achieving the OGD agenda, as by proxy, made the OKFN a required actor-network as well.

By CKAN being available as a platform to store and publish data, the developers were able to take advantage of this, and use the stored data in order to develop additional technologies that could demonstrate the capabilities of the government data:

> *"The real catalyst has been developers taking data, and produce useful applications for the public. That really turned the heads of the public, politicians, [. . .] that was the point at which a lot more people said, this is a really good thing"* [Jonathan Gray: 149]

By doing so, the developers were able to get the attention of actors within the other participating networks, including the academics and UK government. This also gave the developers a chance to demonstrate their capabilities and use within the OGD agenda; effectively made them indispensable within the translation process, and an additional passage point within the translation process.

Technology became a bridge, as well as an artefact to enable new social activities to occur, which in turn develops additional technologies. Effectively, CKAN helped shape the relationships between the academic, government and developer actor-networks, which in turn shaped the technologies that were created by the developers; and underlying all of this was their alignment of a common interests to open up government data.

Whilst CKAN was essential for overcoming technological challenges, the interactions of the actors in various networks was just as critical. The online and offline activities of actors were also crucial to the translation process:

> *"[CKAN] removed a barrier. But if you hadn't had someone like Richard sterling or Andrew Stott, even if you had CKAN it wouldn't of made any difference."* [Chris Taggart: 200]

> *"The developer community in general was very influential. I remember several times, just hoards of developers would be in the cabinet office just talking to Tom Watson and others."* [Emma Mulqueeny: 122]

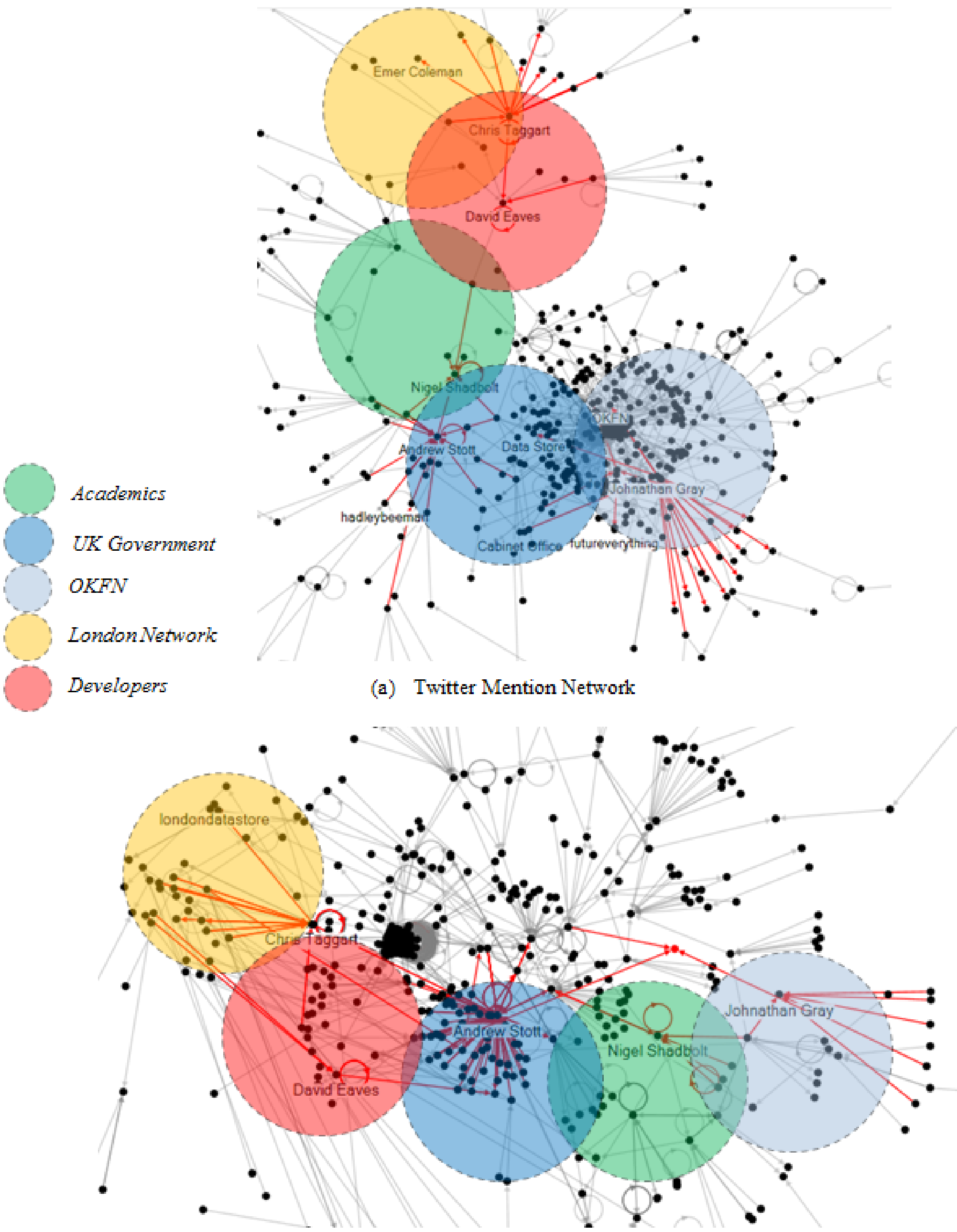
The combination of technologies and human activity represents the socio-technical processes that enable translation to occur. CKAN (which contained the inscriptions of the OKFN actors) provided a way for developers to get other actors attention, but CKAN also required the support of other actors.

The socio-technical interplay between technological capabilities and human activities is also observed through Web communication platforms such as Twitter. These technologies helped actors create and strengthen ties within and between the identified networks. It also provided a medium to communicate the requests and needs of actors who were predominantly only known through their online communications (e.g. the developers), which then could become embodied through offline activities by the bridging actors. As discussed below, by using Web communication technology, these actors were able to identify themselves to other actors (in this case, developers) in order to pursue their goals:

> *"I used that [Twitter] as a means of circuiting government, so when I was having blocks and barriers with agencies, I was Tweeting out, 'I'm going out to GSL today' and all the developers would go, 'Yes, hurry up and get the data!'. That was kind of putting a pressure on the agencies, because they knew the debate was happening in public. So I had a huge amount of freedom, because not only because others knew what I was doing, but to design to a method of putting pressure on the state from the inside out."*
> [Emer Coleman: 56]

As illustrated in Figure 7.5, by tracing the digital interactions of the actors within these identified actor-networks, observations show the developers were situated and communicating with actors in other networks such as the government, academic, and OKFN network. As before, by using filtering techniques (Tinati, Carr & Hall 2012), it is possible to expose highly shared and diffused information between the actors within the network, which include the use of interessement devices to gain the support of the developers, and also developers discussing their achievements and network artefacts:

> *"Calling Open Data developers: we need your help http://bit.ly/QfiLQ #opendata #gov20 #ukgov"* [Twitter Data Stream, 'DirDigEng', 30th September 2009]



(a) Twitter Mention Network

(b) Twitter Retweet Network

FIGURE 7.5: Visualisation of the Twitter (a) mention and (b) retweet network during the time period related to the events discussed during May 09 December 09. The highlighted edges and names represent the different actor-networks discussed Academics, OKFN, UK Government, London Network, Developers

"RT @rewiredstate: UK govt opens data sandpit to developers. http://bit.ly/GoR8N #opendata" [Twitter Data Stream, 'mashthestate', 6th October 2009]

"Cracking the Postcode http://bit.ly/1et9jM free postcodes for all

> *#opendata [wow! Nice Hack Sir!]"* [Twitter Data Stream, 'osbornec', 9st October 2009]

The tweets above provide an example of the capabilities of online communications platforms; actors were able to share information of other actors from different networks (retweeting), and also providing external network artefacts, which in this example linked to official government pages that provided details on how developers can engage with government data ['Cabinet Office', 'Calling Open Data Developers We need Your Help', 30th September 2009].

Technologies such as Twitter enabled actors to connect and communicate with a large network of actors in a quick and responsive fashion. The technologies provided a platform for discussions to be initiated and action to be taken, and as a consequence, it gained the interests of additional actors. As Figure 7.5 illustrates, new networks of actors had formed in comparison to the communications shown previously (Figure 7.2). This is observed in both the mention and retweet network; with an increase in number of active communicating users by 24%. In addition to this, unlike before, the London actor-network is now active in direct communications and sharing information with the developer network.

Twitter also provided a mechanism to share information, and the developers took advantage of this by forming the bridge between the London and other networks, which then enabled them to share the activities of different networks interested in the OGD agenda:

> *"RT @DirDigEng: Please help shape URI standards for the UK Public Sector http://bit.ly/avcIA #opendata #linkeddata #gov20"* [Twitter Data Stream, 'CountCulture', 12th October 2009]

The different forms of interplay between technologies and human activity represent the socio-technical processes that have enabled translation of the actor-networks to occur. This interplay, whether through technical functionality or human activity establish and strengthen ties between the actor networks that were previously unaware of each other's activities, interests, or capabilities.

7.3.4 Obligatory Passage Points and Goals

The translation of an actor-network is achieved by actors passing through the obligatory passage points (Callon 1986*a*), or more simply, actors need to produce outcomes based on an agenda and goals set (Tatnall 2010). However, as it has been shown, UK OGD involves multiple actor-networks, each with their own interests and goals. Despite this, they all share an underlying common interest which involves government data, whether this interest lies within developing the technologies, the social processes, or just access to the data. Consequently, by setting obligatory passage points which was driven by a common goal, the networks became aligned, driven by a purpose, data.gov.uk.

Data.gov.uk represents the goal of the common interests and activities of the academic, OKFN, UK government and developer actor-networks. As Figure 7.7 illustrates, data.gov.uk is the result of these networks working in alignment with each other, which led them to form a variety of obligatory passage points, set by the different networks involved within the translating process.

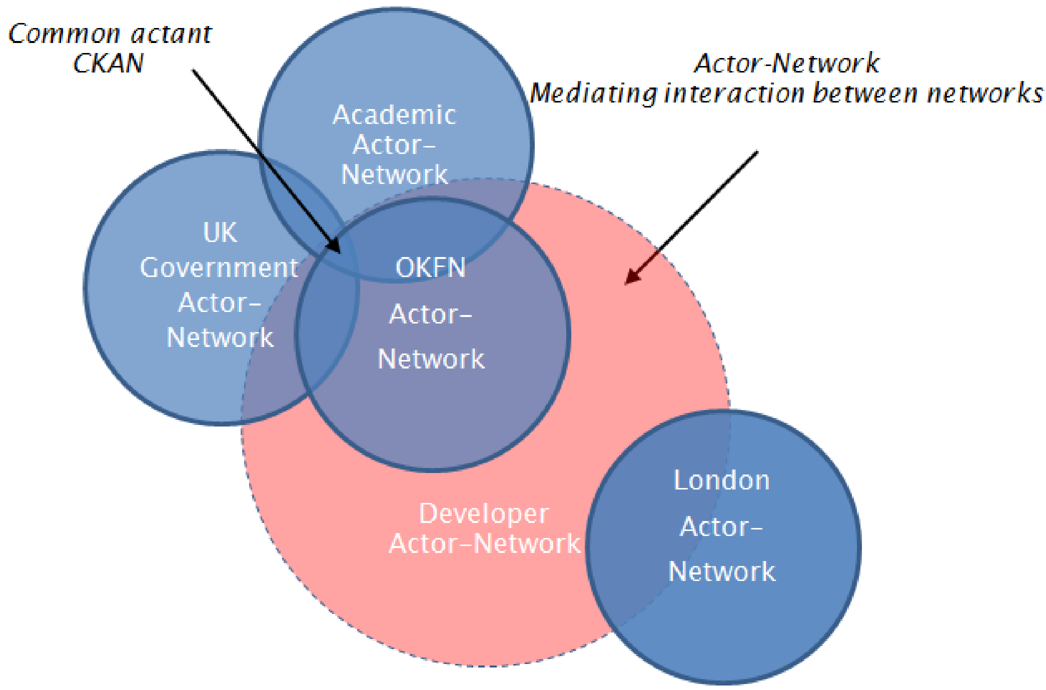


FIGURE 7.6: Illustrating the inclusion of the developer and London Actor-Network into the existing networks, which are already translating towards OGD. The developer network acts as a bridging network connecting the networks together

This goal helped align the different actor-networks towards a common agenda. As the networks pass through the passage points, the common goal of publishing government

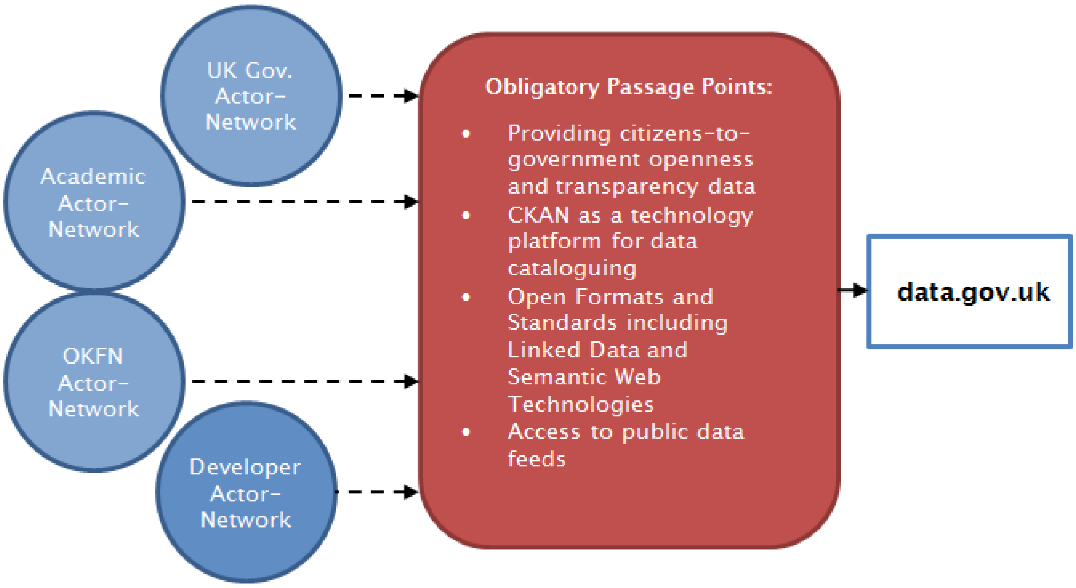


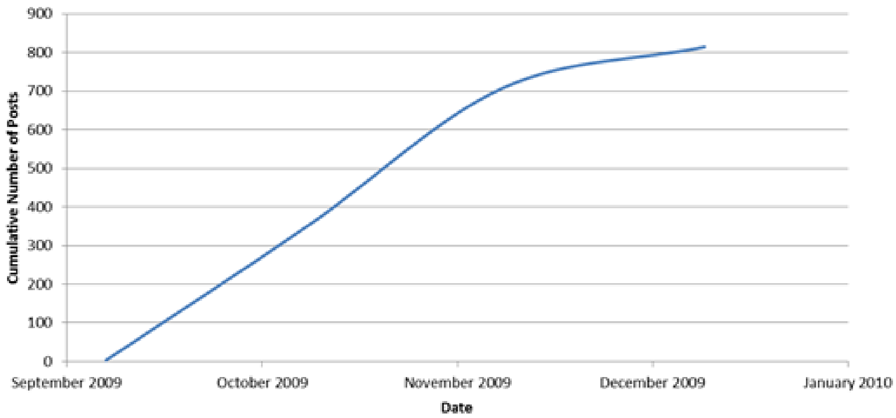
FIGURE 7.7: Passing of the Obligatory Passage Points (set by the different agendas of the networks) by the various actor-networks. The shared interests and goals led to the outcome of data.gov.uk

data should be reached, whilst simultaneously the activities and relationships of the pre-existing networks must remain stable. The successful translation requires all the actors involved to be aware of and committed to the passage points shown in Figure 7.7, which have been inscribed and shaped by the activities of the various networks.

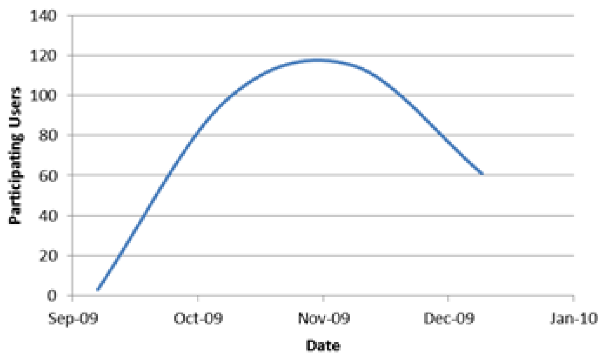
Driven by the goal of data.gov.uk, technical developments and social activities of the actors within all the networks shown in Figure 7.6 led to the creation of the 'developer preview' [Andrew Stott: 112], which was a beta data.gov.uk website. This provided actors access to government data using CKAN as well as other Linked Data technologies and standards. Inscribed in this are the agendas of the different actor-networks part of this newly forming network, each responsible for different aspects to its design and purpose. The involvement of the developers prior to this also played an important part in this process, as they acted as a voice to guide the technical development:

> *"There was hundreds of developers to take data.gov.uk from an Alpha to a Beta release was significant. The fact the community did get engaged very heavily in the early specification of the system."* [Nigel Shadbolt: 84]

During this period, the 'UK Government Data Developers Google Group' formed, which provided an additional online communication forum for actors in the various actor-networks involved in the development of data.gov.uk to discuss: technical and social barriers, challenges, published datasets, and outcomes of their activities. As Figure 7.8 shows, during the developer preview of data.gov.uk, there was a substantial amount of activity, over 800 different posts were made by 277 active individuals posting within the first 3 months of the groups formation.



(a) Cumulative Number of Posts



(b) Number of Users Contributing

FIGURE 7.8: Illustrating statistics of the UK Government Data Developers Google Group based on (a) the cumulative number of posts published (b) the number of unique contributing users per month

The online communications provide a clear example of the impact of the obligatory passage points and the use of CKAN within the translation towards data.gov.uk. Examining the content of the posts from September to December 2009, the most discussed topics concerned government datasets and also Linked Data technologies and references to the functionality of CKAN, the technology that underpins data.gov.uk.

The activities of the actors attempting to achieve the goals set were also effective in producing network objects such as scholarly articles: 'Government Linked Data: A Tipping Point for the Semantic Web' (Shadbolt 2009), which was passed to other networks (including the Semantic Web and Linked Data network), demonstrating the successful activities of the UK academics. As Figure 7.9 shows, the government also responded to the passage points, increasing the number of published datasets for developers to use as part of the developer preview of data.gov.uk. The combination of all the different activities helped lock-in and strengthen the ties between the actor-networks. Effectively, the networks aligned and were undergoing translation towards data.gov.uk, producing both social and technical network artefacts.

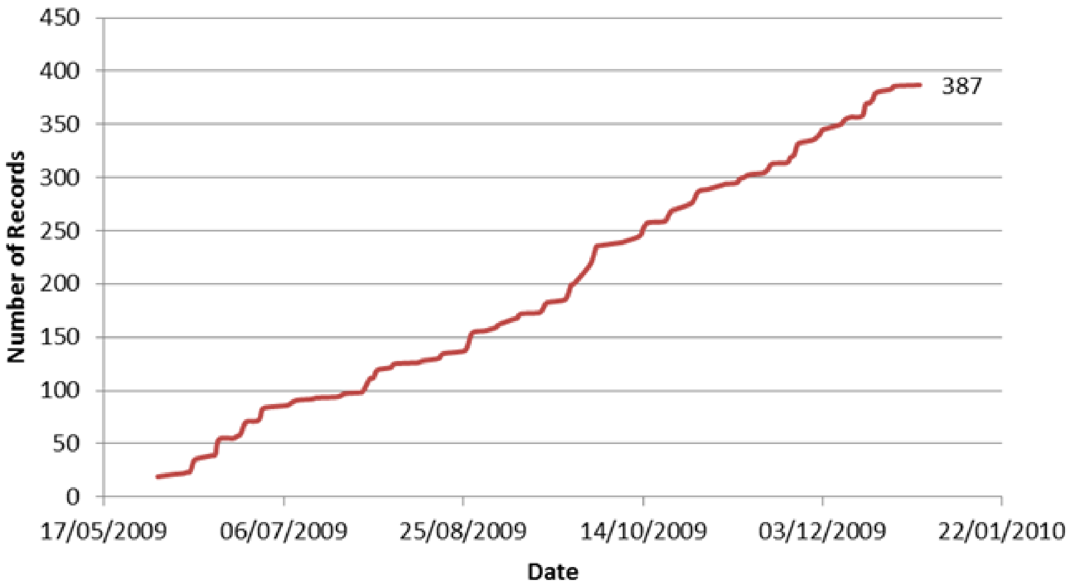


FIGURE 7.9: Published datasets available during the developer preview of data.gov.uk

In summary, discovered in this section:

- Translation is a process which occurs in multiple networks, each which have their own interests and agendas
- Various forms of socio-technical processes are required to bridge the different networks together, including using technologies and human activity
- Networks have been aligned together via obligatory passage points which can help establish common goals in order to translate

7.4 Building upon Network Artefacts

Driven by their common interest of government data, the activities of the academics, government, OKFN, and developer had led to the development of data.gov.uk. Consequently, associations and relations between networks increased, and through their alignment translation occurs, and network outcomes were produced. Technologies were supporting the publication of data from government sources, and developers started to make use of this, creating new technologies.

This section will explore how the development of data.gov.uk enabled additional actor-networks to become interested and enrolled towards the agenda of UK OGD, and how this in turn strengthened relationships of the aligned and new networks, and helped translate UK OGD towards a Web activity.

7.4.1 The Problematisation of Civic Society

Establishing the developer preview of data.gov.uk had a greater effect than just strengthening ties between the actors involved; it led to the involvement of an actor-network which interviewees have described as the "critical people in the middle" [Chris Taggart: 97]. The civic society network contains actors interested and already active in lobbying for increasing transparency between citizens, government and society. Their interests lie within both the social and technical spectrum of transparency, developing technologies and using data with a focus towards raising awareness of the benefits that access to public data and transparent government provide.

Similar to the developers, the civic society network is a distributed collection of actors, yet unlike the developers the agenda of the network was driven by a strong inscription of achieving transparency and accountability. This was partly responsible due to the associations available both online and offline. Civic society also contained actors associated with the developer, government, and OKFN network [Johnathan Gray: 28], thus providing a means to bridge between the networks:

> *"I ran around to all of my geek meet ups that I used to go to and used to just hop on the stage at the end and say 'I'm thinking of doing this government day, and in a nice way, were going to show them what could*

> *happen if data was open and what developers could do with it."* [Emma Mulqueeny: 23]

Through the bridges that the shared actors provided, the interests of the civic society actors were visible to the other networks, and in turn, civic society was exposed to the resources of published public data. Having access to such data enabled them to create Web-based applications and also become involved in discussing the issues of publishing government data:

> *"In the UK we have benefited greatly from a very active civic society and developer community, both in terms of activists and developers doing practical things when it's possible. So on the activism side, there is the Open Knowledge Foundation; there is MySociety, who have been very influential, and also Rewired State. Then we have had developers from rewired state representing, who have come in and used and requested more data"* [Andrew Stott: 179]

Whilst the developers and other actor-networks were working on data.gov.uk the civic society network were actively pursuing the right to public data access, in line with their interests towards providing citizen transparency. These activities of these actors, which were described as 'activists who love data' [Francis Irving: 30] and the 'civic hacker community' [Tim Davies: 106] complemented the agenda of the data.gov.uk network, acting as a catalyst to strengthen ties between them all:

> *"Specific projects such as mySociety, TheyWorkForYou.com and various projects from the Sunlight Foundation have helped to make a policy area which used to be intangible and difficult to understand very tangible, and made the benefits of more Open Data policies more obvious to the public."* [Jonathan Gray: 176]

> *"There were people who were doing stuff around Open Data [...] the MySociety stuff etc., and they had quite a big network of developers that this is simply around the idea of Open Data in all its forms"* [Julian Tait: 103]

> *"There's the Rewired State people, Ema Mulqueeny and so on have been quite prominent in that. And they have done quite a bit with different government departments, organising hack days and the like"* [Andrew Stott: 224]

In all three extracts, civic society actors were described as having goals aligned with the data.gov.uk actor-network (academics, developers, government, and OKFN); which included the development of policies for data access, raising public awareness of the benefits of government data, and working with developers on Open Data technologies. These common interests enabled data.gov.uk to become aware of civic society's agenda, which in turn provided them with the opportunity to become part of the UK OGD translation process.

7.4.2 Bridging Actor-Networks: Online and Offline Activities

The analysis now turns to examine the processes that enabled civic society to become aligned with data.gov.uk and the OGD agenda; and by do so, will highlight the similar socio-technical processes which enabled to this happen.

Actors within the civic society network played a critical role in breaking down the social (and technical) barriers that was faced during the translation of data.gov.uk. Through their associations with actors within data.gov.uk (developers and government), they were able to draw together actors with an interest in OGD in order to overcome some of the challenges and threats that were faced within the translation process:

> *"There was obviously a massive resistance in government, thinking that everybody is going to try and use the fact that there was such little understanding of what Open Data meant, that they [government] thought that if they open the data, and then developers will be able to get through that data and into government. [...] I had these really strange conversations with people who just had absolutely no clue on what this meant, there seemed to be this idea that opening data meant opening government up, to cyber-attacks, and that there wasn't the right security levels etc. So really what I wanted to do when I was talking to all the geeks*

> *was to show them what developers can build and then they can do the sums themselves."* [Emma Mulqueeny: 26]

These actors played a specific role in the alignment and translation of the networks, and are described as:

> *"Goveratcey, so they people that are inside and outside government, but they are highly networked, highly skilled communicators. You know they are a sort of a type."* [Emer Coleman: 168]

These "highly networked" individuals were responsible for providing an opportunity to connect different actors from different networks (developers and government) together. These actors were able to draw together the necessary actors to demonstrate the capabilities of government data, providing a new perspective on the data and what it can be used for:

> *"We scraped a whole load of data, and created our own little database of government data on the day. I think we had about 100 people turn up, it was just one day thing, but it absolutely blew everyone's minds. I invited all of the people that I knew in government, and most of them turned up, and it blew their socks off basically. It was a great day, it one was of those days that you never forget really."* [Emma Mulqueeny: 34]

Events such as this were essential to gain the alliance of the actors that were sceptics to the benefits of OGD, as these were the same actors that would be producing the data or creating OGD policies. These events helped provide a clearer understanding of the social and technical capabilities and implications of ODG, and they also became a forum to enable actors from different networks to form new ties and strengthen existing ones.

As a consequence of events like this, network outcomes were produced, which themselves become artefacts to gain the interest of more actors. In this example, the event provided the developers with the necessary data to create applications which used government data, which then helped gain the support of government actors, who are the ones responsible for more data publications. Ultimately, through the activities of civic society drawing together the developers and government, they turned a set of ideas and inscribed them into something visible and tangible, which become devices of interessement.

Reflecting on these activities and those that occurred in the previous two sections, it is possible to see the step-by-step, multi-networked yet simultaneous translation of the forming OGD community. The different actor-networks have contributed to different roles in the formation of the activities surrounding data.gov.uk, and via the alignment of these networks, various goals and agendas have been inscribed into its functioning.

The activities of the civic society network were supported by multiple forms of interactions, including online communications. Effectively this provided a way to expose their activities and gain the attention of the biggest network of actors possible. Examining the Twitter communications during this time period, actors within civic society were communicating news, information, and media helping share the relevant activities and achievements that were produced due to translation.

> *"RT @DirDigEng: UK Gov sets out pub data principles: reusable; open stds & licence; raw -¿ linked. http://bit.ly/67otOn #smartergov #opendata"* [Twitter Data Stream, 'jaggeree', 7th December 2009]

> *"RT @DirDigEng: Please help shape URI standards for the UK Public Sector http://bit.ly/avcIA #opendata #linkeddata #gov20"* [Twitter Data Stream, 'countculture', 12th December 2009]

> *"datablog from @timberners_lee and @Nigel_Shadbolt: our manifesto for government data http://bit.ly/8GwzCW #opendata"* [Twitter Data Stream, 'DataBlog', 21st January 2010]

These tweets not only represented the 'highly networked' actors, but they were also identified as the most shared tweets within the network of communications during the time period. The content and purpose of these tweets illustrate the kind of function these actors played within the communication network: sharing information between actors, and promoting the activities of the network. This is essential for the diffusion of information to larger networks of actors.

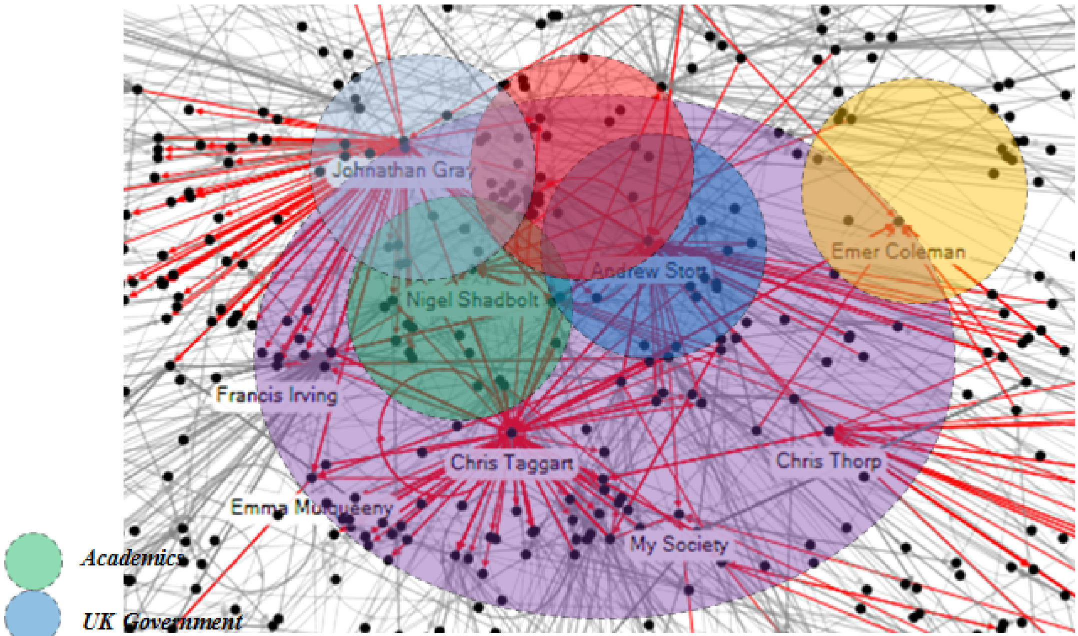
The sharing of tweets embedded with other actors and external URLs are a powerful way to discuss the activities of the network surrounding data.gov.uk whilst providing a reference to the actors responsible for this. It also provides a link (via the URL) to additional network artefacts for more information, and as part of this process, new actors can be introduced. The communications shown in Figure 7.10 reflects this; the volume of online communications had increased by over 600% within 6 months of the launch of the developer data.gov.uk preview, and within these communications, 82% of them contained URLs which pointed to either news regarding activities of the data.go.uk network, or events that were going to take place. The practice of embedding URLs within tweets were identified in both the direct conversations (mentions) and retweets, with 96% of retweets containing URLs.

Examining the communications in more detail, in comparison to the previous snapshot of communications, there had been an increase of 32% new actors who have been part of the retweet communication network, indicating that this news is attracting and spreading to a wider network of individuals. Communications from actors also extended to direct communications with the London network, which in the previous instance communications were only via the developers. However, through the activities of the civic society actors, the network was restructured, and new associations were being made between the London network and data.gov.uk.

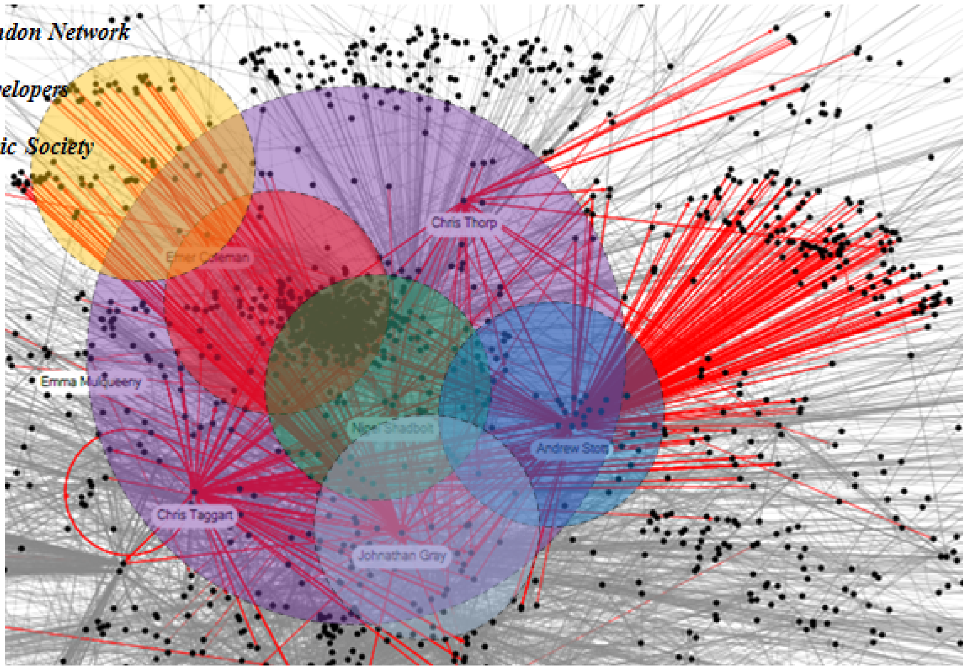
Through various forms of online and offline activities, actors within the civic society network had become aligned and were supporting the activities of data.gov.uk and the agenda of UK OGD. Their activities were also reconfiguring the data.gov.uk, creating new associations with the previously disconnected London network. However, without the activities which led to the developer preview of data.gov.uk, the alignment of civic society would have not been possible.

7.4.3 Alignment of a National Media Actor-Network

This section will explore further how the translation towards data.gov.uk helped form alliances with other UK OGD actor-networks. The analysis will explore the alignment of the national media actor-network, which represents an actor-network with the goal of producing news and journalistic articles. Driven by existing activities, it can be



(a) Twitter Mention Network



(b) Twitter Retweet Network

FIGURE 7.10: Visualisation of the Twitter (a) mention and (b) retweet network during the time period related to the events discussed during Dec 09 May 10. The highlighted edges and names represent the different actor-networks discussed Academics, OKFN, UK Government, London Network, Developers, Civic Society

considered as a network which has been undergoing a translation process before the emergence and formation of data.gov.uk.

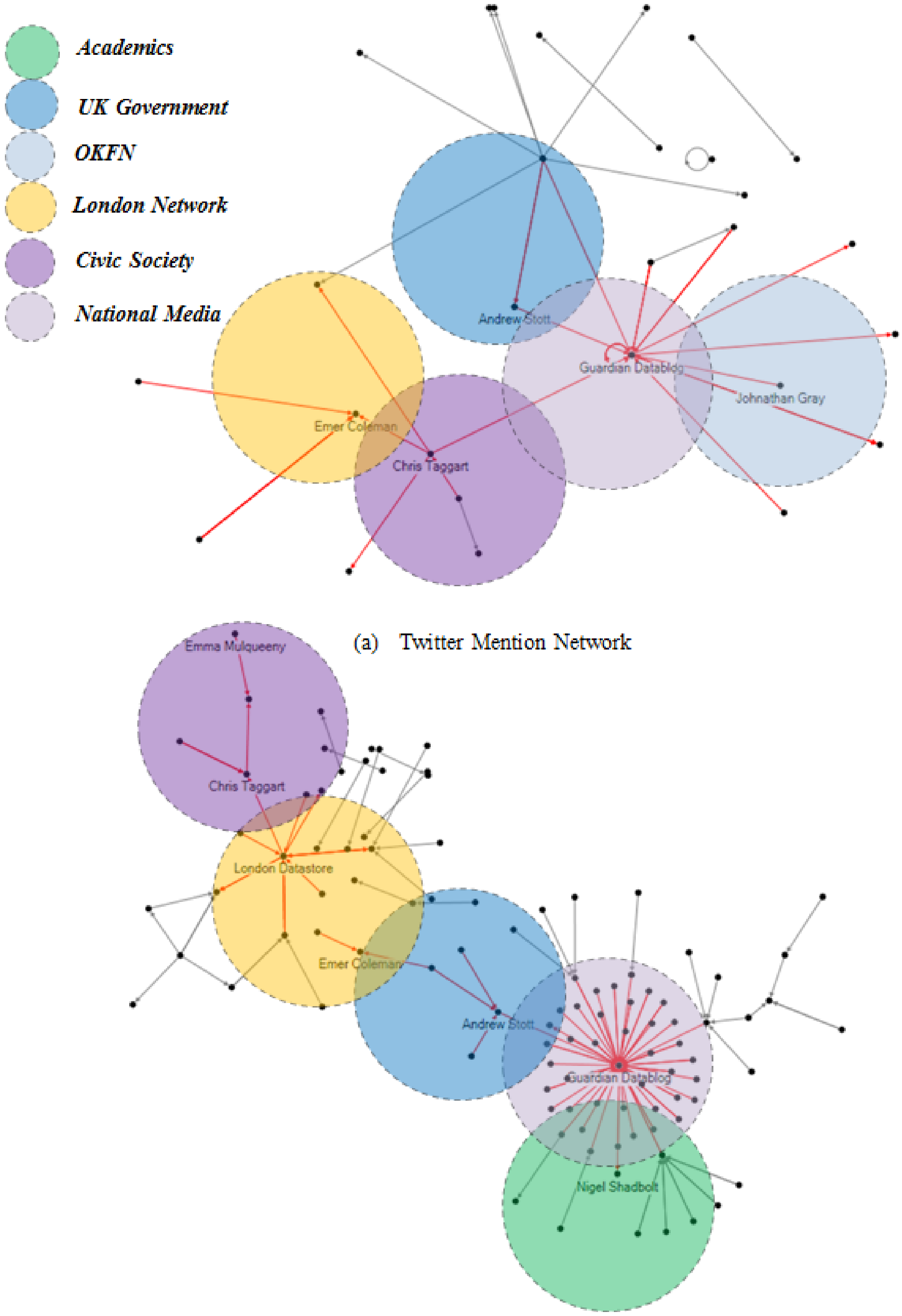
However, by considering the previous activities of actors within the national media network, it is possible to show how their alignment towards UK OGD was a result of the activities of actors within data.gov.uk, with a particular interest in the affordances of reporting government data. As a way into exploring the alignment of the national media network, the 'Guardian Datablog' has been identified (by interviewees and presence within the online communications and friendship networks) as an actor to follow.

The actors within the national media network were actively providing alternative ways to gain attention of other actors-networks using their own network outputs such as journalistic articles, which contained the activities and the requests of other OGD networks. These were used to organise events and take part in other actor-network activities, helping strengthen ties and drive the advantages of government data.

> *"The Cabinet Office has announced an early preview of the site to developers and invited them to comment. The Cabinet Office praised developers for inspiring the government to pursue the data initiative. The developer community through initiatives such as Show Us a Better Way, the Power of Information Taskforce, MySociety and Rewired State have consistently demonstrated their eagerness and abilities to 'Code a Better Country'. You have given us evidence and examples to help drive this forward within government. To get involved, the Cabinet Office invited developers to join a Google Group"* ['Guardian Datablog', 2nd October 2009]

These articles and reports were promoting the activities of the data.gov.uk network, and included communicating the launch of the developer preview and other related activities. In addition to this, articles were published and highly shared across Twitter, and as shown in Figure 7.11, were attracting may retweets from actors interested in using government data, specifically actors within the academic and government network. The sharing of news was a reciprocal relationship. Whilst actors were sharing the articles, the online communications were also being used by 'The Guardian' to share the news of these actors as well, broadcasting the communications of the academics:

> *"RT @Nigel_Shadbolt: Still much to do RT @edent: #OpenData What the UK government give with one hand... http://bit.ly/7hibhG"* [Twitter Data Stream, 'datastore', 22nd January 2010].



(a) Twitter Mention Network

(b) Twitter Retweet Network

FIGURE 7.11: Visualisation of the Twitter (a) mention and (b) retweet network of the network of communications surrounding the National Media actor-network activities, specifically the Guardian Datastore, corresponding to the time period of Dec 09 May 10. The highlighted edges and names represent the different actor-networks discussed Academics, OKFN, UK Government, London Network, Civic Society, National Media

As Figure 7.11 illustrates, the 'Guardian Datablog' functioned as a bridging actor, providing paths between a number of different actor-networks. Examining the mention communication network, the 'Guardian Datablog' connected the actor-networks of the UK Government, OKFN, and the Civic Society, including communications that allow them to act as a mediator between networks:

> *"Would be great to chat at some point about adding more #opendata from the Guardian to #ckan @datastore!"* [Twitter Data Stream, 'jwyg', 7th December 2009]

The 'Guardian Datablog' were not only operating through online means of communication; given their associations with other actor-networks, they were also responsible for setting up events to enable different actors from different networks to work towards the UK OGD agenda:

> *"We thought conferences and hack days would be good, and one of the early hack days, one of the first external ones that was done at the Guardian, [. . .] I went and told the Guardian, asking if they could provide a venue for this. We did that and it was interesting because it was confrontational in a way, as we were kind of telling government that you are really rubbish at IT, we think we could probably do better if you let us, but it was just saying that, but it was showing the sort of things that could be done."* [Chris Thorpe: 39]

This event drew together actors within the developer and civic society network, and also the UK Government and the London network, which up to this point, their interactions and communications only occurred through the mediation of technology or a shared actor. The result of these activities helped establish new associations and strengthen existing ones, which enabled translation to proceed, where new actors were aligned and enrolled to improve data.gov.uk.

The hackathons organised and ran by 'the Guardian Datablog' produced artefacts from the interactions of multiple actors-networks, including actors within the civic society, national media and developer networks. In turn the product of these events worked as interessement devices for data.gov.uk, which helped strengthen the ties with actors in government:

> *"We had a number of developers in the guardian building up in kings cross for 2 or 3 days developing some proof of concept style applications, doing famously the post code paper, which showed what could be done by pulling data from different government data sources, and showing that the product didn't have to be a website."* [Andrew Stott: 187]

Specific artefacts, such as the 'postcode paper' demonstrated how joint activities between actors in different networks could be successful. The production of this had many consequences: from providing an opportunity for developers to show their capabilities, to enabling government to understand the capabilities that publishing government data has:

> *"You could use the data to make a physical paper, which was great for the cabinet office meetings and the like, as no screens allowed in there. It also showed how important the post code data was."* [Andrew Stott: 187]

As Figure 7.12 illustrates, the network artefacts, and online and offline events became devices of interessement, mediated through the relationships of actors within the civic society and national media network to reaffirm the goals of data.gov.uk. The 'postcode paper' artefact worked as a driver to allow translation of data.gov.uk to proceed, whilst at the same time, it offered competitive incentives for the London Network to translate towards their own goals.

During this time, the London network was undergoing a process of translation led by an agenda not part of data.gov.uk, even though they were translating towards a similar original problematisation of openness and transparency. Although these two networks were not autonomous or working exclusive to each other as they share common actors, the decisions made by the actors involved during the earlier stages of translation has led them to approach OGD in different ways, both technologically and socially.

> *"we knew around the data.gov.uk stuff, [...] pushing for the linked data agenda, and I took the decision not to go that route as it is just one further barrier".* [Emer Coleman: 218]

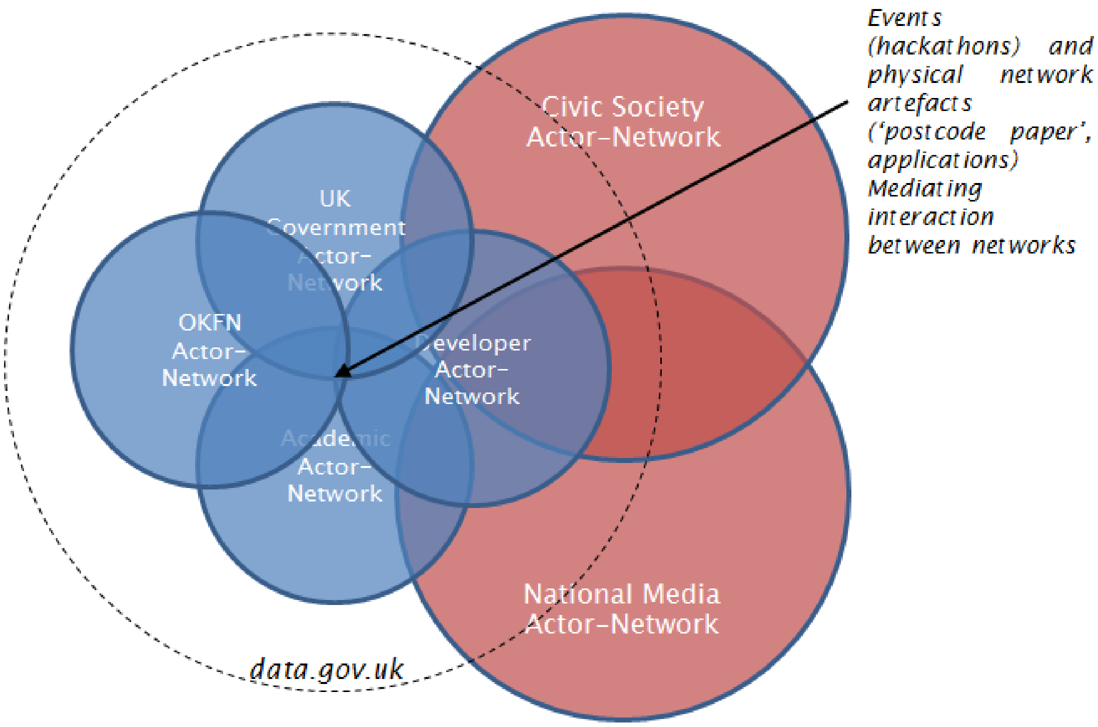


Figure 7.12: Illustrating the strengthening of ties between actor-networks via the use of online and offline events, and physical network artefacts passed through the networks via the mediated process of the Civic Society and National Media Actor-Networks

> *"we could have just gone down the data.gov.uk route, and put our data in there. [. . .] I wasn't convinced that they would necessarily go the same route as we wanted, which was to say, we will publish everything."* [Emer Coleman: 40]

The activities that drove the translation of the data.gov.uk and London network led to the public launch of the London Datastore on January 7th 2010, and the public launch (based on the developer preview) of data.gov.uk beta on the 21st of January 2010. Both launches were considered as "really key events in the whole thing of data.gov.uk" [Chris Gutteridge: 67], not only because they were underpinned by different technologies and policies, but as they were the first public facing announcements of the commitment towards UK OGD.

7.4.4 The Translation of data.gov.uk and the London Datastore

Competition between networks is often described as a threat to network translation and stabilisation (Callon 1986*a*) (Latour 2005). Yet despite the competition between the actors in the data.gov.uk and London network, both networks managed to successfully pass through their passage points.

The competitive activities of the different actors were important for driving their own translations process; and the competition for achieving a public launch first helped speed up both networks translation [Emer Coleman: 40]. Aiding and supporting this process of competition were network artefacts produced by other actor-networks, reporting on the current state of success and translation that actors within the networks had achieved:

> *"Today marks another step along the path: making the 'beta' of data.gov.uk available to everyone, whether in the wider community of interest or one of our registered developers"* ['Guardian Datablog', 'Our manifesto for government data', 21st January 2010]

> *"Is the data open? Though they don't use a license or legal tool to make the data open, their Terms and Conditions appear to make the data open as in the Open Knowledge Definition. Nevertheless it would be good if they made this more explicit by using a legal tool such as the PDDL, ODbL, or CC0!"* ['OKFN Blog', 'New Open Data from London Datastore', 11th January 2010]

Whilst these artefacts helped the translation, they also strengthened the ties between existing actor-network, such as the ties between the developer and data.gov.uk actors, or the London and OKFN network. Actions like this help strengthen the associations between networks, allow translation to proceed, which form competition between actors.

As Figure 7.13 shows, a large number of datasets were published three weeks prior to public launch of data.gov.uk, the amount of datasets released during their launch provided an incentive for developers and those interested in examining the data to continue working within the network. In comparison to this, the London Datastore took an alternative approach and incrementally released datasets after its launch, and although fewer datasets were made available, this reflects the inscriptions made during the early stages of London's translation.

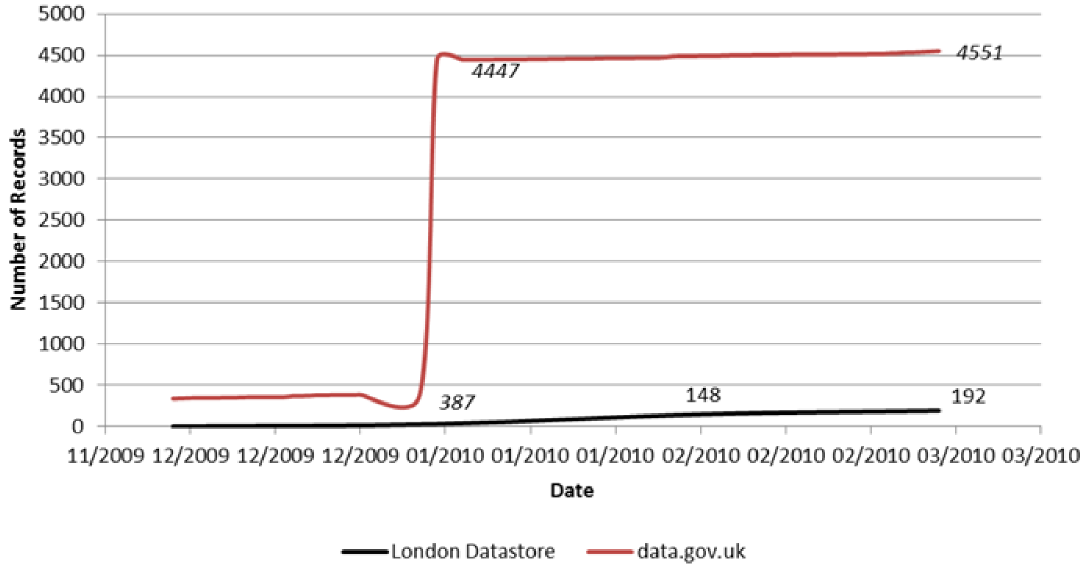


FIGURE 7.13: Total Number of datasets available in data.gov.uk and the London Datastore before, during and subsequently after their public launch

Despite the different approaches to data publication, both data.gov.uk and the London network had aligned and enrolled actor-networks in order to successfully translate and achieve their goals. This success, as described, was uncharacteristic of government projects, which typically fall behind schedule, and overspend:

> *"historically over specified and have run on for vast of amounts of money and time and always go over budget and underperform"* [Chris Gutteridge: 69]

However, as the interviewee later described, their success was down to their interactions and activities with the required actors, and the adoption and development of technologies and policies to support the OGD agenda. By attracting and aligning actors that had the necessary skills, using capable technologies, and also competition between networks, it was possible to achieve their goals using the resources available:

> *"data.gov.uk was very much thought out from first principles, it was scoped as what could be done in 6 months, and we had the plan and put it together,*

> *[...] we needed to do something pretty quickly, and knew it would be fairly high profile in the community, so we didn't want to go horribly wrong. Also, we didn't have any money at that stage, and doing this in a classical, Government IT project kind of way, would have meant that after 6 months, we wouldn't have even finished writing the requirements, let alone done the yearlong tendering process"* [Andrew Stott: 110]

Contributing to the success of data.gov.uk and the London datastore was the structure of the actor-networks, the 'top down, bottom up, middle out' [Andrew Stott: para. 170] formation of actors was not generally the accepted methods used for projects involved with the UK Government; yet this structure produced by the alignment of a distributed set of actor-networks, with their own interests in OGD.

> *"It's had very strong political backing by both the current prime minister and previous prime minister, and in both cases been part of their political zeitgeist [...] We have benefited greatly from a very active civic society and developer community, both in terms of activists and developers doing practical things when it's possible. [...] So on the activism side, there is the Open Knowledge Foundation; there is MySociety, who have been very influential, and also Rewired State.* [Andrew Stott: 180:184]

The translation of data.gov.uk and the London actor-network is the result of gaining the support of new actors-networks by enticing them to join the network by using interessement devices, and also by enabling their own agendas to become a passage point to pass through. However, this acted as a double-edged sword; it may have enabled new actors to become enrolled to the overall goals of publishing data, but the original relationships between the Semantic Web and Linked Data networks were weakened.

Emerging here is a process of how new networks form. In order for new activity to emerge, previous networks must remain in translation and not lose sight of their original goals. In this case, the academics had to continue to do what academics do: research, increase knowledge, and publish papers, the government had to continue to create policy and legislation, national decisions, and offer financial support, and the developers had to continue to create software and new technologies.

During the emergence of data.gov.uk and the London Datastore, there has been a shift from a loosely connected set of actors and actor-networks, connected via online communication, and offline activities towards the current state which is the result of physical interactions, technological developments, and alliances and competition between actors from various networks.

However, it is important to note that the agency of some actors may have been hidden from the analytical lens. For example, the quote above describes how the UK Government had a strong backing and drive towards the agenda of UK OGD; these actions and activities had an impact on the growth of the Web activity, however they were not captured as they were either not engaging with the same technologies as other actors, or were interactions that were hidden from the observational lens. However, whilst these unseen actors (e.g. the UK Prime Minister) and activities may not be captured by direct observation, it is possible to expose the outcome of these activities and actor interactions and expose their effects and impact on the Web activity, which is the primary purpose of the analysis.

In summary, discovered in this section:

- How the translation of a set of actor-networks driven by a common goal provides a platform for other networks to build upon, and become involved
- By new networks becoming involved with the already translating network, the goals that were previously defined adapt and change with the interests of the newly aligned actor-network
- The alignment of actor-networks is achieved through online and offline activities and communications. Used together they provide a method to attract networks of actors that were not originally identified as part of the agenda
- The role that competition between networks and actors have towards enabling translation to proceed, and its relative effects on the speed of translation
- The importance of the network structure as a result of the activities and inclusion of specific actor-networks, and its relation to the speed to which translation can be achieved

7.4.5 The Emerging Web Activities of UK Open Government Data

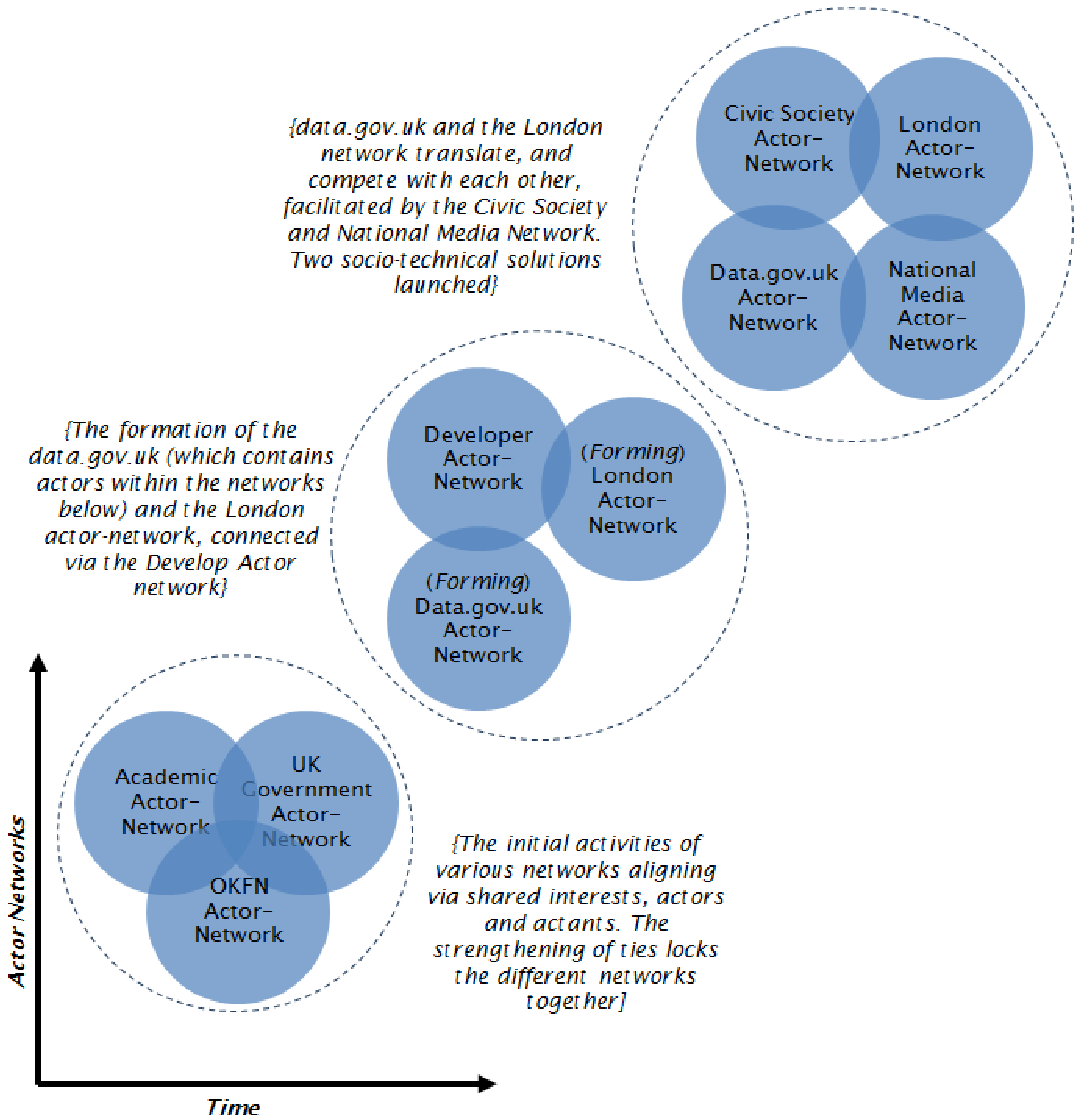


FIGURE 7.14: An overview of the development of data.gov.uk and the London Datastore, plotted (not to scale) against the time and number of actor-networks. This is a representation of how the different actor-networks involved have come together to enable a translation process

By establishing common goals that all the actors could agree on, the different networks involved enabled actors to become aligned and enrolled, and as Figure 7.14 illustrates, for new actor-networks to emerge. The formation of data.gov.uk, and subsequently, London datastore is an illustration of two substantial socio-technical network artefacts that have resulted from the network activities. However, these artefacts require the commitment of not only the current activities that enabled them to emerge, but also the stability of previous networks.

This chapter has demonstrated how the problematisation of an idea, shared across multiple actor-networks can emerge into a network of translation, driven by passage points, and connected via shared actors. Moreover, the ANT-driven observational lens unpacked how social and technical processes were entangled within the growth and structure of the emerging artefacts of data.gov.uk.

Based on these discoveries, the next chapter will explore how the network of activities grows and establishes itself as a Web activity known as Open Government Data.

Chapter 8

Establishing UK Open Government Data

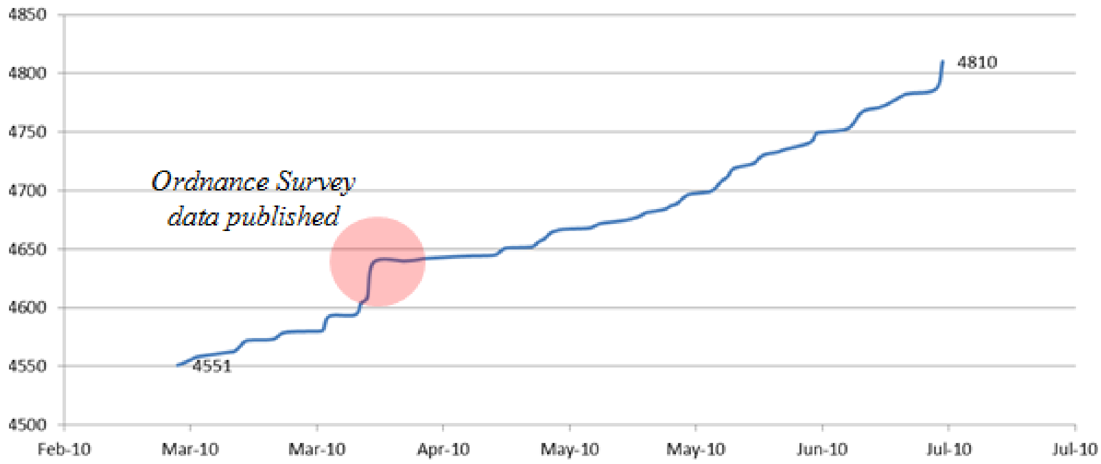
8.1 Expanding UK Open Government Data

The previous chapter exposed the activities of actors that led to the emergence of the Open Government Data Web activity. This chapter will now examine how the activities and commitment of the actors responsible for the UK OGD Web activity have provided a platform for further adoption of Government Data and for new types of Web activities to emerge.

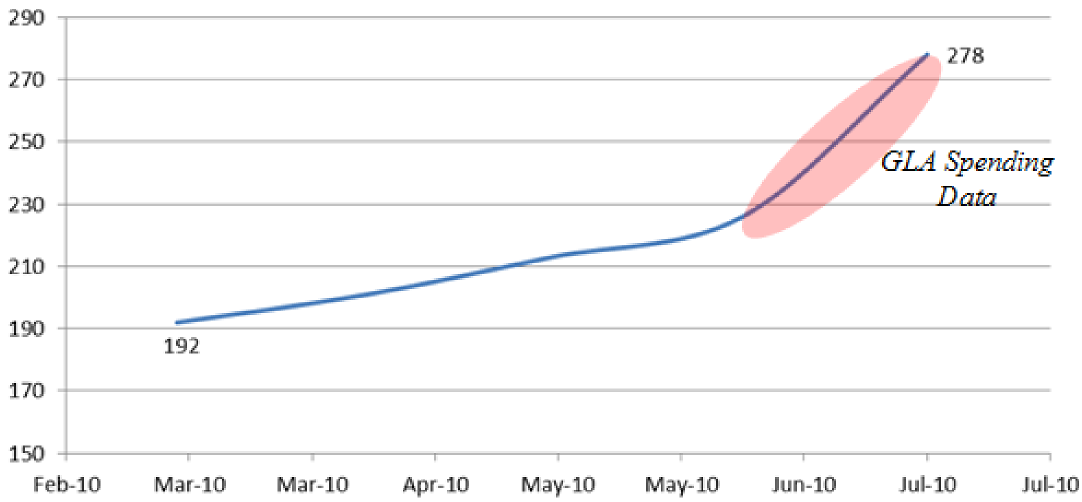
Given the emergence of data.gov.uk and the London Datastore, the following section will examine how the production of different network artefacts and additional socio-technical processes helped strengthen the associations of existing networks and gain the alignment of new actor-networks.

8.1.1 The Activities of Enrolled Actor-Networks

From an initial set of distributed heterogeneous networks of actors, a common goal was established, and new actor-networks emerged and produced data.gov.uk and the London Datastore. The translation of these socio-technical systems were inscribed with the interests and needs of the actor-networks involved in their production, and through their commitment and stability, new actors became engaged with the translating technological and social activities of opening up government data.



(a) data.gov.uk – Available datasets published between March and July 2010



(b) London Datastore – Available datasets published between March and July 2010

Figure 8.1: Number of datasets published between March and July 2010 in both data.gov.uk and the London Datastore

As the activities and commitment of actor-networks continues, the growth of OGD continues to occur; online and offline. As shown in Figure 8.1(a), the number of available published datasets continues to increase, and in addition to this, specific datasets were published:

> *"I'm delighted that the Ordnance Survey is releasing this data for free re-use. It will help people make fuller use of other government data on*

> *data.gov.uk, as well as stimulating innovation in mapping itself."* [Guardian Datablog, 'Ordnance Survey launches free downloadable maps', 1st April 2010]

Articles were also written and shared using Twitter in order to promote the activities of the publication of more government data on data.gov.uk. This demonstrated the commitment of government and the capabilities of the technologies that underpinned data.gov.uk:

> *"Confirmed that some Ordnance Survey data will be free for reuse without "derived data" hassle from April 1 #bbdf #superfast #opendata"* [Twitter Data Stream, 'guardiantech', 22nd March 2010]

> *"RT @guardiantech: Ordnance Survey launches free downloadable maps http://bit.ly/bu37pc #opendata"* [Twitter Data Stream, 'McGarryConsult', 1st April 2010]

These tweets were in the top 5% of shared messages published between March and July 2010. They had been retweeted by 63% of actors who were new to the online communications during this time period; this being actors who were not identified within the #opendata communication channel before March 2010. As suggested, the publication of new data was a "really clear advance" [Francis Irving: 53] for data.gov.uk, helping demonstrate actor commitment and enabling translation to continue. The availability of new kinds of data became a driver for developers and other actors to create new technologies, which was a "key milestone" [Nigel Shadbolt: 89] for OGD. Government were publishing their data for free, providing access to datasets which were of extremely high value:

> *"you have royal mail, upwards of £5000 to get address information for postal code information. Some of it is free, some of it is open, but a lot of it is isn't."* [Malcolm Barclay: 122]

As Figure 8.1 shows, the publication of OS's data was just an initial step in government illustrating their commitment towards the UK OGD agenda. Between January 2010

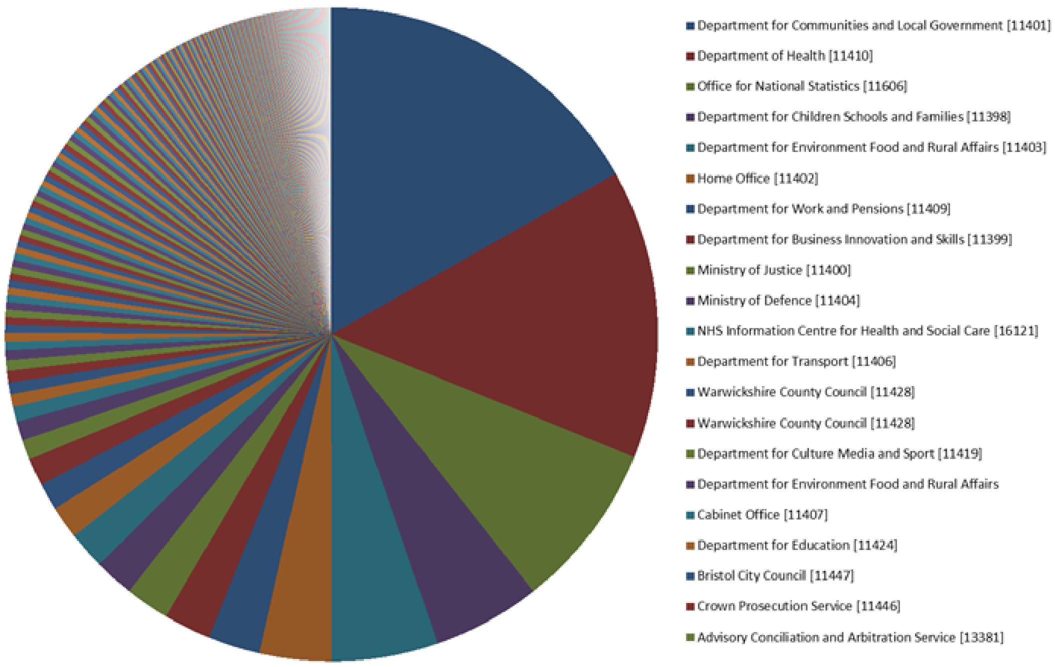


FIGURE 8.2: Published datasets by 336 different Government labelled groups in data.gov.uk between the public launch and July 2010

and July 2010, the number of datasets published within data.gov.uk and the London Datastore continued to increase, and as Figure 8.2 illustrates, there were 336 different government groups responsible for the data published, each contributing to different types and quantities of data.

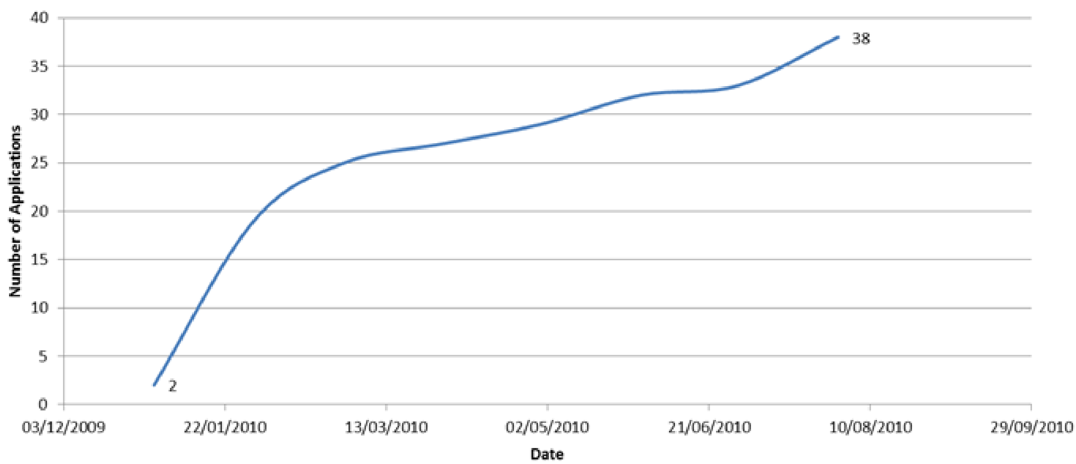


FIGURE 8.3: Number of Web applications and services published on the data.gov.uk data portal between the public launch in January 2010 to July 2010

By publishing more data, it provided the developers with the necessary resources to achieve their own agenda. Developers were able to create applications and services which utilised the data. As Figure 8.3 shows, between the public launch of data.gov.uk in January 2010 to July 2010, there was an increase in the number of applications published; 36 different applications were produced by 34 different actors associated within the developer network. The developers were using a range of different datasets to create applications, and by doing so demonstrated their commitment and technical capability, but also showing that the efforts gone into publishing the data was not wasted. The publication of more data did not just confirm that the government were adhering to the policies of publishing data, but it proved to be a valuable asset for other networks to build upon. Take for example the publication of the COINS (Combined Online Information System) dataset, which contains UK treasury information:

> *"Today we have lifted that veil of secrecy by releasing detailed spending figures dating back to 2008. We will not stop here we plan to release more data in the coming months that will be easier for the general public to understand"* [data.gov.uk blog, 'Finance data (COINS) goes live', 6th June 2010]

Whilst this demonstrated the Government commitment which included a letter from the Prime Minister regarding this as an important milestone reached, [PM Letter, 'Letter to Government departments on opening up data', 31st May 2010], it also helped OKFN achieve their own goals of accessing such data sources, overcoming their failed attempts:

> *"When we first asked for the COINS data (which is the UK's most detailed database of government spending) - we couldn't get a copy of the data [...] We finally received a huge slab of paper in the post, which is obviously not very useful when it comes to exploring the data on a computer."*
> [Johnathan Gray: 244]

The alignment of actor-networks to form data.gov.uk demonstrated how the formation of different actor-networks along with their own agendas can enable network translation. The use of online communications helped support the publication of datasets, being used by the actors to a tool to announce and discuss the relevant successes of published data and by embedding extra hashtags such as '#COINS' and '#freeourdata' promoted the activity of UK OGD to a wider community of actors:

> *"Gov spending database (COINS) now live on data.gov.uk http://data.gov.uk/dataset/coins #opendata" [Twitter Data Stream, 'Rchards' 4th June 2010]*

> *"Congrats to the Guardian team http://bit.ly/b8XQGB - its a hard mother lode to mine - but great start #opendata #freeourdata #COINS"* [Twitter Data Stream, 'Nigel_Shadbolt' 10th June 2010]

Figure 8.4 represents the Twitter communications and sharing of content during the publication of the COINS dataset. As illustrated, whilst associations continued to be made between the previously identified actor-networks within data.gov.uk, the use of additional hashtags helped draw upon actors that were not part of the earlier online communications. The number of individuals participating in the communications had increased by 35%, of which 46% had been using the hashtags '#COINS' or '#freeourdata', indicating that the adoption of them helped draw upon a wider network of actors.

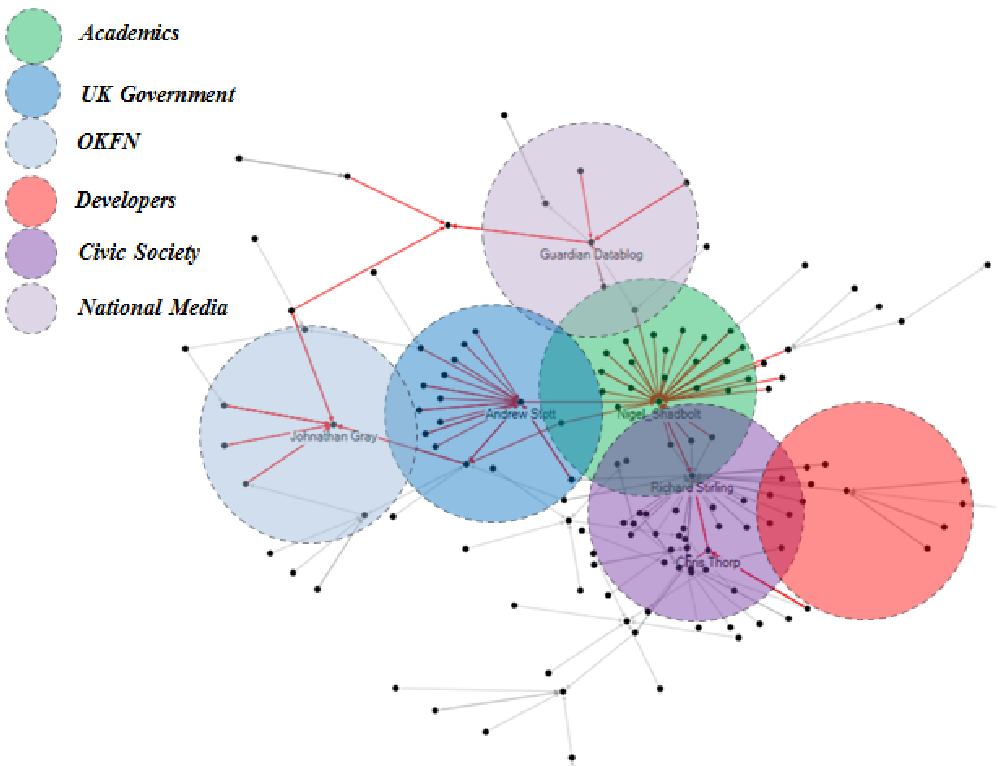


Figure 8.4: Illustrating the Twitter Retweet network based on tweets mentioning the string 'COINS' during the time period of March to July 2010

However, online communications were not the only method used to engage with a more actors interested in government data. The academics used other methods of engagement in order to announce new government practices of publishing data and the availability of datasets. Presentations and talks were given (Shadbolt 2010*a*) (Shadbolt 2010*b*), providing a forum for public engagement and response.

Offline activities were not only used as a broadcast mechanism to show their commitment, they were also used to gain feedback and offer support to actors that had interests with using the data. Seminars and workshops were given providing details on the technical and social processes that are involved in publishing government data. These seminars were aimed at providing tuition for using the published data efficiently and were intended for individuals who were not familiar with the processes involved:

> *"Developers and others that are using the released data to generate systems or other on-line capabilities to allow the public to access the released data"*
> [data.gov.uk Team Blog, 'COINS Coding A Seminar', 21st June 2010].

Whilst these seminars and events were ran to promote the activities of OGD, they were also providing mechanisms for actors who had already engaged with the data to provide feedback of the processes involved in using the specific technologies:

> *"I thought you said transparency. It turns out I need a degree in computer science to even open it up to have a look. Is this a new way of cover up?"*
> [data.gov.uk Team Blog, 'Finance data (COINS) goes live', 6th June 2010]

As part of the co-constructed translation of CKAN, OKFN, and in response to the problems identified, additional support and documentation for developers and 'data users' were given, which included the introduction of new file formats to support the range of datasets being published in data.gov.uk ['OKFN Blog', 'CKAN releases Drupal module for developers', 21st August 2010]. Due to the responses and changes towards technical implementation and the social engagement, the agenda of OGD was adapting to the involvement of new actors and unexpected forms of feedback and uses of the data. This is an illustration of the co-constructive relationship of adoption and change, both technically; through new data formats and changes to technological design, and also social processes and forms of interaction.

The analysis will now examine how the changes in the agenda and the production of network artefacts led to the emergence of a new actor-network of enrolled actors.

8.1.2 The Emergence of the Public Transparency Board

Driven by the activities and commitment of the actors within the networks associated with the development of data.gov.uk and the London Datastore, a new agenda was forming. Inscribed within this agenda was a letter from actors within the UK Government which sets out a new set of goals towards widening the reach of the UK OGD Web activity:

> *"To oversee the implementation of our transparency commitments, a Public Sector Transparency Board will be established in the Cabinet Office, which will be chaired by the Minister for the Cabinet Office Francis Maude. Board representation will include a mix of external experts and data users, and public sector data specialists; [...] The Board will provide support to departments as they deliver on the Government's transparency commitments set out in this letter. The Board will also be responsible for setting open data standards across the public sector, publishing further datasets on the basis of public demand, and in conjunction with the Ministry of Justice will further develop the Right to Data and advise on its implementation."* ['Prime Minister Letter', 'Letter to Government departments on opening up data', 31st May 2010]

The Public Transparency Board (PTB) emerged from the inscription of different agendas from actors already translating towards OGD; including actors from the government, academics, civic society, OKFN, developers, and the national media networks. The letter containing the inscriptions of the government was a way to make clear their intentions, and help enrol actors towards achieving their goal. It also reinforced the commitment of actors towards providing openness and transparency. This letter not only demonstrated the commitment of the UK government towards openness and transparency, it reshaped the UK OGD agenda, drawing upon the local government organisation structure within the UK; this addition to the agenda was promoted under the name of "Local Government Spending Transparency":

> *"New items of local government spending over £500 to be published on a council-by-council basis from January 2011 [...] New local government*

> *contracts and tender documents for expenditure over £500 to be published in full from January 2011"* ['Prime Minister Letter', 'Letter to Government departments on opening up data', 31st May 2010]

The translation towards enrolling local government was further reinforced by statements from the academic network which were presented at the first PTB meeting. These statements supported the government's plans to widen the agenda of UK OGD to align and enable local government to publish their data as well as align other networks which had pools of unpublished data:

> *"Transparency of data is not just an issue about central government. Local Authorities, the National Health Service, schools and police all have useful data which should be public. There is important work to do here to ensure that the same standards and energy we are seeing in central government are embraced in local government and elsewhere."* ['Letter from Nigel Shadbolt to Francis Maude', 'Transparency and Open Data: Next Steps', 14th June 2012]

This enabled the academic network to continue the pursuit of their own agenda of Linked Data technologies (formats, protocols, data publishing standards), and also the use of open licences in order to allow for free, reusable data sources. Effectively, this letter enabled the academic network to situate itself as one of the indispensable focal actors within the PTB which made their agenda part of the PTB agenda. By means of becoming a focal actor in the network, the academics were able to set passage points for the PTB as well as demonstrate their activities to the international network of OGD:

> *"We must promote and support the development and application of open, linked data standards for public data, including the development of appropriate skills in the public services. [...] This is exactly what is needed for the data that Government is publishing and this is exactly why Tim and I have been promoting this approach and others (in particular the US) are now following [...] There is a wider context of course. We should work with the Government to engage with the leading experts internationally working on transparency, public data and standards, and to promote*

> *international liaison and global standards setting."* ['Letter from Nigel Shadbolt to Francis Maude', 'Transparency and Open Data: Next Steps', 14th June 2012]

The academics involvement did more than support the government's interest towards publishing local government data, as Figure 8.5 illustrates, the agenda enabled UK OGD to demonstrate to the international OGD community its achievements, and unlike before, UK OGD had now undergone translation towards a Web activity of enrolled actors who are producing outcomes.

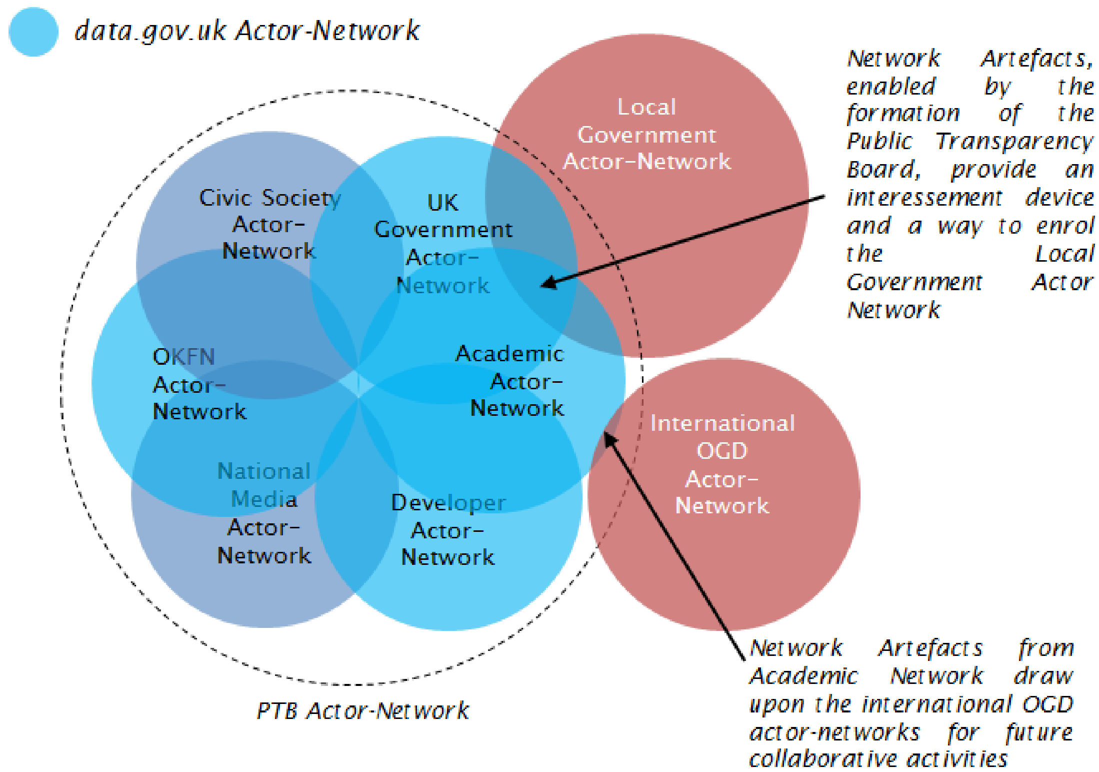


FIGURE 8.5: Illustrating the formation of the Public Transparency Board (PTB) Actor-Network and the use of network artefacts to interest and gain the enrolment of the Local Government Actor-Network

The agenda of UK OGD had re-problematised; the production of data, applications, and the continued support of actors led to the formation of the PTB, which had a common agenda of getting local government to publish data. The following section will explore the socio-technical processes involved to achieve this.

8.1.3 The Interessement of Local Government

The alignment of local government begins by conceptualising it as a network that had already undergone a translation process with an agenda formed by exiting local council and local authority legislation. The alignment of local government with the translation of UK OGD relies on a complex agenda; partly set out by central government and partly by their own agenda. However, unlike the alignment of networks previously discussed, associations between central government and local government already existed through shared actors and network artefacts such as central government acts, legislation, and policies.

Achieving the alignment of local government was considered as much as a social as a technical process which required the support of the already enrolled actor-networks producing and using the published data. The PTB meetings provided a way for aligning the activities of local government to the goals of OGD. These meetings were used to help teach the cultural and technical practices required for local government to publish their data:

> *"The work of the Department for Communities and Local Government and the Local Data Panel to deliver transparency and open data was discussed and includes Ensuring local authorities and other bodies are part of the journey on transparency, complying with the spirit of open data and engaging in dialogue with developers and others about how best to move forward in a process of continuous improvement; Discussions around how to standardise local authority data [. . .] encouraging citizens, developers, public and private sector organisations to access and use local data and information."* ['PTB minutes' 'Transparency Board Minutes', 15th September 2010]

The discussions between actors within the PTB helped establish a suitable working environment to help teach the cultural practices and technical infrastructure needed for local government to become an enrolled actor-network. Multiple networks within the PTB supported this process including civic society who helped local government expose local governments and councils who were not conforming to central government's plans to publish local data.

> *"What I wanted to look at was local data, specially the local pension funds, and then I found a problem, which a lot of the other people that got into this - the open data experience - encountered. They want to do something, and the data that should be available wasn't available. In my case I wanted to look at local authority pension funds, not [. . .] I discovered that this type of data was really hard to get hold of, it was scattered around in odd places, as PDF's. [. . .] I started to try and organise it and you think 'let me get the list of the councils, and work out which councils belong to which pension funds and so on.' I soon discovered that there wasn't even a list of councils available."* [Chis Taggart: 25]

The activities of civic society actors to support the PTB agenda of publishing local government data led to the development of OpenlyLocal, a Web service (website) that listed local governments which were and were not publishing their data. OpenlyLocal contained the inscriptions of the interests and activities of the civic society and built upon the previous activities of data.gov.uk and the PTB network. OpenlyLocal offered a technological platform for individuals to engage and find information that was initially published in "horrid excel spread sheet with loads of macros" [Chris Taggart: 53]. The technology was simplifying the social activities needed to view and made judgement of the data, and therefore became incentives to change cultural and social practices. In addition to this, combing the exposing capabilities of OpenlyLocal with communication technologies such as Twitter provided a way to broadcast the findings of OpenlyLocal to the #opendata network. This exposed both the council under question, as well as OpenlyLocal as a platform for UK OGD announcements:

> *"Council Spending data now added to OpenlyLocal.com http://bit.ly/aRgqFD #opendata #localgov #spending"* [Twitter Data Stream, 'OpenlyLocal', 18th June 2010]

> *"East Sussex County Council has been added to OpenlyLocal #localgov #opendata http://bit.ly/berDr4"* [Twitter Data Stream, 'OpenlyLocal', 7th July 2010]

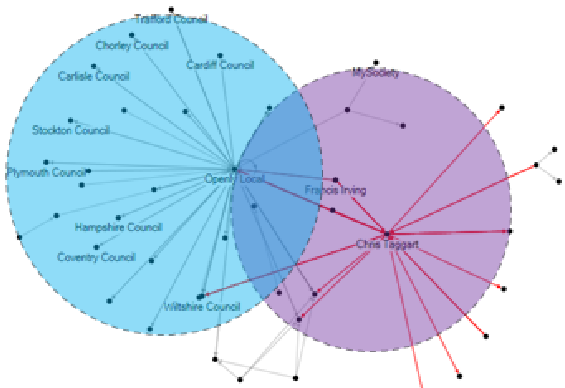
> *"More #localgov spending data +d 2 @OpenlyLocal http://bit.ly/bxAySQ . Now 24 Councils 29875 Suppliers 148090 payments all #opendata"*
> [Twitter Data Stream, 'OpenlyLocal', 24th August 2010]

As the tweets above and Figure 8.6(a) illustrate the activities of local government were supported by actors communicating, which included the URLs link to particular datasets. This helped local government actors who were using the #opendata channel to identify other actors who had published data and also enticed those who had not published their data (or made it obvious that they had published their data) to align with the OGD agenda:

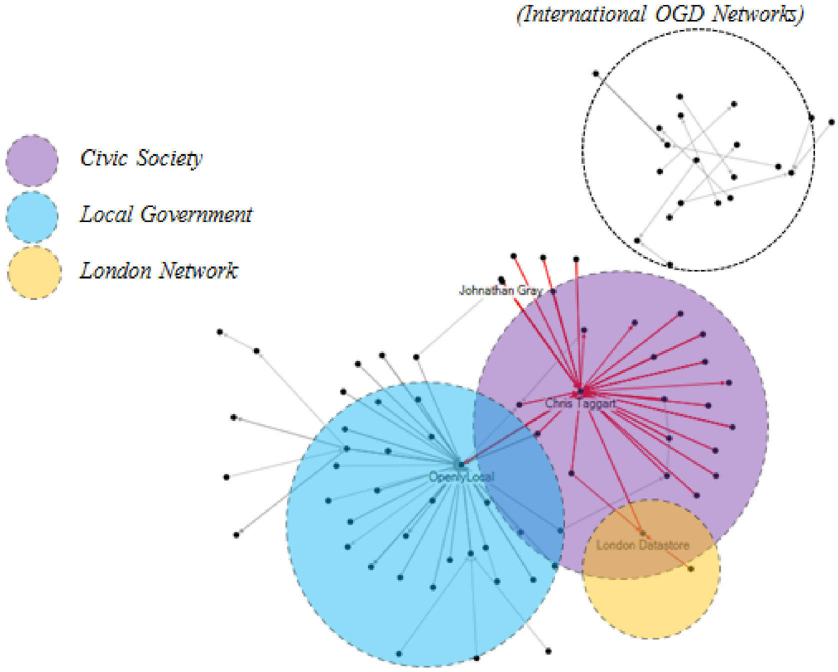
> *"a couple of councils asked if they can they be included in it [OpenlyLocal] and the Cabinet Office said 'we like what you're doing, would you like to come and sit on this panel advising us about opening up local data'."* [Chis Taggart: 46]

The enrolment of local government was partly due to different social and technical incentives provided by other actor-networks in the PTB; it was also due to the local government actor's commitment and acceptance to introduce new social and technical practices. The translation of local government was a co-constructive relationship of central government's agenda, and local government's willingness to change. Unlike previous central government strategies [Chris Taggart: 24], local government actors such as Warwickshire County Council had been active in developing the necessary technological and social solution to publish their data. As part of this, these networks had established their own set of goals and requirements in order for successful translation to occur. These goals which included defining formats and standards, building websites, and citizen engagement, were aligned with the OGD agenda, thus achieving them would affect all other networks.

> *"In the last month or so we have shamelessly used the iPhone project as a way of kicking off the slightly less sexy, but ultimately much more fulfilling work on open data [. . .] The main outcome I want us to achieve is establishing a web presence to provide access to the data as well as a framework of standards and processes for getting the data up there, making*



(a) Illustrating the Twitter mention network based on tweets mentioning the string 'OpenlyLocal' during the time period of June 2010 to December 2010



(b) Illustrating the Twitter retweet network based on tweets mentioning the string 'OpenlyLocal' during the time period of June 2010 to December 2010

FIGURE 8.6: Visualisation of the Twitter (a) mention and (b) retweet network of the network of communications containing the string 'OpenlyLocal' during June Dec 2010

sure it is in the right formats and ensuring that it is properly maintained."

['ICT Strategy at Warwickshire County Council Blog', 'Open Data: Building the tools to help you Hack Warwickshire', 26th February 2010]

Driven by multiple agendas and supported by many actors, the translation of local government actors was successful, and in a short period of time, technological platforms were produced, and data was published. Take for example Warwickshire County Council

(WCC), due to the support of government, developers and available technologies, a prototype data portal was produced, and within a month, the launch of the WCC open data website was announced and made public:

> *"As of now the open data site is available at http://opendata.warwickshire.gov.uk. the site will be officially launched next week but if you want to kick the tyres and check out the data that is already available please feel free. Any WCC people may find that this URL doesn't work properly within our network don't panic we will be sending round an alternative address via other means. [...] Currently there is a small collection of data sets available, but by next week there will be more information available and in more formats. We will also be gearing up the community side of the site and have already set up a new twitter feed to help keep the wider world informed @wccopendata."* ['Warwickshire Open Data Project Blog', 'Warwickshire Open Data Site is Now Live', 16th April 2010]

As the extract from the blog above explains and Figure 8.7 illustrates, the launch of the WCC data site included an initial spike in activity in terms of the publication of data, with datasets released during the first month of the data portals launch, then subsequent instalments of datasets accompanied by blog posts providing detail on the actors' activities until July 2010.

Warwickshire's was not the only actor within the Local Government network active in translating towards the PTB's passage point of 'openness and transparency'. Other councils were also publishing data, and similar to Warwickshire, they had created their own technical platforms and social processes:

> *"It's been a rollercoaster ride in the world of government data over the last 12 months, and, thanks to a change of government, it looks to be getting more exciting with David Cameron's recent letter to government departments urging them to open up more of their data. [...] we've launched our brand new Open Data section, allowing you to search for datasets, view data by category and also see what other people have built with our data. In the next few months, we're looking at opening up even*

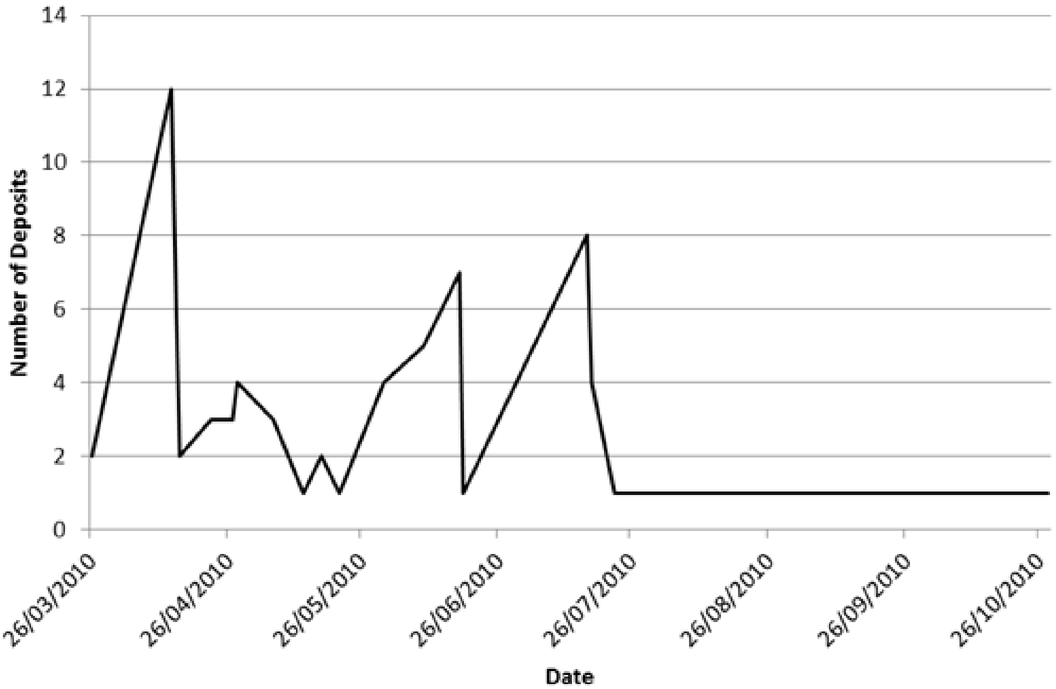


FIGURE 8.7: Dataset deposits within the Warwickshire Open Data Website Between March and December 2010

> *more datasets, as well as organising a hack day for local developers to get together and build a useful application (or applications!) in one day."*
> ['Lichfield District Council Blog', 'New Data site Launched', 11th June 2010]

As this blog describes, networks such as the developers were identified as part of the translation process, which reinforced the original agenda of UK OGD that was set during the formation of data.gov.uk. By providing more data, developers could continue to create applications, and in turn, demonstrate the value of making local government data available. Effectively, the alignment of local government actors was strengthening existing networks, which helped maintain their commitment and stability.

> *"we did have quite an interesting request the other day, somebody was asking for the geo-location of bins to put dog waste in, and people were quite amused with that required. Then someone said, "I bet they're building an App, they want that data so they can build an App". When you walk your dog, you can find the nearest bin if you want to be a responsible citizen."*
> [Philip Jones: 116]

Gaining the attention of additional actors was supported by other Web services; Blogs and communication platforms such as Twitter provided a medium to share news of local government activity. Actors in other networks such as the civic society and central government used these technologies to share the achievements of local government activities, which provided visibility to a larger network of actors, particularly useful as local government actors were trying to entice new actors such as developers and data users:

> *"RT @WCCOpenData: It's (almost) official - Warwickshire's data is 'open'; for business http://bit.ly/ckTiIu #opendata"* [Twitter Data Stream, 'countculture', 16th April 2010]

> *"Warwickshire County Council election data from 2001 onwards now available http://bit.ly/9zSOCD #opendata"* [Twitter Data Stream, 'WCCOpenData', 22th April 2010]

> *"RT @premierkissov: RT @WCCOpenData: Soon be publishing details for our #hackwarwickshire competition launch on Monday 17th May - http://bit.ly/aYMtDQ #opendata"* [Twitter Data Stream, 'ambrouk', 5th May 2010]

Similarly, online communications were also used by actors within the Lichfield network to discuss their activities with other networks including the civic society network, who helped announce the news of the activities to their own network of followers:

> *"RT @OpenElection: Lichfield DC becomes first council to publish election results as #opendata as part of @OpenElection Data project. Thanks @pezholio"* [Twitter Data Stream, 'countculture', 29th March 2010]

> *"Our webmaste @pezholio has just blogged about how we've used #opendata on the LDC website http://is.gd/6JA7z"* [Twitter Data Stream, 'Lichfield_DC', 21st April 2010]

> *"Councils to receive guidance on migrating data into the public domain http://bit.ly/9WzGd6 ¡– mentions Lichfield DC #opendata"* [Twitter Data Stream, 'jaducms', 24th April 2010]

Twitter was a mechanism to share the news of local governments (who individually did not have a large follower count in comparison to actors within the civil society network). Take for instance the tweet below, it became subject to an extended chain of retweets which diffused amongst those actors already known from previous communications and also was shared by a collection of actors that were previously not identified as active within the #opendata communication channel:

> *"All data from @lichfield_dc is now on data.gov.uk! http://bit.ly/aaWjd0 #opendata"* [Twitter Data Stream, 'pezholi', 1st April 2010]

> *"RT @WCCOpenData: Warwickshire's #opendata is now also available via central gov at data.gov.uk: http://bit.ly/9ivXwS"* [Twitter Data Stream, 'DirDigEng', 21st May 2010]

Sharing this news provided a simple and direct way to engage with not only the actors that had already worked on and contributed to the agenda of Open Government Data, but also aligned actors that potentially could aid the translation of OGD. Web communication platforms such as Twitter has been instrumental in the growth and success in network translation, as an "incredibly important" tool [David Eaves: 159]. Twitter was suggested as a social and technical mechanism that enabled focal actors to:

> *'find the people who gets and cares about open data in an organisation whether it is a government department or a local authority, or a local council'* [Chris Taggart: 103]

It was even suggested that without these Web communication platforms, the technological and social processes involved in publishing data "would not exist" [Emer Coleman: 67].

The goal set by the PTB to get local government actors to publish local government data had come to fruition, and had been supported by other networks and technologies. As illustrated in Figure 8.8, through various network artefacts in forms of incentives, communications, and the setting of passage points, actors within the local government had published data and aligned to the PTB agenda. As a consequence of the local governments' alignment, the stability of the PTB was strengthened. The PTB had become mobilised, it had enrolled other actors-networks and provided a voice for the passive actors within the network. The PTB were able to harness its assocations and actors to set and attract other networks.

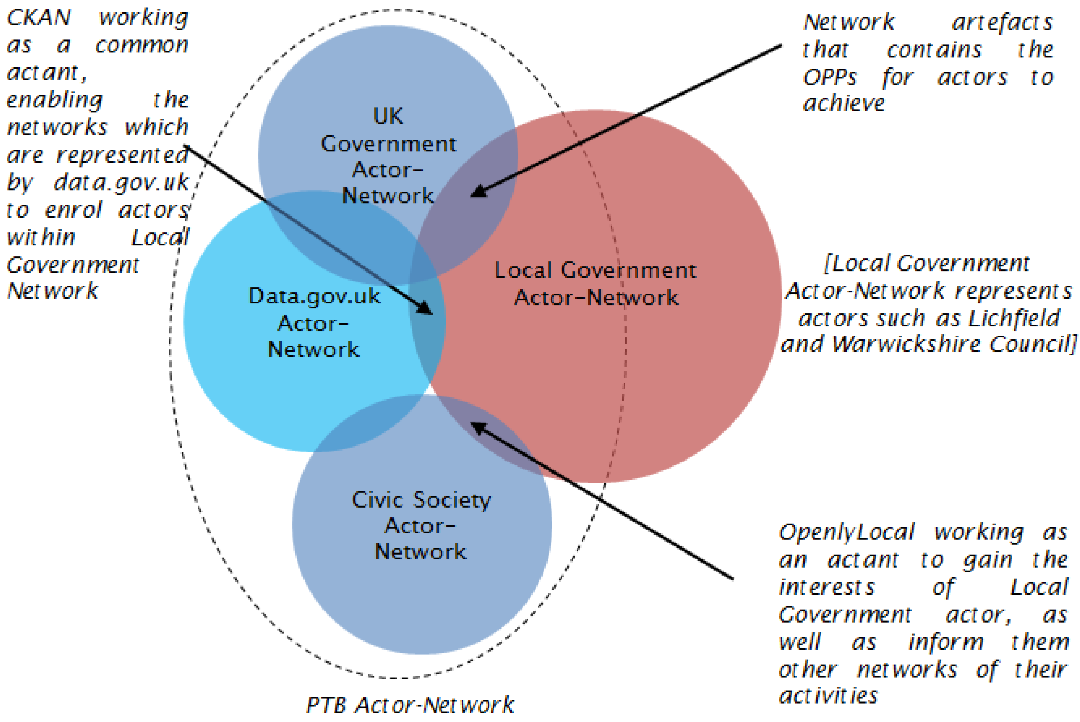


FIGURE 8.8: Illustrating the Interessement and Enrolment of the Local Government Actor-Network through shared actants and network artefacts. Data.gov.uk represents the formation of actor-networks (data.gov.uk has been de-punctualised)

In order for the translation of the OGD network to continue, the PTB had to maintain stability and commitment to the goals of previous networks. Take for instance the academic; their previous alignment to the Linked Data and Semantic Web community was still part of their own agenda, and this the agenda of the PTB. This agenda had developed into the '5 Stars of Open Data', which as Figure 8.9 shows, was set by the academics to define and determine the 'openness' of data in terms of the license, format, and linkability of the data. The academics set this as a passage point within the PTB network, requiring networks such as the UK government and local government to publish their data in specific formats. To this end, the 5 stars of open data was offered as an interessement device, passed between actors, which helped align their activities and efforts of publishing data. Offline activities reinfornced the passage point of 5 stars of open data; events ran by the OKFN such as the 'Open Government Data Camp' brought together actors from the already known and enrolled actor-networks and also those interested the potential benefit that Open Government Data could provide, including “data experts”, “policymakers”, “journalists”, “citizens”, “advocates” and “reusers” ['EventBrite.com', ‘Open Government Data Camp Registration Page’ 2010].

★	Available on the web (whatever format) *but with an open licence, to be Open Data*
★★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people’s data to provide context

Figure 8.9: The ‘5 Stars of Open Data’ as provided by the Academic-Network via a number of presentations at international conferences, helping attract a wide range of actor-networks

By taking a more critical observation of the activities discussed, similarities with the emergence of data.gov.uk become apparent in terms of the processes required to enable new agendas to be problematised and acted on. The formation of the PTB is the result of multiple actors working towards similar agendas; the academics, OKFN and government, all worked on different parts of an agenda that had the common factor of openness, whether this was developing technologies, changing cultural views, or even the creation of policies to support it. In a similar fashion, the activities of the different actors within the local government, helped by the PTB network, led to a network inscribed by different goals and agendas of the newly enrolled actors-networks.

8.1.4 Actor-Networks Becoming Mobilised

The OGD network had widened, new actors became enrolled, and the activities surrounding the UK OGD Web activity were changing. As Figure 8.10 shows, the numbers of actors partaking in online communications had grown steadily during 2010, which in comparison to 2009, had risen by over 1200%, as shown in Figure 8.11. The offline activities had also increased, demonstrated by the number of actor-networks involved in participating and aligning with the OGD agenda of publishing Government Data. The formation of the PTB and the increased online and offline activities led to more datasets being produced locally and nationally. A point was reached where the strength of ties of the previous actor-networks had created a snowball effect; the strength of the ties between actors provided the conditions for new networks with common interests, actors and artefacts to become part of the translation towards OGD.

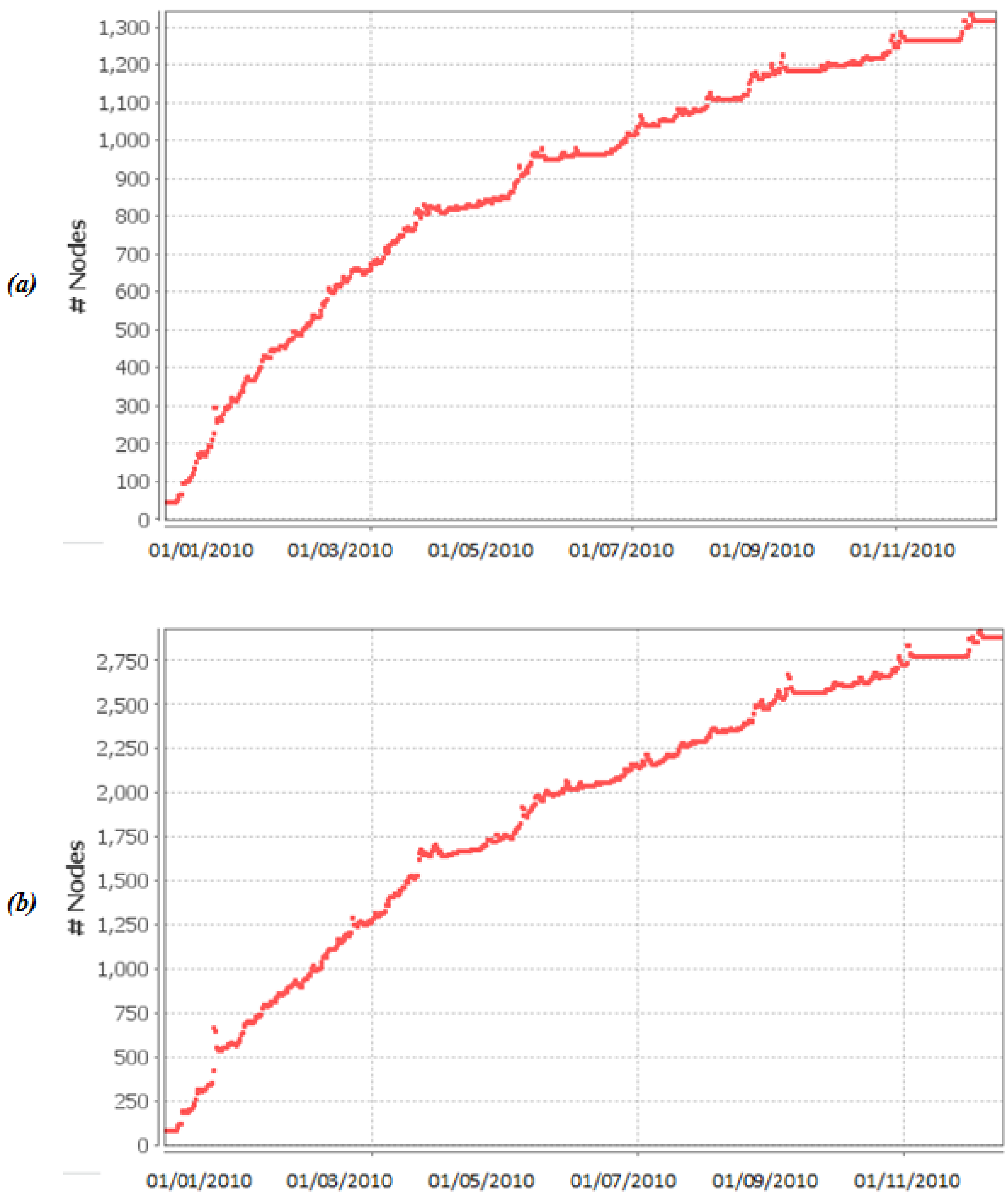


FIGURE 8.10: Number of unique individuals participating in Twitter (a) Mention and (b) Retweet communications between January and December 2010

In summary, discovered in this section:

- Translation of data.gov.uk led to the emergence and formation of the Public Transparency Board (PTB), a network formed around a new agenda driven by the

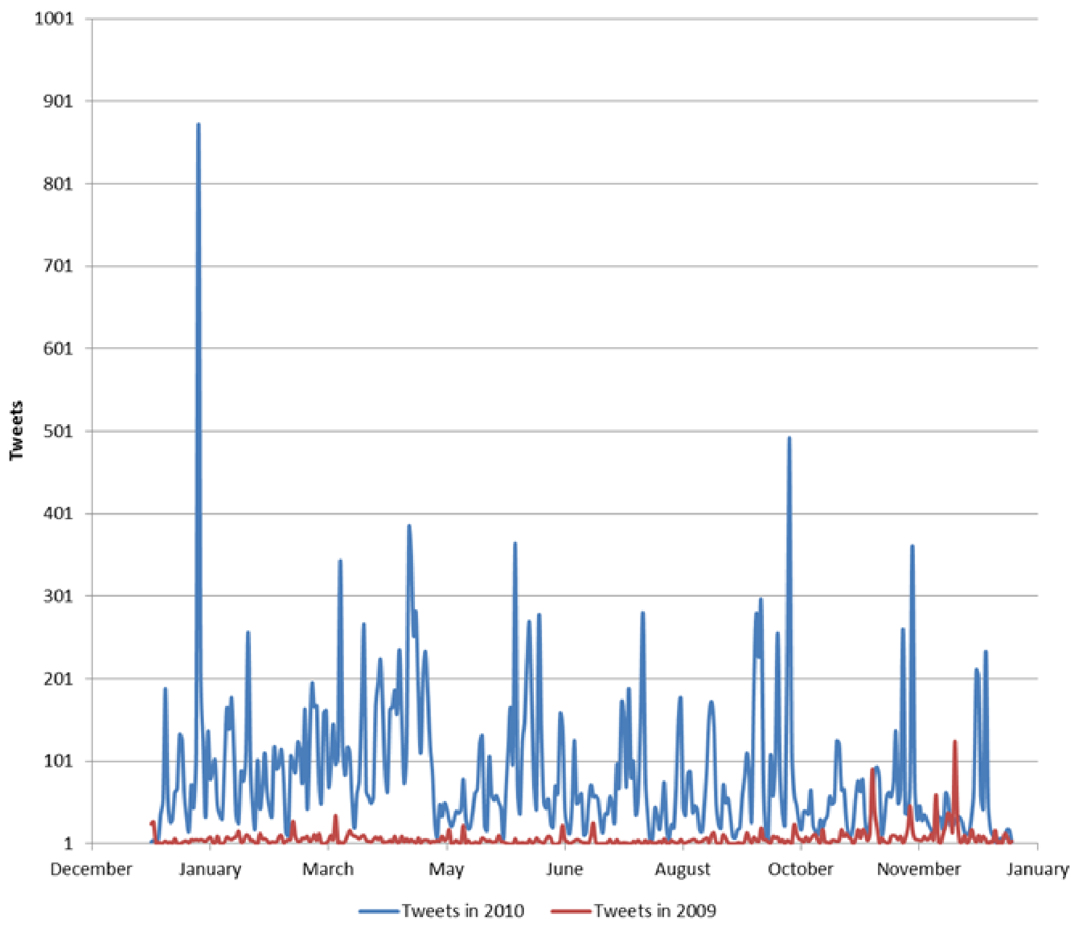


FIGURE 8.11: A comparison of the volume of Tweets (including mentions and retweets) made during 2009 and 2010 regarding #opendata

actor-network.

- The stability of existing networks helped new networks to become interested and aligned with other networks agendas.
- The enrolment of new actors changed the network's configuration as new passage points were introduced, along with new agendas.
- Technological artefacts facilitated the enrolment process, and the combination of technological artefacts is extremely powerful in gaining the support of new actors.

8.2 Sustaining a Network of Actors

Translation of an actor-network is an on-going process, and even when it reaches mobilisation, commitment of the enrolled actors is required in order for it to continue to achieve its goals. This section examines how a network operating within a mobilised state of translation can continues to translate, and how socio-technical processes are used to overcome threats and barriers which may impede their stability. The following section will examine the barriers encountered during the aligning local government, and the socio-technical processes used to overcome them.

8.2.1 Barriers to Enrolment

The enrolment of local government to UK OGD was driven by an agenda set by actors within the PTB network. However, despite the pre-existing associations between the PTB and local government, gaining their enrolment was faced with challenges, which included overcoming the barriers of local government's physical location:

> *"when open data came on the agenda they just seem to be, it was a very much in London based initiative, even now I suppose what we were doing was a little bit more impactful"* [Julian Tait: 198]

As described above, due to geographic location, local government actors not positioned close to central government or more specifically, London did not receive the same level of support to achieve the goal set by the PTB to publish their data:

> *"there isn't really any support as such, and I do think that being in London, or having easy access to meetings in London, as a lot of stuff does get formed on the basis of discussions and meetings that do happen within the capital, [...] There have been support through legislation, and that cannot be downplayed at all. Because if they say this is something that you have to do, through government edicts, that states certain things that can be done in a certain way, then you've got to say 'guys you've got to be doing this'."*
> [Julian Tait: 218]

Figure 8.12 illustrates the location of the different local government actors in terms of their geographic location within the UK. The colour of markers on the map indicate whether or not they are publishing their data; a red marker represents a local government not publishing data and a yellow represents a local government publishing open data. As Figure 8.12(b)reveals, the local governments that had published open data were geographically situated closer to Greater London, confirming the interviewees' views. Given these barriers, the following section will now examine how social and technical processes were used to overcome the alignment of local government actors despite their geographic location.

8.2.2 Reducing Barriers Using Web Technology

The Web as a technological platform is geographically neutral, which was critical for local government actors to align to the OGD agenda that set by the PTB. Web communication such as Twitter supported the translation of the local government actors within the OGD translation; facilitating actors with tools to identify and share with other actors, gaining support, and strengthening existing associations. However, whilst Twitter was an excellent broadcasting technology, it was also the network of actors that used it that made it such a powerful tool for aligning local government. For instance, the national media network used Twitter to disseminate and share news of their activities as well as the OGD networks they were associated with. As shown previously, actors within this network were essential during the earlier translation of data.gov.uk, functioning as an intermediary by disseminating news to and between associated networks such as the academics, government, and OKFN. However, without having an audience to view this information, their efforts of sharing content would have achieved very little.

As Figure 8.13 illustrates, the size of the audience that 'The Guardian' had obtained in comparison to actors within other networks is substantially larger, with the potential to broadcast and mediate messages to over 30,000 actors. In comparison to this, other individual actors with the associated networks have a maximum reach of less than 500 actors. As 'The Guardian' was already involved in the activities of OGD, using their network of associations provided a way to gain the attention of a large number of actors:

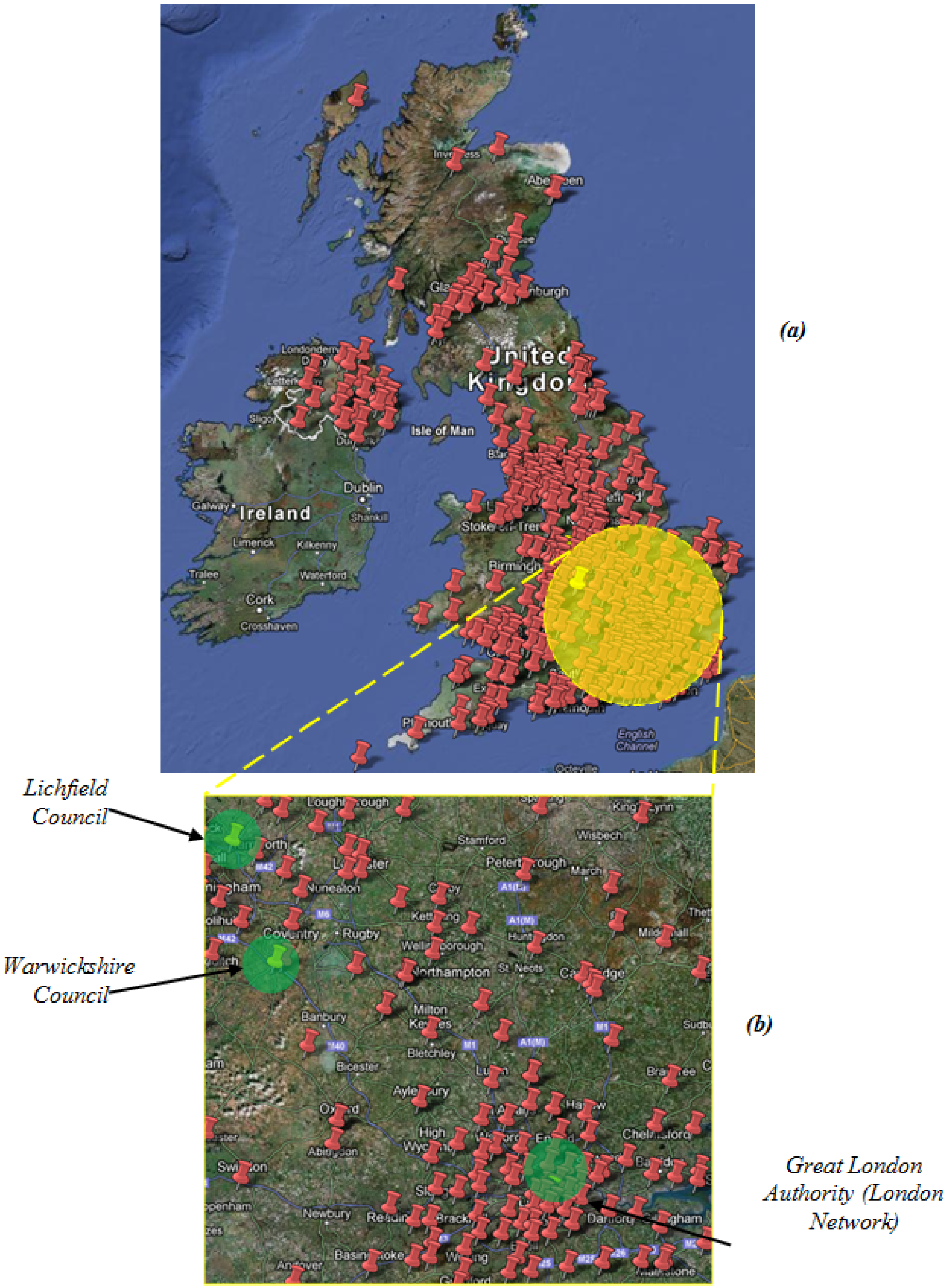


FIGURE 8.12: Visualisation of the actors in the Local Government Network, Red placemarkers represent 'Closed Actors' and 'Yellow' represents those that are producing Open Data. Image (b) provides a more granular view of the London geographic area. Snapshot of January 2011

"[The Guardian] are very hot on seeing how the open data is being used and finding what people have done and telling their stories [. . .] they had lobbied

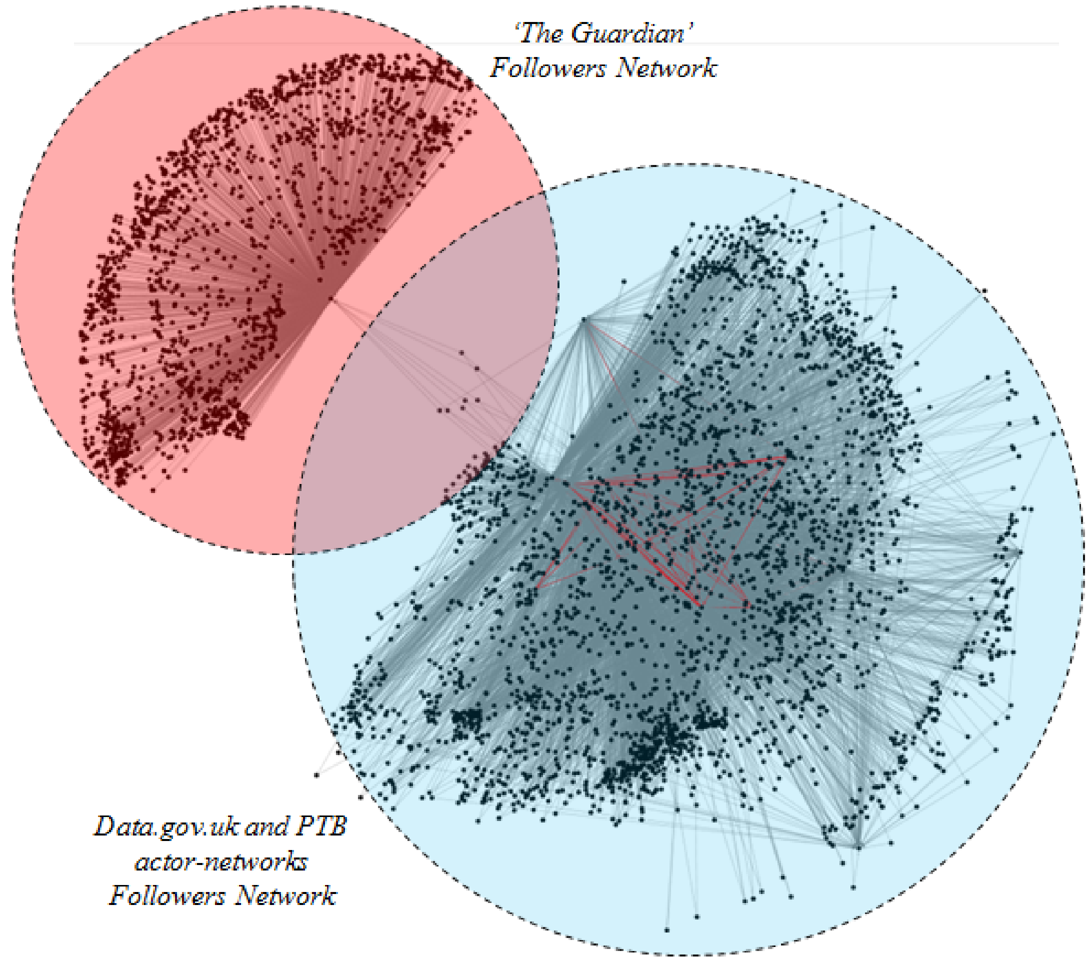


FIGURE 8.13: A visualisation of the static Twitter follower communication network connections of 'The Guardian', an actor within the National Media Network, and other actors (connected by red edges) in the data.gov.uk and PTB network

quite hard, and then they were very supportive" [Emma Mulqueeny: 74:116]

The national media network took advantage of their network and produced articles such as 'A year of data.gov.uk' ('The Guardian Datablog', 21st January 2011), which described the success of data.gov.uk. Articles such as this described the enrolment and activity of different government department and also the aligned actor-networks part of the PTB:

> *"Everything from local authorities spending to the whereabouts of the nation's bus stops, from energy consumption in ministries to infection rates in hospitals. [. . .] data.gov.uk itself was originally designed for developers, and is pretty spartan. There is now a wider interest in government data*

> *and its visualisation"* ['Guardian Datablog', 'A year of data.gov.uk', 21st January 2011]

Supported by Twitter's broadcasting mechanism and 'The Guardian's' large network of associations, actors within the academic and government network were able to reinforce the ties between existing networks of activity and restate the agenda that had been set:

> *"Improving the UK's information landscape needs this kind of still constant attention. One of our next phases of work is to amend Freedom of Information legislation and to create a "Right to Data" to give the public extra means to obtain data from public bodies. [...] They are achievements that are being emulated and copied around the world. This open government data revolution could be as important as any we have seen in the Web era."* ['Guardian Datablog', 'A year of data.gov.uk', 21st January 2011]

The article highlighted the commitment to publish more data and enrolling local government actors, which was further reinforced by additional network articles published by the national media network, which included statements detailing the requirements set out for local government to fulfil:

> *"in the next two months yet more data will be released, including street-by-street crime data, details of all government contracts over £25,000, and details of every item of spending over £500 by every local council in England. So far, 200 out of around 360 have done so."* ['Guardian Datablog', 'Government data UK: what's really been achieved', 21st January 2011]

> *"Sir Eric Pickles informed local authorities that they should publish items of spending over £500 by January 2011. [...] sticking to the common features in these guidelines helps to make the data easier to gather and analyse, and people are keen to do this. Take for example Chris Taggart from Openly Local. Taggart, who has so far imported 91 councils' spending,"* ['Guardian Datablog', 'Local council spending over £500', 1st February 2011]

The technologies used to publish these articles also provided a forum for actors interested in OGD to discuss their activities and experiences with the data. The discussions were a demonstration of continued engagement and support towards the translation of OGD, it was also demonstrating that the activities of previous networks were being built upon as a platform for new developments:

> *"We are busy ploughing quite a lot of time and money at Placr into building an open data aggregation platform capable of supplying data feeds through an API. We have only got UK Travel Options for iPhone and Tube Radar to show for it so far but there is a bunch of other stuff in the pipeline. It is going to take time to gear up to exploit open data... the backend operations are complex. [...] open data will change the public sector in due course, by giving it new more informed and independent drivers. This is especially important where there are monopoly providers"* ['Guardian Datablog', 'User Comment: Government data UK: what's really been achieved', 21st January 2011]

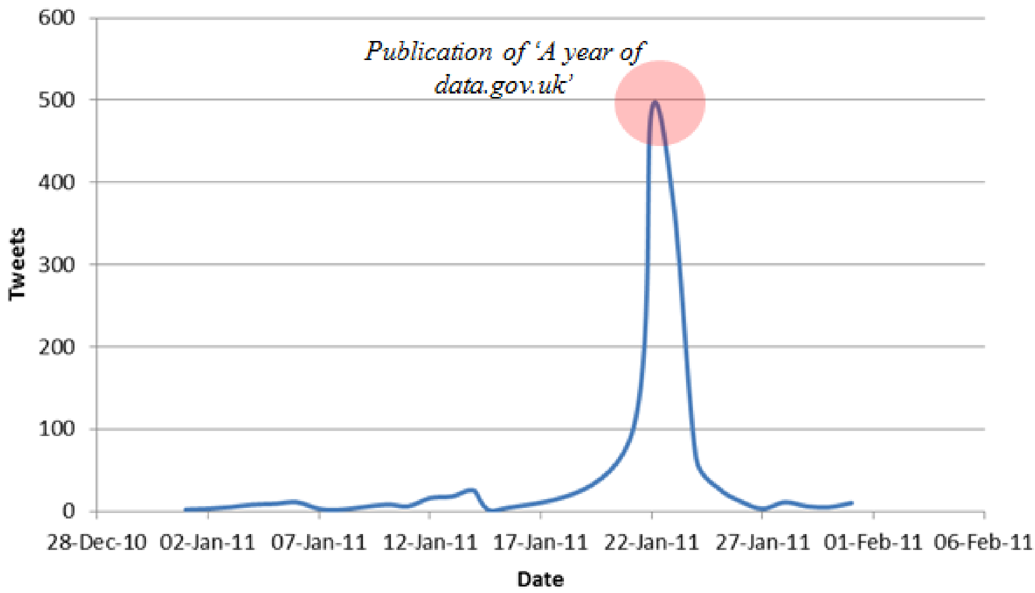


FIGURE 8.14: The volume of Tweets that were made surrounding the articles produced by 'The Guardian' discussing the success of data.gov.uk

Twitter once again was instrumental in disseminating the news online as Figure 8.14 and Figure 8.15, and within these tweets were the use of embedded links to the articles, which resulted in tweets which were part of long retweet chains, including tweets such as:

> *"RT @paul_clarke: Data.gov.uk 1 year on. Some general thoughts before #ukgc11 http://bit.ly/e9Ml7i ¡-My comment http://j.mp/h4qvWO"*
> [Twitter Data Stream, 'countculture', 22nd January 2011]

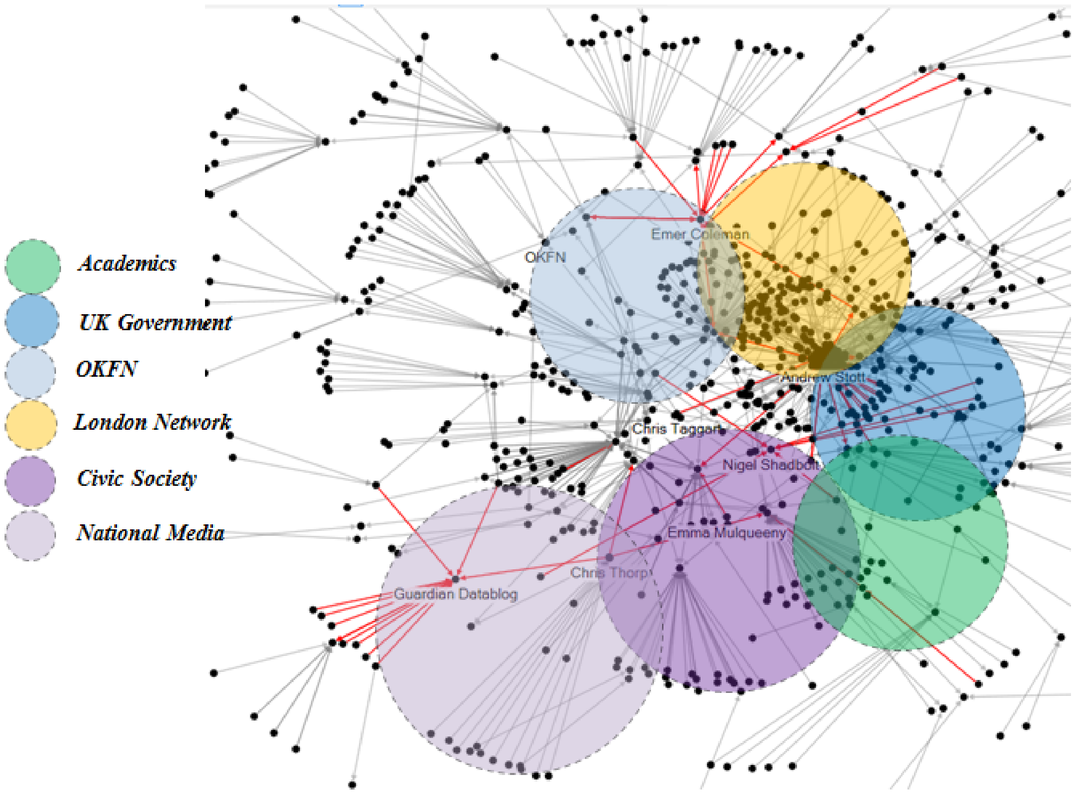


FIGURE 8.15: The communication network of retweets that contained the terms 'data.gov.uk' and 'Guardian' during January 2011. Overlaid are the de-punctualised actor-networks that represent data.gov.uk and the PTB networks

The peak in activity identified in Figure 8.14 represents tweets that contained similar content to 'countculture' tweet, and were part of extensive retweet chains across the communications network. These highly retweeted messages accounted for more than 70% of the communications made during the time period shown in Figure 8.15. Similarly, the translation of the Greater Manchester Open Data (GMOD) actor-network was supported by extensive retweets regarding the success of their activities. As Figure 8.16 illustrates, tweets that announced the success of the GMOD's OGD activities became subject to multiple retweets, including retweets by actors within networks such as the UK government, academic, civic and OKFN:

"RT @MadProf RT @billroberts: http://bit.ly/k2U62M Transport linked #opendata for Manchester Great stuff!![lucky bugrs]" [Twitter Data Stream, 'DirDigEng', 18th May 2011]

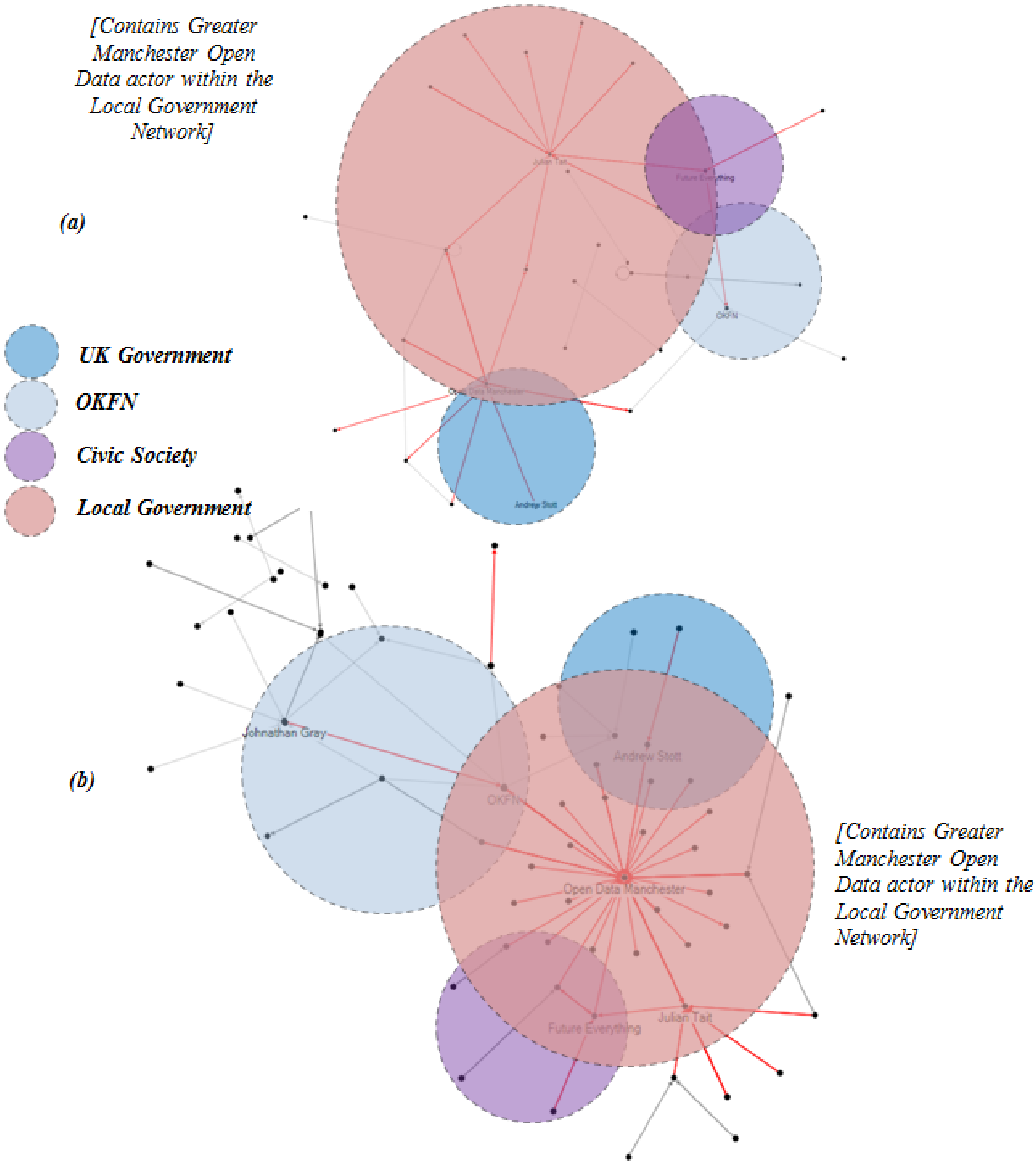


FIGURE 8.16: Twitter communications network of (a) Mentions (b) Retweets extracted from the OpenData dataset containing the String 'Manchester' during the time period of January to December 2011. Overlaid are the different actor-networks represented by associated actors

Web technologies and artefacts (articles, publications) facilitated the associations and communications with actors already aligned to the goals of OGD, and they also provided a way to raise awareness to actors that potentially were not aware of these activities.

Essentially, the Web enabled actors to engage with each other aiding the alignment of local government. Online communications had a direct effect on the networks undergoing translation, and as a consequence, helped local government actors as well as PTB actors, translate.

Technologies were only one way to overcome barriers. The translation of individual local government networks also helped align them with the OGD agenda. Demonstrating this, the Greater Manchester Open Data (GMOD) local government became aligned with OGD as a result of previous activities similar to the PTB:

> *"There has been a development path of open data practice within Manchester [...] basically back in the beginning of 2009 I started to develop a project of Future Everything [...] It was basically an enquiry asking how cities would develop if all data was made open and whether they would evolve and develop with the same asymmetry as you would see at present. [...] it wasn't originally started as an open data project, but it was seen as an interesting area of enquiry and we were looking at where else there was any examples at that time of where cities had adopted open data practices"*
> [Julian Tait: 7]

The activities of the GMOD network were important precursors for helping align their agenda towards OGD. Similar to situation of the UK government before their alignment with other actor-networks (academics, OKFN), actors in GMOD identified that they needed to publish data despite the lack of technologies or knowledge, and struggling to overcome social barriers of actor alignment:

> *"we started working with Trafford Borough Council, and they started wanting to develop an intelligence hub, which is all about close data, and close analysis. So we said this really has to be open, and this is where we started the Data GM project together. [...] you get some people who are trying to block it by saying, 'we'd understand what the point of this is, or how do we do it, and in three months' time we will still have jobs' etc."*
> [Julian Tait: 126]

The social barriers of adopting OGD were overcome by the successful translation of technologies and social practices instilled by data.gov.uk. As a consequence, data.gov.uk helped GMOD gain the technological capabilities and establish social practices in order to achieve their goals:

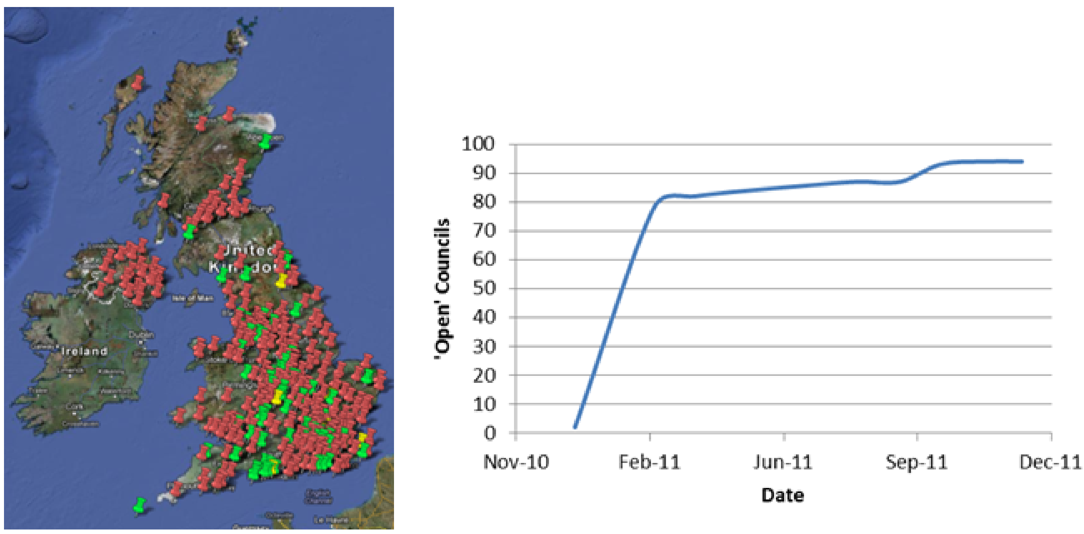
> *"Data GM did evolve and it did become something, and it came something quite quickly. Because we used CKAN as a cataloguing system, and within three months of the project starting we have something to show them [...] The reasons why we chose it was because it was being used data.gov.uk, and this process had to be quick, because as soon as you go through procurements, it becomes a very lengthy process, and so we have to go through a single tender action. The reason we went through the single tender action was because basically the CKAN system was the only one around"* [Julian Tait: 142:291]

As the extract above suggests, the translation of GMOD was due to the availability and capability of CKAN, along with the social processes and barriers that it overcomes. However, the use of the CKAN technology by GMOD was not only a way to help another local government actor to enrol, it also strengthened the OKFN commitment, ensuring the alignment with OGD:

> *"it is based on a platform that no one in the local authorities have any experience with know how to use, it is a Drupal platform, and Drupal developers within local authorities, probably in the UK 5 or 6? There are just not used to it, and so they become very dependent on CKAN to look after the instance."* [Julian Tait: 298]

The enrolment of the GMOD network required the commitment of networks that were already enrolled. The UK government network were active in providing incentives and network artefacts, and commitments were demonstrated by the publication artefacts documenting the current success and next-steps of OGD, including the requirements for local government to publish their data ('PM Letter', 7th July 2011).

GMOD was not the only local government actor in translation. Open Kent followed a similar translation process to that of GMOD, sharing the common interests in publishing council data, and aligned with the PTB agenda. However unlike the translation of the GMOD network, Open Kent did not use CKAN, instead, they created their own technological platform, Kent Connects, which was designed by their own team of developers ('Open Kent Blog', 'Councils Open the Door on Transparency', 27th May 2011). Despite not using CKAN, Open Kent still received support from actors enrolled in the OKFN network, they disseminated the success of Open Kent's ('OKFN Blog', 'Open Kent', 24th August 2011). This demonstrated the commitment of OKFN extended beyond just the adoption of CKAN, but was driven by the desire to help OGD translate. Between January 2011 and December 2011 the number of enrolled local governments had grown by 86 and as shown in Figure 8.17(b), despite the geographic barriers previously discussed, the alignment of local government was no longer limited to the boarders of London, as show in Figure 8.17(a).



(a) (b)

Figure 8.17: (a) Visualisation of the actors within the Local Government Network, Red placemarkers represent 'Closed Actors' and Yellow represents those that are providing some Open Data, Green represents 'Fully Open'. Snapshot of December 2011. (b) Number of 'Open Councils during 2011

8.2.3 The Mobilisation of Enrolled Actors-Networks

The alignment and translation of local government actors also affected existing networks involved with OGD which helped strengthen common goals between networks and in turn required their agenda to adapt to accommodate the new goals that had been identified. This was illustrated by the renewed associations and alignment between the London and PTB network, which was helped by the incentives of publishing valuable transport data:

> *"The Transparency Board agreed that it would be good to see more transport data opened up in London but that a lot is already happening under the leadership of the Mayor; and suggested that the fact that transport services outside of London are not in direct political control may actually increase the value of the prize. This is because releasing the data gives people a chance to present and appreciate a whole-system view."* ['PTB minutes' 'Transparency Board Minutes', 8th February 2011]

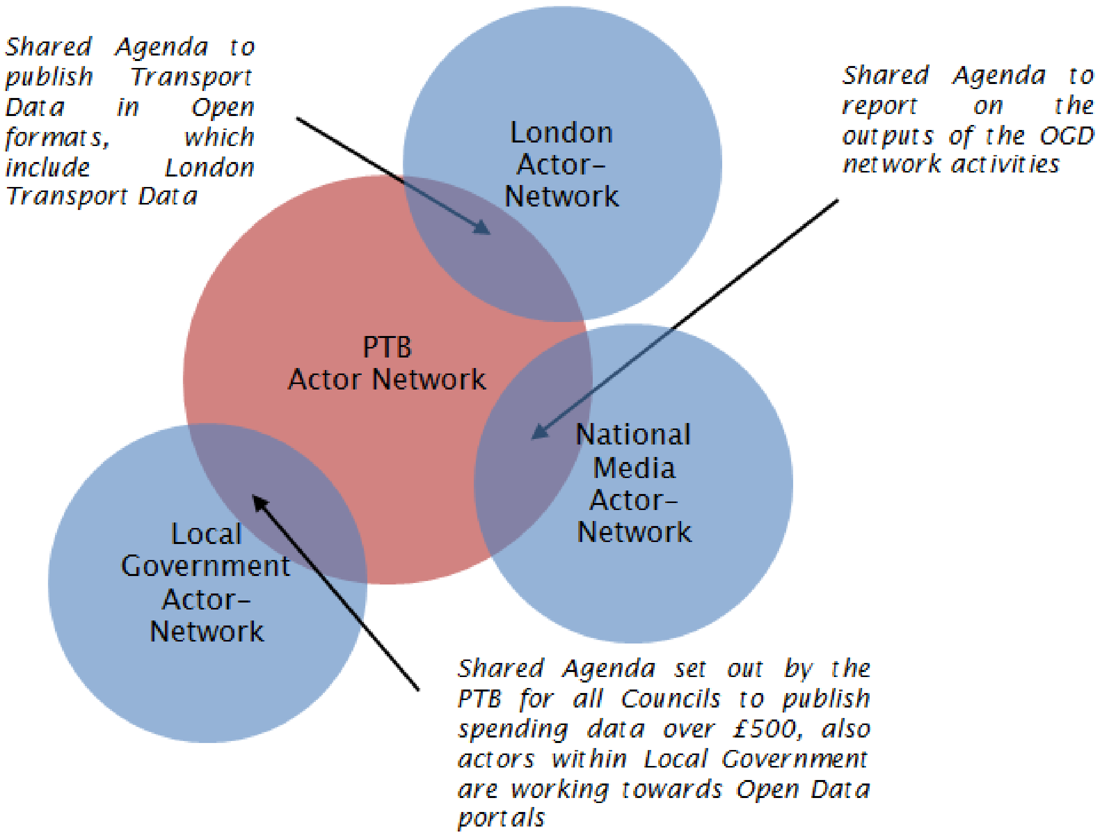


FIGURE 8.18: Illustrating the agenda and passage points between the PTB actor-network and the Local Government, London, and National Media network. The PTB represents a punctualised actor-network which includes the actors-networks within data.gov.uk

As Figure 8.18 illustrates, the PTB was supporting a number of networks, which before its formation, were only connected via shared actors and transient links. Through the production of network outcomes and additional networks building upon their success, the PTB operated in a mobilised state of translation, which helped establish common goals to enable networks to translate towards OGD. As part of this mobilisation process, other networks were demonstrating their commitments; academics, the national media, and OKFN were involved with running international events ('OGD Camp 2011', 'Open Government Data: Making it Real', 21st October 2011), which provided a method for these networks to discuss their successes and achievements of OGD. They also enabled other networks such as civic society and developers to reinforce their commitment and presence within the translation of OGD.

As identified during different stages of the translation process, online communications were used to support the events, and during the time frame of 1st October to 30th November 2011, 12% of the 76,000 tweets made were related to the events. As Figure 8.19 shows, various identified networks were involved in either creating or sharing communications (63% were retweets and mentions), with certain tweets diffusing across them all:

> *"RT @DataGovUK: Hear @DirDigEng on why it is important to submit views to the #opendata #openuk consultation: http://t.co/BPbrt160"* [Twitter Data Stream, 'anked', 13th October 2011]

> *"RT @okfn: Cool! Just confirmed @DataGovUK and @datastore as partners for #ogdcamp 2011 in Warsaw: http://t.co/W8IU8yX"* [Twitter Data Stream, 'datastore', 21st October 2011]

> *"The future of open data? http://t.co/14PqI0kQ va @guardian #opendata #openGov #oGov #Gov20 #OGP #OGDCamp #GobiernoAbierto"* [Twitter Data Stream, 'redmatriz', 21st October 2011]

> *"RT @bethnoveck: Tim Kelsey and UK Gov launch consultation on transparency and open data. Replies sought by Oct 27. http://t.co/qafYmO4e #opendata #opengov"* [Twitter Data Stream, 'zarino', 21st October 2011]

As the tweets above illustrate, the mobilisation of the PTB helped other networks translate towards the agenda of OGD. By local government actors publishing data, they

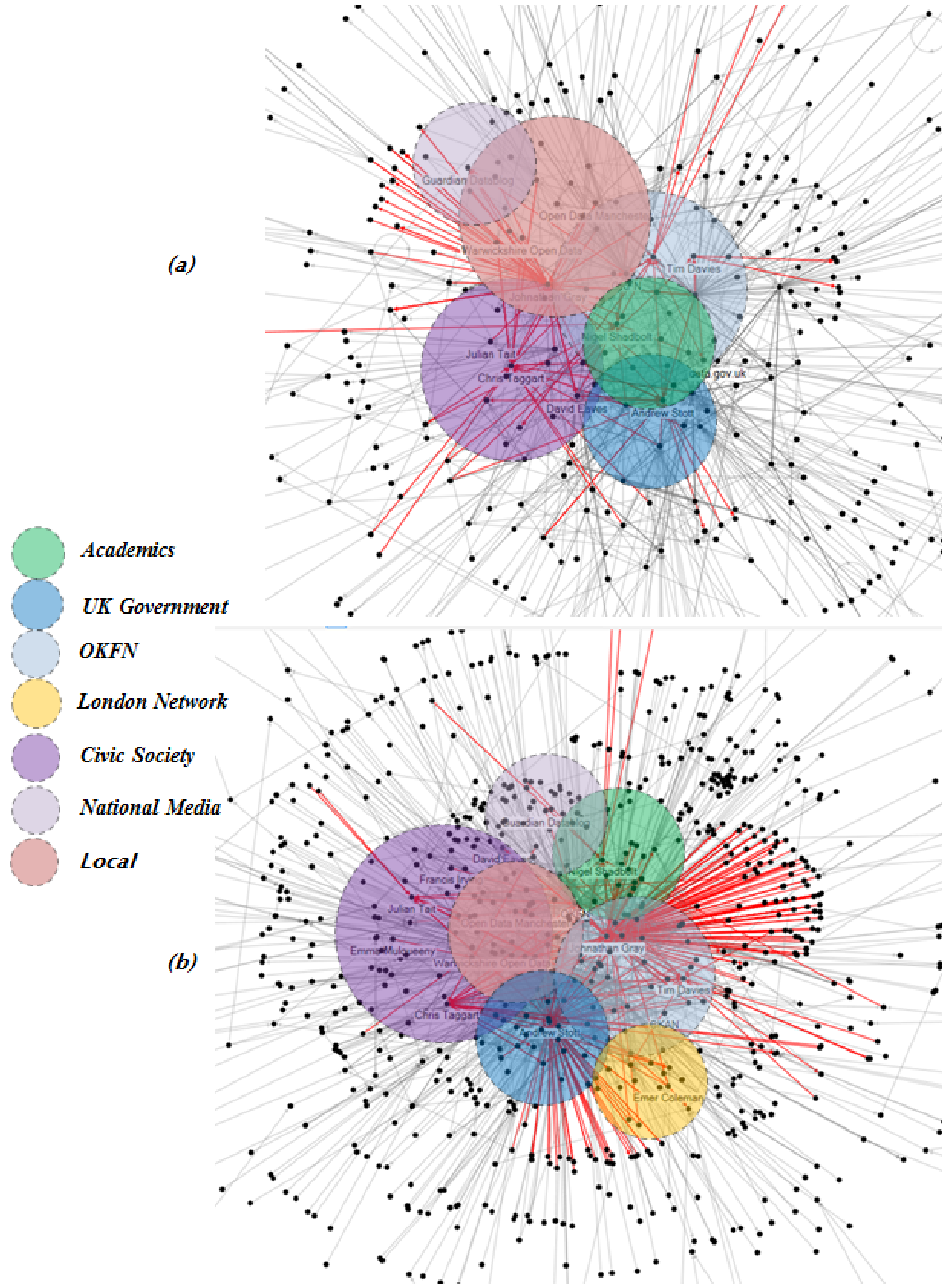


Figure 8.19: Twitter communications network of (a) Mentions (b) Retweets extracted from the Open Government Data Camp 2011 Twitter Dataset during, which contained tweets using the #OGDCamp hashtag. Overlaid are the different actor-networks represented by associated actors. 1st October 30th November 2011

passed through the passage points set by the PTB and translated as enrolled actors aligned with the goals of OGD. Supporting this was the use of on- and off-line activities, including events, online communications, and the production of network artefacts for incentives and signs of commitment.

The publication of data and the development of applications and presentable network outputs not only helped reinforce the relationship between actors, but were catalysts to gain the interest of actors to become aware and interested in the benefits of Open Data. The UK public Open Data consultation was responsible for gaining their interests, as it was aimed at "citizens, businesses, public services themselves, and other interests groups" in order to understand and improve how the OGD network can "best embed a culture of openness and transparency":

> *"Empowering citizens with openness and transparency with the aid of government data"* ['Cabinet Office', 'Making Open Data Real: A Public Consultation', 27th November 2011]

The alignment of a new network of actors, who were originally problematised as part of the original UK Government agenda during the initial translation of OGD, was being realised. However, as the translation process has shown, the alignment of citizens has relied on the different activities that led to the enrolment of local government network, their ability to achieve the PTB goals of publishing their data, and the socio-technical processes involved in making this visible and known to actors outside of the translation process. Consequently, these processes enabled citizens to be able align with OGD.

8.2.4 The Alignment of Citizens

The activities of the mobilised actors, specifically the publication of data and publicity of Open Government, worked as incentives for a new network of actors, the citizens, to become aligned. Whilst these actors were problematised earlier within the translation process by the UK Government, the enrolment of citizens was not possible until the stability and translation of other actor-networks. The production of network artefacts such as "Making Open Data Real: A Public Consultation" aided this enrolment process, which was a report written by government describing the successes and future plans of UK OGD. This report led to 240 written responses and another 217 online responses which contained feedback towards the development of data.gov.uk; as the 'summary of responses' suggested, in terms of engagement from the citizens:

> *"The volume of responses submitted is indicative of the strength of interest in the Transparency and Open Data agenda. In 2012 the Government will set out its strategic vision for the agenda and its response to the evidence submitted to the consultation."* ['Cabinet Office', 'Making Open Data Real: Summary of Responses', 30th January 2012]

This report also provided a means for actors to express their view which in turn reconfigures the OGD agenda. Responses spanned from discussing privacy of data, to data quality and commitment of data publishers at the national and local level:

> *"A number of respondents, particularly those from across the public sector, stressed the importance of the data controller retaining control over the data - i.e. whilst data may be signposted to via a central portal, it should continue to be hosted locally (on the data controllers website)."* ['Cabinet Office', 'Making Open Data Real: Summary of Responses', 30th January 2012]

In terms of network artefacts, the 'Open Data Consultation' played a critical role in reinforcing the commitment and agenda between the actors within the previous translations. This was supported by the publication of data (at a local and national level) and the engagement with citizens, to which their communications had been facilitated by online communication platforms and technologies such as Twitter.

> *"There is a high level of engagement with some of them, I happily engage with them and some of my colleagues across the chamber do. It's quite nice when you have an intense discussion with them [...] So there are about 7 or 8 councillors that are regular tweeters, and we tweet not just in general, but also in meetings. At our last general meeting, which was the budget council, I tweeted 50-60 times, one of my labour colleagues did roughly about that, and some of my conservative colleagues did something similar."* [Gareth Anderson: 131]

This particular use of these technologies was an illustration of a co-constructed relationship between the actors and actants as well as the relationship between the online and offline activities. The ability for actors in different networks to communicate with each other appeared critical in strengthening the ties across networks, which reduced the barriers of translation, and the combination of multiple technologies enabled actors within the citizen and local government network to interact and engage with each other:

> *"We also web cast a lot of our meetings live, so people can see them. So in the last meeting at the council, people were watching it online and sending in comments, [...] We are Web casting and we are looking at ways to use a large screen behind the council, which the video is displayed, but we are also looking to have live tweeter feed, so anyone that uses the correct hashtags will have their comments appear, our media team tweet during councils and during planning of various meetings that are web casted. We have hundreds of people watching these meetings live, so our assumption in everything is to make everything as we can, give access to public to all information."* [Gareth Anderson: 147]

8.2.4.1 Enablers for Network Alignment

Similar to the geographic layout of the local government network, the citizens was a decentralised network with actors located across the UK. Consequently, these Web technologies decreased the barriers for them to engage with the OGD agenda. Citizens were able to interact directly with actors in the government (local and national) network, removing the layers of hierarchy and reducing time of response. In addition to being able to contact new actors, citizens were able to access and interact with the published data through the various data portals developed, both at the national level via data.gov.uk and at the local level, such as the Kent or Manchester Datastore. These interactions provided new opportunities for citizens to engage with the OGD activities. As actors associated with the local government network described, the ability for citizens to access data had affected the existing processes and complexity of the interaction between them and the local government:

> *"we have a small group of people, who I would describe as watchers, and they operate quite widely on Twitter and many are critical of our council and our predecessors and that has continued with us. [...] what they do is*

> *they examine all that is going on in the council all the time, and me and a couple of my colleagues debate quite a lot on twitter with them a lot. Some of them complain about the speed of our Freedom of Information requests, where specific information that they after is not publically available. But a lot of these requests are answered by saying the information is already available and here it is."* [Gareth Anderson: 100]

The publication of data was beneficial for citizens and local government, it reduced the time spent on responding to Freedom of Information (FOI) requests [Carl Haggerty: 272] and it also helped improve the internal operations of local government as it removed the barriers of obtaining data without specific organisational processes and forms [Julian Taint: 266]. The national media network had also explored the possibilities of using the local and national data and as part of this process of experimenting with the data they developed and shared their experiences with new methods and techniques, which responded to the requests made by citizens for tutoring and support:

> *"Alongside the development of an open data standard(s), many respondents asked for more education concerning the use of data, in order to ensure users understand the importance of the context of data (i.e. its limitations) and are able to derive benefit from it."* ['Cabinet Office', 'Making Open Data Real: Summary of Responses', 30th January 2012]

The national media network wrote articles and offered training for citizens interested in using government data, providing example applications and toolkits for citizens to immediately access and examine the data ['The Guardian Datablog', 'Facts are sacred: the power of data out now on Kindle and iBooks', 31st January 2012]. These actions helped align and strengthen the relationship with the citizens, ensuring they had the support required to make use of the data. The alignment of citizens was a multi-faceted process, it required them to be interested in using government data, but it also relied on the previous activities and incentives of networks involved in UK OGD. As Figure 8.20 illustrates, gaining the support of citizens can be partly accredited to the activities and stability of the previous networks; it required data.gov.uk to be established, for data to be published and developers to create applications, for the PTB to form and for local government to enrol. However, from the citizen's perspective, these different stages in the translation of OGD appear as black-box processes. Yet, these black-boxes represent a complex network of actors and agendas, which have translated due to the various socio-technical processes, commitments of actors, and production of network outcomes.

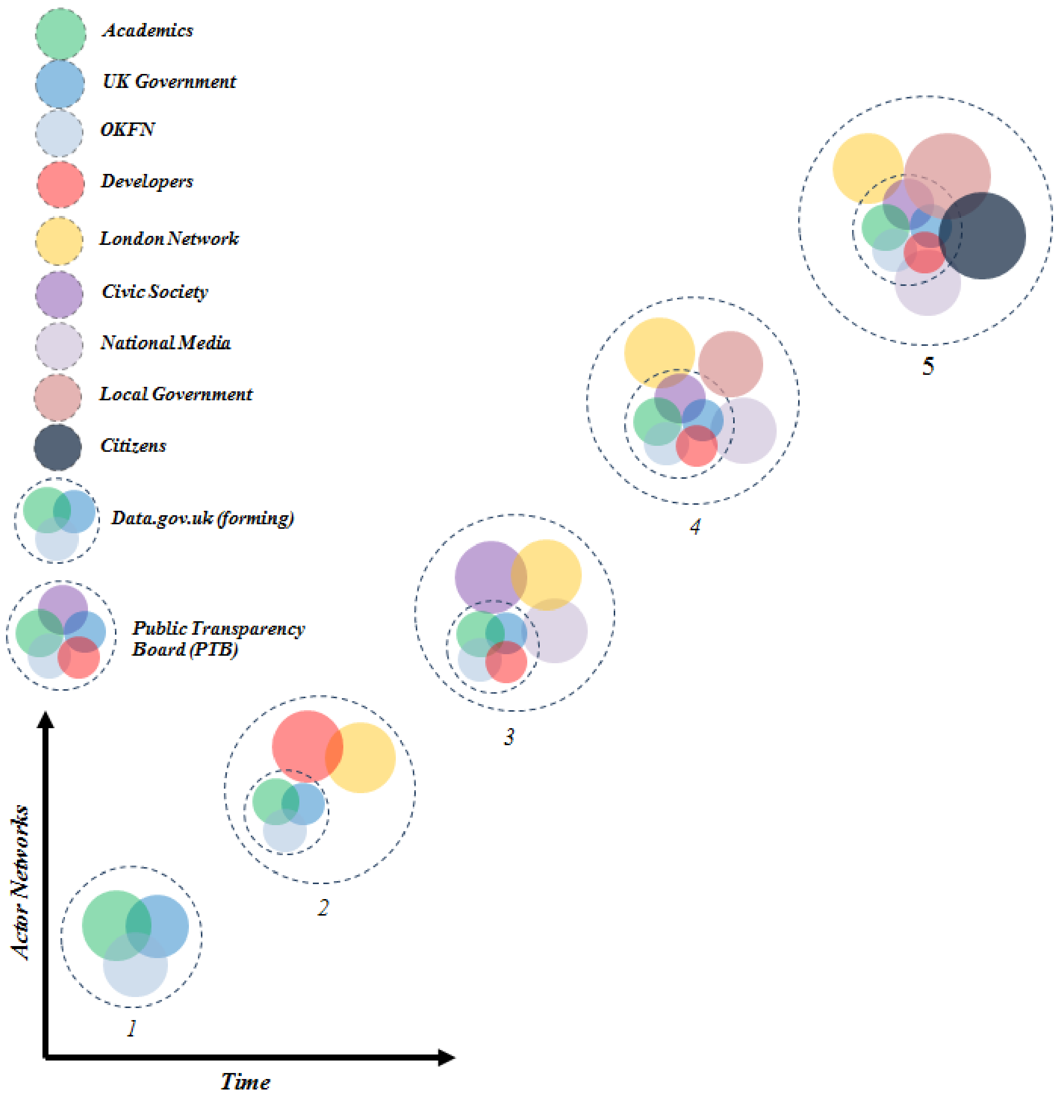


Figure 8.20: An overview of the development of OGD plotted (not to scale) against the time and number of actor-networks. This is a representation of how the different actor-networks involved have come together to enable a translation process, which is spreading across translations

As the analysis of the growth of the OGD Web activity has shown, OGD is not one, but many things and through the interactions of actors, the formation of agendas, and the socio-technical processes used, translation led to a Web activity operating in a mobilised state, enabling new networks to emerge and translate.

In summary, discovered in this section:

- How the barriers that are faced for sustaining a network operating in a mobilised state of translation
- How the use of socio-technical processes to overcome barriers to the enrolment of new networks, and also to ensure stability with networks already enrolled
- How the mobilisation and stability of existing networks enable new networks to become aligned and begin translating.

8.3 Mobilisation of Open Government Data

The previous sections have shown how the interactions of actor-networks have translated to a new form of Web activity, namely UK Open Government Data. As part of this process of translation, agendas had been established, goals had been achieved, and networks become mobilised. However, translation is an on-going process, and stability is only ever a temporarily state (Latour, 2005).

Given that the UK OGD Web activity state had reach a state of translation that appeared to be stabilised, the following section will examine the new kinds of Web activities that emerged from the mobilised OGD network and the processes used to ensure that OGD remained in a stabilised state of translation.

8.3.1 Enabling New Kinds of Web Activities

For UK OGD, the networks were passing through the passage points of the original (and adapted) agenda. Government data was being published locally and nationally, which enabled actors to engage with it in various ways, for software development (developers), for informing reports (national media), for advancing techniques in data management (academics, OKFN), or for improving government-to-citizen transparency and accountability (civic society). The production and consumption of data by actors within OGD was beginning to form its own set of activities which were supported by the 'data instructions' provided by the national media and OKFN network. Citizens were able to obtain, explore and made sense of the data. This was made an even more attractive activity by providing financial incentives for those able to use the data in new ways, which effectively helped demonstrate that the data had value for exposing government transparency and accountability. These activities subsequently became known as 'Data Journalism':

> *"Data journalism has seen a steady increase in recognition among the media community in recent years, but today marks a significant step in its growth as the Global Editors Network (GEN) launches its Data Journalism Awards with support from Google and the European Journalism Centre (EJC). [...] Two first prizes will be given for each section - one for the winning national or international media organisation and one for a regional or hyper-local submission. Each winner will receive US$10,000, which will be shared in the case of a prize going to multiple individuals."* ['The Guardian Datablog', 'First ever international data journalism awards launched', 31st January 2012].

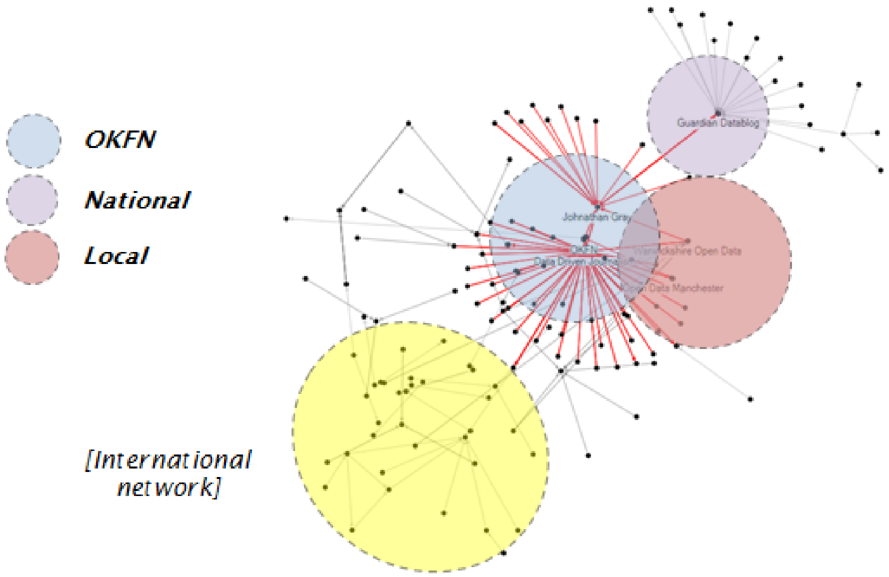


FIGURE 8.21: Twitter communications network (Combined view of Mentions and Retweets) of tweets extracted from the Open Data Twitter Dataset which contained tweets with using the String 'Data Journalism Award, or 'Journalism Award', Overlaid are the different actor-networks represented by associated actors. The yellow overlay represents an international network of actors communicating about the Data Journalism Award. Communications between January and June 2012

By offering citizens and other interested actors financial and social incentives such as the 'Data Journalism Award' provided a method to expose the value of the published government data. As Figure 8.21 illustrates, communications of incentives were shared using Twitter, and the announcement of the Data Journalism Award accounted for 6% of the 129,000 #opendata tweets made between January and June 2012. Tweets of this content also accounted for 3 of the top 5 most diffused messages during this time period and were retweeted by actors that were previously not part of the #opendata communications channel. The activities of data journalism drew upon a network of actors who were not directly involved with the OGD agenda:

> *"RT @jwyg: First ever international data journalism awards launched: http://t.co/8GFqTJ5E (via @datastore) #opendata #ddj #dja"* [Twitter Data Steam, 'ddjournalism', 20th January 2012]

> *"Official Google Blog announces the launch of the first international Data Journalism Awards: http://t.co/IbR69mYr #ddj #dja #opendata"* [Twitter Data Steam, 'ddjournalism', 20th January 2012]

> *"International Data Journalism Awards Recognize a Crucial Field http://t.co/qwTifT9X #OpenData #OpenGov #Gov20"* [Twitter Data Steam, 'digiphile', 9th February 2012]

The association with established networks was an interessement device for the translation of OG, the activity of data journalism demonstrated that OGD was becoming a Web activity that was worthy of investment from other networks, which indirectly acted as a catalyst to enable new actors to engage with OGD:

> *"Impressed that there have been more than 700+ entries to the Data Journalism Awards. Some interesting projects. #ddj #opendata"* [Twitter Data Steam, 'jwyg', 14th April 2012]

Alongside the activities of the data journalism award, the 'Data Journalism Handbook' was established using the knowledge and network outputs created by the developers, OKFN, and national media networks. As part of the activities that surrounded the launch of the handbook, the online presence of Data Journalism became part of the #opendata communications channel. As Figure 8.22(a)and Figure 8.22(b)illustrate, the involvement of 'ddjournalism' functioned as an intermediary for connecting OGD networks (Figure 8.22(a)) with communications concerning Data Journalism, and also sharing information and news (Figure 8.22(b)), including highly shared messages such as:

> *"Help us finish the Data Journalism Handbook! Read here how you can contribute: http://t.co/uXkcqmaf #ddj #opendata #datajournalism #dataviz"* [Twitter Data Steam, 'ddjournalism', 6th February 2012]

As Figure 8.22(c)shows the launch of the data handbook and the announcement of the journalism award contributed to an increase in online activity and were both within a short timeframe, and 60% of actors communicating about the handbook were also communicating about the data journalism award. By providing news coverage of both of these events, the OGD networks offered another mechanism to entice actors to participate in Data Journalism activities. However, actors interested in Data Journalism were not only new actors, but also actors already aligned with OGD, including actors within the citizen network. These actors were already involved in OGD activities, and showed specific interest into using local government data for reporting and investigative purposes; as described by an actor within the local government network:

> *"Their [Citizens] interest is in what I would call as data journalism, and we've seen a lot of it recently like in the Guardian, and also related to that are 'hyper local' data journalism."* [Adrian Slatcher: 64]

In addition to the data being made available, the launch of the Data Journalism Handbook provided citizens with knowledge, training, and simple technologies which were previously not available. The tutorials and tools produced by the developers, national media network, civic society and OKFN provided citizens with the opportunity to engage with the published datasets in order to fulfil the goals originally set:

> *"The ambitions that have been met around armchair auditors that have taken the position of real auditors who have been checking up on authorities and other organisations"* [Jos Creese: 87]

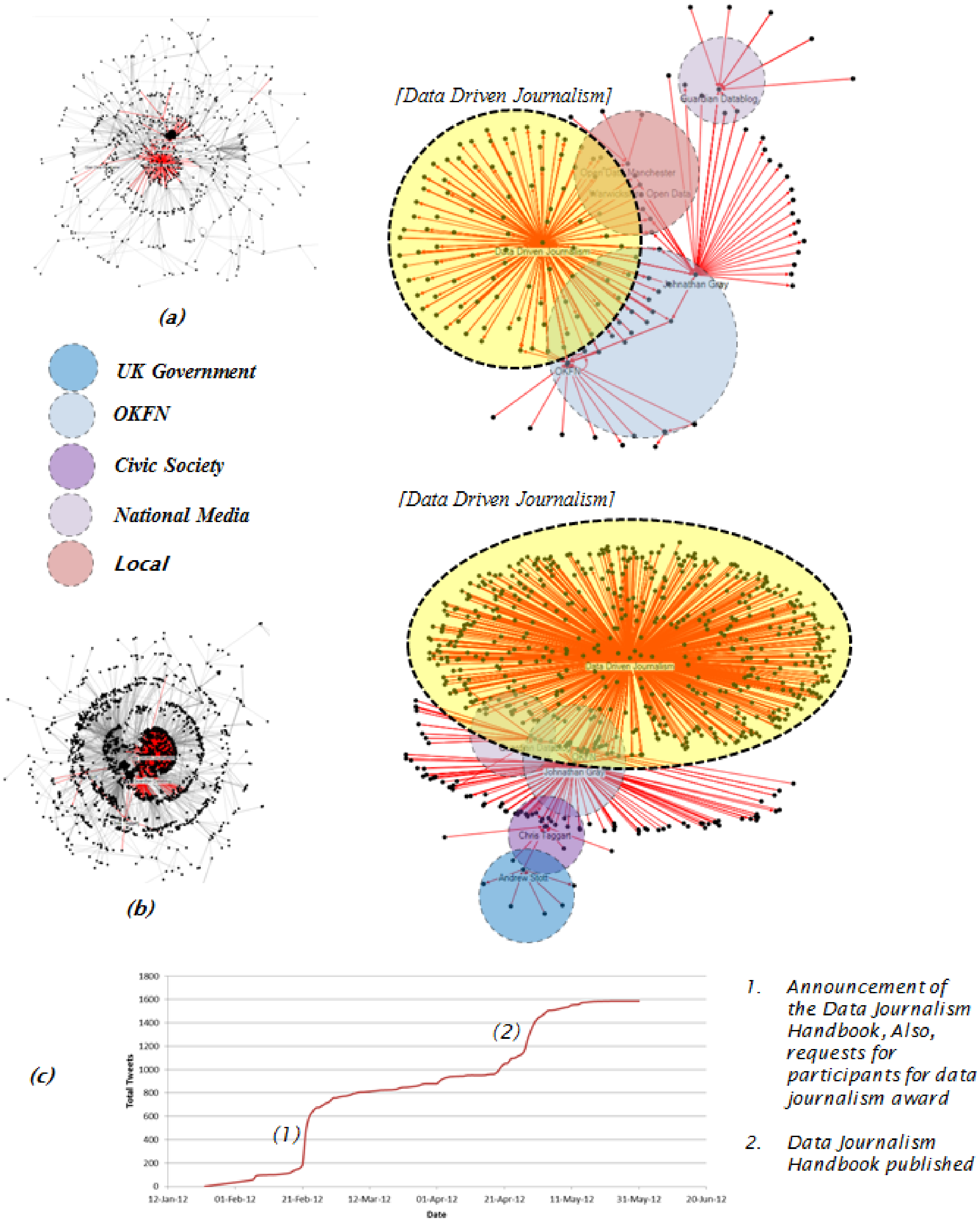


Figure 8.22: Twitter communications network of (a) Mentions (b) Retweets extracted from the Open Data Twitter Dataset which contained tweets with using the String 'Journalism Handbook', 'Data Journalism Handbook', or ''DDJ Handbook' (based on Interviewees' reference to these keywords). Overlaid are the different actor-networks represented by associated actors, and also the identification (Yellow overlay) of the Twitter User 'Data Driven Journalism'. (c) The total volume of tweets associated with this dataset during Jan June 2012

The 'armchair accounting' [Chris Thorpe: 310] activities by the actors in the citizen network was the emergence of a new social process not possible before the availability of Open Government Data. As Figure 8.23 illustrates, providing data, technologies and tutoring, the OGD networks helped citizen's gain the skills required to independently examine and scrutinize government data. These new skills enabled citizens to engage in new forms of activities which fulfilled the original OGD agenda of improving government transparency and accountability across national and local government. This was reducing the layers of interaction between government and citizens and as described by a local government actor:

> *"They [citizens] can see that information straight away, and they don't have to take my word for it"* [Gareth Anderson: 99]

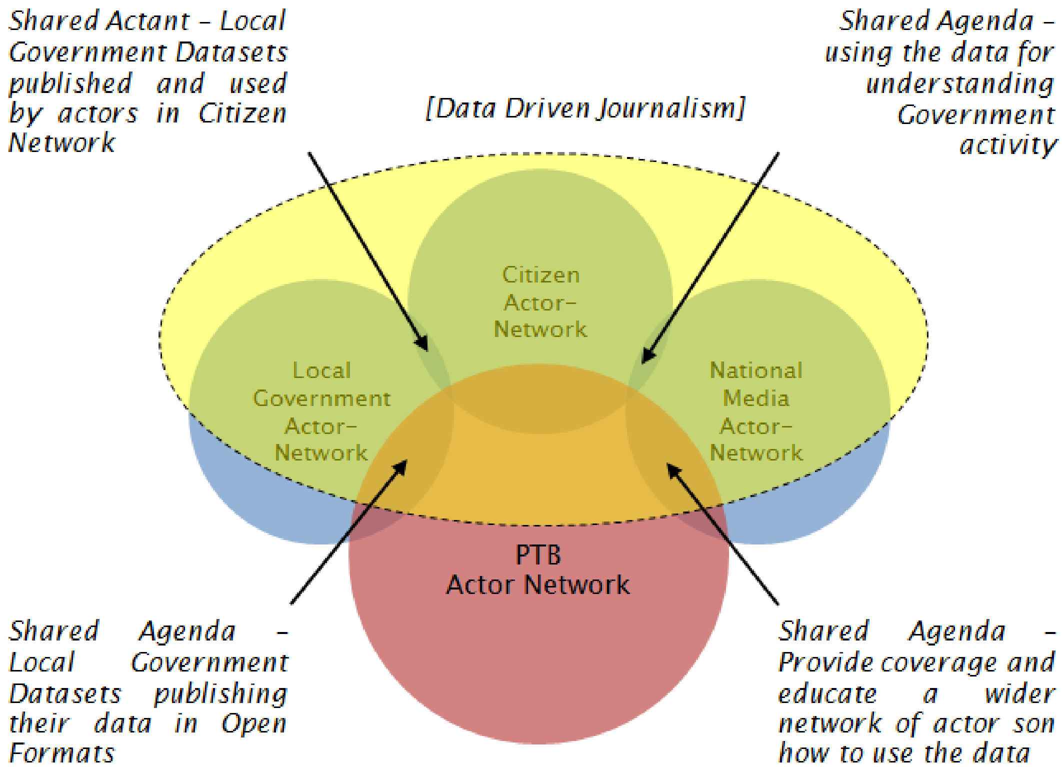


FIGURE 8.23: Illustrating the alignment and strengthening between the actor-networks within UK OGD, according to their shared agenda and actants. The Data Driven Journalism actor-network encompasses the Local Government, Citizen and National Media Actor Network, sharing a common agenda and actants of using published Government data and reporting about transparency and government activity

8.3.2 Reinforcing the Stability of Open Government Data

Stability of OGD can be considered as the achievement of the OGD agenda; social processes were agreed, technologies were developed, and government data was being published at the local and national level. The activities of data journalism and armchair accounting made use of this data, adding value and illustrating its capabilities, which in turn reinforced the stability of the OGD Web activity. OGD actors were able to celebrate their successes, and also lay out plans for the future of OGD to ensure that it remains a stable Web activity:

> *"Our Open Data commitments cover health, education, transport, crime and justice - as well as central government spending. We've already released over 35,000 datasets on data.gov.uk - the largest resource of its kind in the world."* ['Cabinet Office', 'Information Commissioner's Conference Francis Maude keynote speech', 6th March 2012]

> *"The Digital Age has made transparency an irresistible, unstoppable force. It is easier for the public to demand data - easier for governments to provide it. Easier for citizens to hold Governments to account - on a day to day basis not just at election time. It is easier for people to use data to inform their choices [...] we now have regular publication of central department spending data over £25,000 and local government spending over £500. [...] Civic Dashboard is a field-leading example of how open data can give local people an online, accessible way of understanding how their city is doing and how government is responding to their needs and wishes."* ['Cabinet Office', 'Open Data Innovation Community - Francis Maude speech, 14th March 2012]

As identified before, online communications using Twitter helped spread the announcement of OGD success and future plans. As Figure 8.24 illustrates, the UK government's announcements was shared to the network of actors that the academics, civic society and national media networks had obtained during the translation process. Tweets such as those by the 'cabinetofficeuk' and 'Nigel_Shadbolt' were extensively retweeted (representing the most retweeted messages), and 28% of these actors that were part of the diffusion of the selected tweets were not previously part of the #opendata communications:

> *"Full text of Francis Maude's #odinnovation speech plus #opendata press release available here: http://t.co/T4JpYxA".* [Twitter Data Steam, 'cabinetofficeuk', 14th March 2012]

> *"Francis Maude's commitment to UK supporting Innovation via Open Data here http://t.co/jmngLOhb #opendata #digitalLON #ogp"* [Twitter Data Steam, 'Nigel_Shadbolt', 14th March 2012]

> *"Francis Maude: public data owned by the citizen not the state http://t.co/maFuQtnD #opengov #opendata #transparency"* [Twitter Data Stream, 'cabinetofficeuk', 19th April 2012]

The commitment for future OGD plans was reinforced by the PTB with the publication of the 'Open Data White Paper: Unleashing the Potential #opendata' (HM Government 2012). This report, written by the actor-networks within the PTB, contained a statistical dialog of the current success of UK OGD, along with the future agenda of data publication, use and engagement.

The 'Open Data White Paper' embodies a multitude of agendas, interessement devices (statistics of outcomes), and commitments to future OGD goals. As part of this, it helped PTB actors document the success of their approaches to OGD, including the academics 5 Stars of Open Data (pg. 24) and the development of technology due to the efforts and commitment of civic society and developers (pg. 24). The report also enabled government to reconfirm the importance of the activities of those that supported the OGD agenda, both socially and technically; specifically the "civic-minded developers and related communities" (HM Government 2012). In addition to this, The 'Open Data White Paper' readdressed the goals that were yet to be achieved; and as a local government actor described, helped ensure that enrolled actors stay committed towards the current and future agenda of OGD:

> *"As the governments just published this White Paper, I'm talking to colleagues across authorities at the moment, particularly colleagues involved*

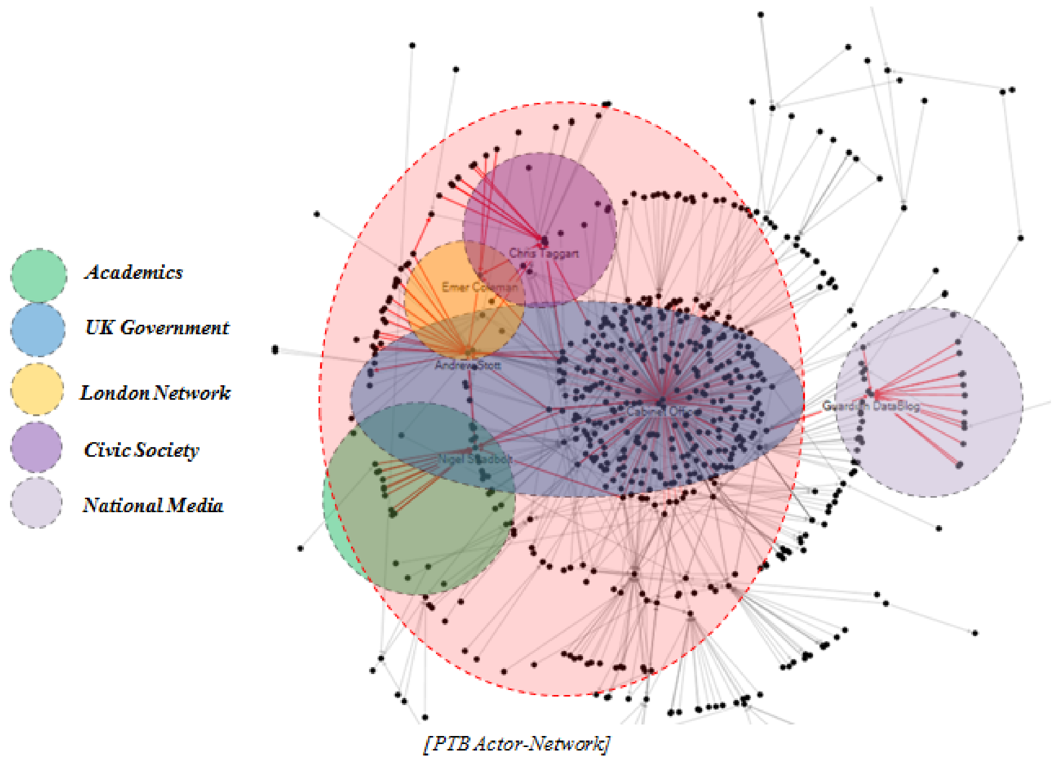


FIGURE 8.24: Twitter retweet communications network of tweets extracted from the Open Data Twitter Dataset related to the Speeches given by Francis Maude regarding the current progress of OGD in the UK. Tweets extracted using the query 'Francis Maude AND speech'. Overlaid are the different actor-networks represented by associated actors. The Red overlay represents the actor-networks associated with the PTB actor-network.

> *with our local enterprise partnership, and asking the question, is this something that we should be pursuing more actively."* [Philip Jones: 181]

Online communications also helped disseminate this news. However, unlike previous uses of tools such as Twitter, the communication technology was being used to facilitate the engagement and interaction with new and existing actors, rather than just a broadcast mechanism to announce news. Tweets such as the 'cabinetofficeuk's' shown below asked for feedback and participation, providing a social and technical process to enable actors to engage.

> *"Do you have a question on the #opendata white paper? Please post - @tkelsey1 will be answering Qs shortly here: http://t.co/TH7jfYA5"*
> [Twitter Data Stream, 'cabinetofficeuk', 28th June 2012]

Due to the emphasis on engagement rather than broadcasting, 83% of the total tweets made related to the 'white paper' contained 'mentions' to other users, which is an indication that actors were communicating rather than just sharing tweets with each other. As Figure 8.25(c) illustrates, these communications peaked during the government's speech on the 'white paper', and as Figure 8.25(b) shows, this was across many networks, who were conversing about the government's commitment to OGD with tweets such as:

> *"Tim Kelsey (@tkelsey1) and Francis Maude commit to opening up postal address file at launch of #opendata White Paper"* [Twitter Data Stream, 'Will_Tanner', 28th June 2012]

> *"Very happy to note UK Gov's increased commitment to open standards coming out of today's White Paper launch incl in NHS. #opendata"* [Twitter Data Stream, 'philarcher1', 28th June 2012]

> *"People need a forum to shout back at government. Yes! @bengoldacre #opendata White Paper Launch"* [Twitter Data Stream, 'haideebelly', 28th June 2012]

The 'white paper' also provided an opportunity for the aligned networks, including civic society, developers, OKFN, and academics to introduce a new agenda and set of activities that would help translate UK OGD beyond its current capabilities, expanding its agenda towards new actors and new activities:

> *"The Open Data Institute (www.theodi.org) will demonstrate the commercial value of Open Data and work closely with the public and private sectors as well as academia in developing its exploitation. It will be led by Sir Tim Berners-Lee and Professor Nigel Shadbolt."* (HM Government 2012)

The Open Data Institute (ODI) was produced through the translation of OGD, and represented the extensive list of activities and goals that the actors involved in OGD had achieved. As the Cabinet Office article announced, the ODI is "the first of its kind in the world" and was led by the academics and had the commitment and support from the

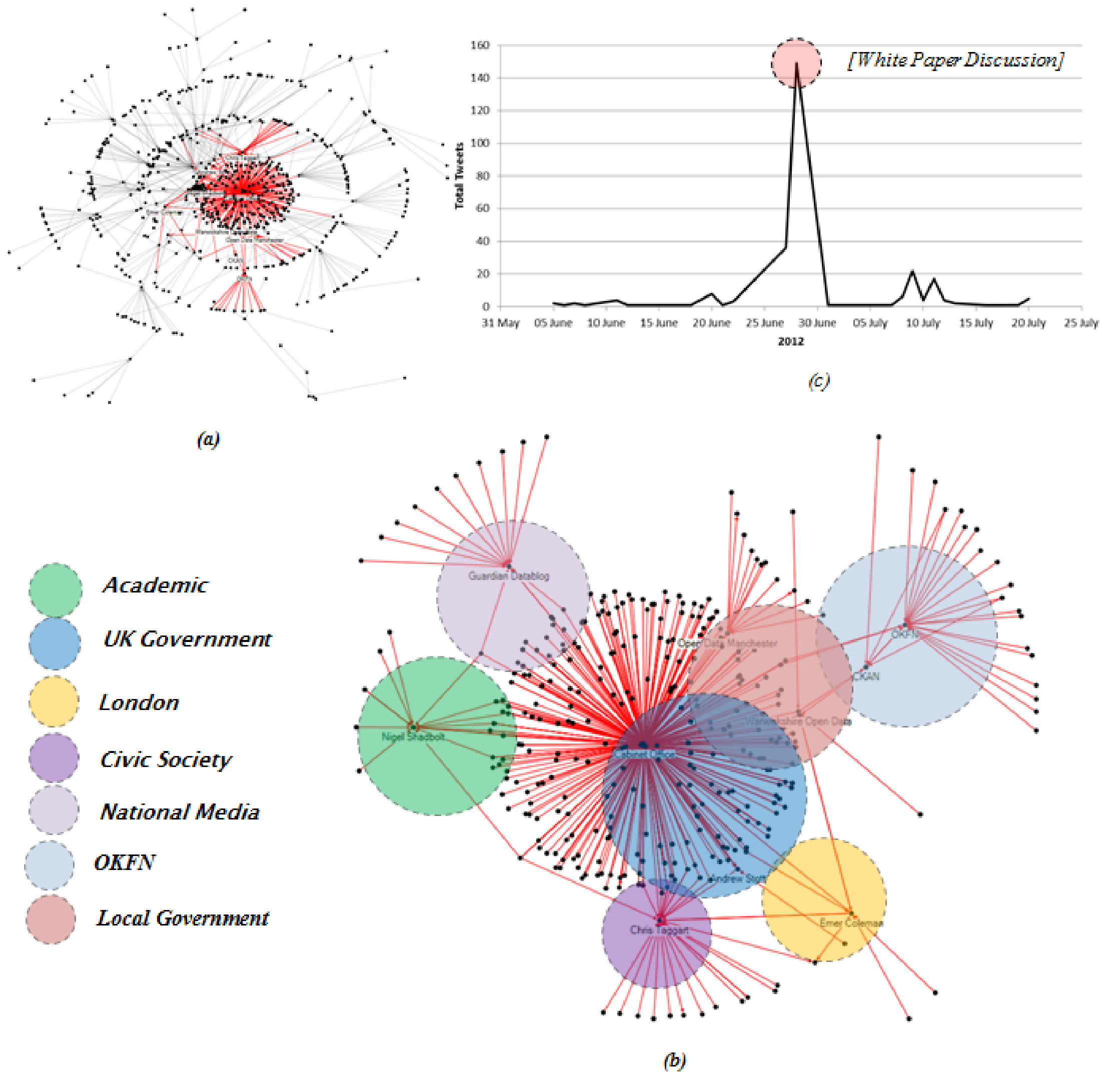


FIGURE 8.25: Twitter communications (retweet and mentions) network of tweets extracted from the Open Data Twitter Dataset related to 'white paper' published by the UK Government. Tweets extracted using the String 'White Paper'. Overlaid are the different actor-networks represented by associated actors. (a) represents the entire network communications, (b) represents the Twitter accounts associated with the different actor-networks identified in the previous stages, (c) a time series of total number of mentions during the announcement of the 'white paper'

UK Government, which was demonstrated by the £10 million funds provided ['Cabinet Office Blog' 'Plans to establish Open Data Institute published', 22nd May 2012].

Furthermore, the announcement of the ODI in the 'white paper' helped establish and share the agenda with the actors that were already aligned with OGD, such as the data producers, the developers, and civic society. This was also the case for engaging with new networks such as "Open Data technologists and entrepreneurs" that were problematised by the ODI as actors who were able to set up businesses and find commercial use and value within the data. However, despite a new venture of OGD, the academics were still committed to ensuring the development of "open standards and commissioning research relevant to Open Data exploitation", with the vision to make UK an "international leader in Open Data" (pg. 16).

As Figure 8.26 shows, the news of the ODI caused a rise in online activity during the Cabinet Office's announcement and also during the launch of the 'white paper', and via other sources of coverage and publicity by actors within the OGD network. The activities of the ODI was a catalyst for new actors within the national media network to become interested and report the OGD agenda. Web-based news publishers such as 'Wired UK' was an example of this, providing publicising and reporting the activities and future plans of the ODI, ['Wired UK', 'Professor Nigel Shadbolt outlines plans for Open Data Institute', 22nd May 2012], and subsequently reporting the development of the ODI actor-network, which caused the spike in activity shown in Figure 8.26(c) ['Wired UK', 'CEO of Open Data Institute: we need to make open data non-threatening', 8th November 2012].

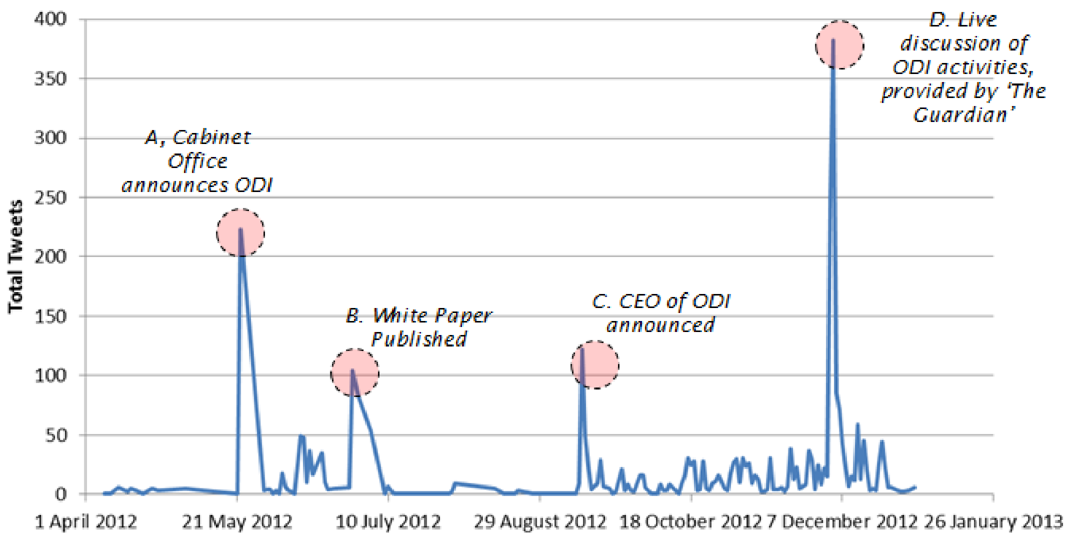


Figure 8.26: A time series of total number of tweets during the announcement of the 'white paper'. Tweets extracted from the Open Data Twitter Dataset related to Open Data Institute (ODI). Tweets extracted using the String 'ODI' or 'Open Data Institute'. Overlaid are the events related to the spikes in tweets made

The formation of the ODI was more than just the inscription from the actor-networks involved in the previous activities that has led to the current state of OGD. The ODI became a device to ensure that the current actor-networks committed and mobilised with the agenda of OGD continue to remain focused on their goals, and it was a chance to widen the current network to a wider set of actors, aimed at aligning new businesses, developers, and citizens with the goals of the OGD agenda.

> *"I think things like the Open Data institute will have a really important role in convening people with special interests and discussions"* [Chris Thorpe: 272].

8.3.3 An Established Web Activity: Open Government Data

The Web activity of OGD was established, and it was operating in state of stabilised translation. As a result of this stability, new forms of Web activities emerged along with their own agenda and translation process.

This section has demonstrated the strengthening of ties between the citizen and national media network, and how their interests with the OGD network helped establish 'data driven journalism' and 'armchair accounting' as new kinds of activities which build upon OGD. The new social and technical activities were part of a co-constructed relationship, and through the interactions and of the actors, the agenda of OGD was altered as was the activities that were emerging. These new activities and the evolving agendas laid the foundations to enable more participation from citizens, which in turn helped fulfil the changing goals of OGD.

The publication of the 'white paper' and the subsequent announcement of the ODI was the confluence of the different agendas of the actor-networks coming together to form a new common set of agreed goals and passage points. As Figure 8.27 shows, the OGD network has evolved, and the actions and goals of the ODI (actors) become part of the OGD network. The number of actors involved had increased, and so had the activities supporting OGD.

Given that the networks that are active in OGD are mobilised and stable, as long as they continue to remain stable and committed and aligned to OGD, they become black-boxes within the network of OGD and to the networks that enrol in the future.

In summary, discovered in this section:

- Networks which operated as stabilised became black-boxes which enabled new networks to build upon their successes and outputs. The stability of the OGD Web

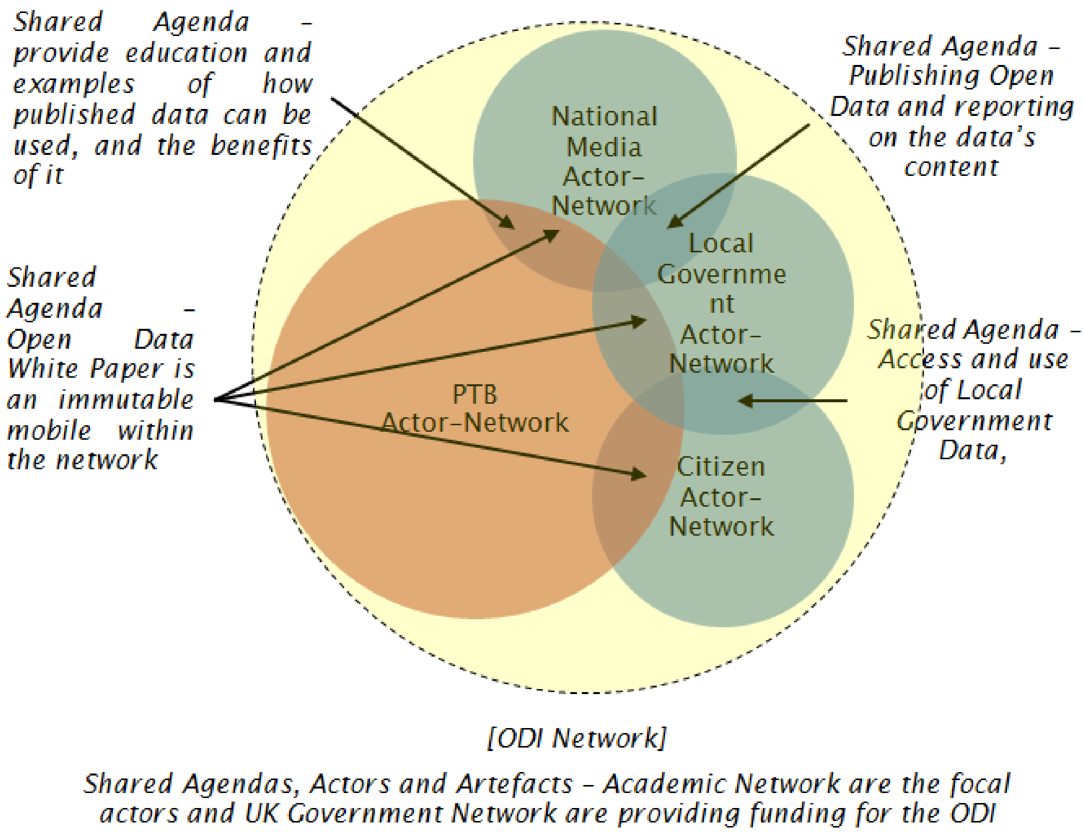


FIGURE 8.27: An illustration of the ODI Actor-Network, including the addition of the Citizen Actor-Network. The ODI Network represents all the actor-networks with a shared interest and set of passage points in alignment with the agenda of the 'white paper'. Note: Not represented in this figure is the London Network, as this has not translated along the same path as the other actor-networks in the UK OGD Network, there are no shared passage points in the 'white paper', nor do their activities (social and technical) correspond to that of the other networks

activity enabled new kinds of socio-technical activities, including 'data journalism' and 'armchair accounting'

- Reinforcing the stability of a network provided a way to ensure the aligned actors remained committed which helped form a new agenda and enlist new actors
- Ensuring stability can be achieved using social and technical mechanisms to ensure as many actors as possible are aware of the activities and new agendas being set.

8.4 Summary

This analysis conducted over the last two chapters has demonstrated the emergence and translation of a Web activity from a problematised state, to a functioning, mobilised set of actor-networks, that converged and configured around a shared set of interests, agendas and goals. By unpacking the evolution of the OGD Web activity through an analytical lens supported by mixed methods and framed by Actor Network Theory, the actors associated with the OGD activity were exposed, as were the social and technical processes and interactions that enabled OGD to translate.

By exposing the social and technical processes involved with the translation of the OGD Web activity, the growth in terms of the associations between actor-networks was revealed. Tracing the activities of the actor-networks at the granular level revealed complexities beyond a single interactive process of translation. What has been shown is a Web activity that has grown as a result of multiple actor-networks translating and interacting around a shared agenda. The evolution of the Web activity is the manipulation of this agenda plus the commitment and stability of enrolled networks; and underpinning this were the socio-technical processes that helped forge associations between actors.

The analysis has also unpacked the co-constructive relationship between technologies and humans, revealing their interactions and impact they have on the translation of an agenda. By using mixed data sources this understanding is further developed by exploring the relationship between the online and offline activities, exposing how technologies facilitate human activity, but also how they are shaped by the outcomes of human action. The findings have highlighted that understanding the translation of an agenda via socio-technical processes is a complex, multi-staged, and multi-network process; and how the stability of networks result a black-boxing process, providing a platform for new activities to emerge and translate.

By considering the findings of this analysis, the next chapter will readdress the research questions, specifically the implications for understanding the growth of the socio-technical Web.

Chapter 9

(Re) Integrating the Macro and the Micro

9.1 Introduction

The previous chapters have presented an analysis of the UK Open Government Data case study and revealed the social and technical processes that enabled it to emerge, grow and become successful. The analysis has pushed beyond an understanding that aligns itself towards understanding just Web traffic or social interaction. In contrast to existing literature, this thesis is now in a position to show how the Web's growth is more than technical innovation, or human interaction; it will also demonstrate that it is not sufficient to only describe the Web as a socio-technical or co-constructive phenomena; because this shorthand does not adequately capture the interactions of human and technological activity.

This chapter will draw together the findings of the analysis of UK OGD to expose the processes that enabled a disconnected network of actors to form and grow into a new kind of Web activity at the micro level. Developing the analysis the chapter shifts to a macro level perspective to discuss the implications of the findings for understanding the growth of the Web in broader terms. By taking the processes discovered in this case study that led to the translation of UK OGD, this chapter shifts in emphasis to understand how they relate to the Web more generally, and how they can be used as a set of mechanisms to offer a new way to conceptualise and interpret the growth of the Web.

To begin with, the following section will examine the findings of the UK OGD Web activity analysis, and consider the socio-technical processes that emerged due to the capabilities and the development of technology, and engagement and response of humans.

9.2 Implications for Open Government Data

The growth of the UK Open Government Data Web activity has been analysed to explore the underlying socio-technical processes involved in the emergence and formation of this Web activity. The analysis drew upon a combination of quantitative and qualitative data sources, and exposed the interactions between actors and actants which represented the 'socio-technical' processes that enabled the Web activity to translate. The analysis also revealed the complexity and multi-network structure of Web activity, which is due to the operation and interaction of different technologies and human interests, and the production and reflection of network artefacts. It is a result of these activities that the OGD Web activity achieved its goals such as publishing data, and enabled new forms of Web activity such as 'data journalism' to emerge.

By assimilating the findings of the previous chapter, this section will discuss how OGD Web activity was able to successfully translate, gain actors, produce outcomes and achieve its goals. These properties which represent the functionality, operation and success of the UK OGD Web activity are then considered in terms of their characteristics which embody the socio-technical processes of how the Web activity translated.

9.2.1 Socio-Technical Interaction and Agency

By drawing out the social and technical interactions using an ANT inspired approach the analysis demonstrated what it means for the Web to be socio-technical, produced as part of co-constructive process. The empirical analysis has shown that a co-constructive relationship between humans and technologies is more than just a way to identify the interplay between each other; co-construction is embodied in the way technology is used, how technology evolves and how it changes and is changed by human activity. It is embodied within the relationship between online and offline activities, but also more subtly; the role that technology played in shaping the agenda of the growth of UK OGD.

A clear example of the socio-technical growth of the network was in the facilitation of communications by the use of online communication platforms such as Twitter and other communication technologies. These helped humans to discuss issues of OGD with each other, and also provided actants with the ability to share information with each other to a large network of individuals. As examples showed, the use of specific features such as 'retweeting' enabled individuals associated and aligned with the goals of OGD to share both messages and content (URLs, images, videos) with their own network of associations, and in many cases, these actors were part of a much larger network, thus providing the environment for information to diffuse at scale. However, technologies such as Twitter did more than just allow people to communicate with each other; they had a profound effect on the operation and direction of the OGD growth.

Technologies like Twitter provided alternative methods of interaction, which effectively had a direct impact on the functioning of a network; without the use of these technologies, would the UK OGD Web activity be able to achieve its goals in the same amount of time, and would the agenda of UK OGD be as diverse in terms of actors and goals, and would the network configuration be the same? There are also questions regarding the development of technologies like Twitter, such as the added functionality to embed media or automatically shorten URLs to occupy less character space. As the analysis demonstrated, embedded URLs became increasingly popular within the tweet helping the community share more information, directing traffic to government data portals, thus providing a mechanism for more individuals to engage with them. Whilst these features were external to the translation of UK OGD, their impact was not. Take for example the creation and use of the #opendata and #datagovuk hashtag, these were a critical component to enable a community to form around the communications of Open Government Data, offering a dynamic forum of information exchange and up-to-date news of the activities of the community. The adoption and use of Twitter hashtags represents a co-constructive process between technological functionality and social interaction. As the analysis demonstrated hashtags provided a communications channel that was 'always on' and 'open' for all to contribute and participate, there was no invite required and individuals were free to communicate independent on time or location. These self-managed communication channels helped contain and shape the agenda of

UK OGD, and provided a way to monitor the community. By taking part in communications using these hashtags, it offered new individuals to become part of the Web activity and have the potential to reconfigure the network.

These technologies were also affecting how individuals could engage with each other, online, and offline. As analyses have shown, an individual's position (or role) within the online network was an important characteristic within the use of online communications technologies; the more associations that an individual obtains across multiple networks, the more likely that information could reach a different collection of individuals. This in itself changed the way individuals could communicate, and contact others; unlike other technologies such as the telephone, it is possible to diffuse information to an unknown population of individuals, without needing a priori knowledge of their details. Yet it is also important to consider an individual and their activities online and offline, and both affected their use of the technology, and their ability to interact with others.

Although technologies such as Twitter are an authority and hierarchy free platform for all users, in multiple cases, an individual's 'bridging' role within the online communications network (authority or amplifier of information) was influenced by their role in their offline network of associations. The individuals that had a multiple associations with different networks beyond online communications were just as important for the growth of UK OGD. These offline bridges enabled new individuals to become aware and aligned to the agenda of OGD; which helped increase their online networks, gaining more associations, which reinforced their role as a 'bridge'. Consequently, by harnessing their offline and online associations with the previous networks they were involved with, these individuals were in the position to act as an intermediary to draw together different network agendas and share network artefacts, both which helped the growth of UK OGD. Furthermore, Web communication platforms provided a way to surface and uncover individuals who might be potentially useful for aiding process of translation, whether this was in terms of developing software, campaigning for access, or simply supporting the agenda.

Inscribed by human actors, the features and constrains of technologies were part of the interplay between online and offline activities. Take for instance Twitter's technical architecture, it provided a mechanism to enable highly shared information to rise to the top of conversations, and offered recommendations to connect with the individuals responsible for it. Metrics such as an individual's number of friends or followers provided an indication of an individual's connection and potential influence within a community (Anger & Kittl 2011) (Rao et al. 2010). As shown in the analysis, actors that bridged networks tended to have a much larger proportion of followers compared to others in the network, thus making it possible for these individuals to share information to a large network of individuals.

The interpretive flexibility of technologies such as Twitter represented a dynamic socio-technical relationship; the technical functionality of the technology was not only being used beyond its original scope, but the technology was also changing the way the human actors were operating. Equally, the adoption of these technologies by the UK OGD Web activity was not without consequence, the functionality and their use were also affected. Take for instance Twitter, OGD engaged with it as a tool to share news and communicate as a community, despite being a system previously riddled with spam (Pear Analytics 2009) (Chu et al. 2010) (Hurlock & Wilson 2011). Effectively, the actions of the OGD Web activity reconfigured Twitter for themselves as a technological system which was driven by those seeking context-rich information.

The interplay between humans and technologies extended beyond just the use of Web communication platforms; the features and functionality of data publishing technologies such as CKAN were part of a socio-technical process responsible for Web growth. CKAN's agency extended across all actants within the network: the data and the individuals curating the data, and those wanting to make use of the data, such as developers or citizens. CKAN required them to publish and obtain the data a central location, namely data.gov.uk or the other data portals created; and by doing so, it caused highly distributed and de-centralised network of individuals into working together to produce a Web activity that was operating as a coherent network. Effectively, the technological agency did helped UK OGD to grow as a coherent network, unifying activity, and helping achieve its goals.

Technological agency also affected the growth of OGD in multiple ways, including how the standards and publication formats inscribed by technologies influenced the way further technologies were developed, or how individuals could engage with the data. Technical formats helped shape the publication processes of OGD; they defined how data was made available, and how data could be obtained and what it was capable of. For instance, comparing data.gov.uk and the London datastore, both were inscribed with different features, the former led by strict publishing formats, the later with a much more relaxed set of guidelines. As a consequence of these differences, data.gov.uk and the London datastore operated differently: socially and technically. Whilst data.gov.uk strived for more 'Five Star' data (HM Government 2012), the London Datastore focuses on a limited number of reliable and timely data sets [Emer Coleman: 25]. Ultimately, this affected the social processes of publishing data, affecting both data publication and data use. On the one hand, the availability of linkable data in data.gov.uk provided developers with the capabilities to create applications and technologies with features that data in the London datastore could not provide. On the other hand, these standards meant that data curators, namely government employees, had to be trained to publish this data, and citizens had to learn how to use it.

The features and capabilities of technologies such as CKAN illustrate the knock on effect that result from the inscriptions of human actors within technological design. Through the process of developing a technology, human interests inscribe technological capabilities and functionality. However, once a technology transitions from the controlled development environment to the Web, it is subjected to various socio-technical processes. During this transition, the inscriptions of actors are challenged and tested, exemplified by publication of data formats subsequent to the launch of data.gov.uk, or the adoption of Twitter and the #opendata hashtag as a communication channel. Technology is no longer bound by its original design, its capabilities and functionality is questioned and adapted, and is operated in new and unintended ways. However, it is these characteristics of technological and human interaction which have helped the UK OGD Web activity to grow.

9.2.2 Network Growth and Emerging Activities

Common views of the Web consider its growth and structure as a graph of nodes and edges, as hyperlinks between documents, or as a network of technologies and physical infrastructure. However, as the analysis has shown, Web growth, and by that, the emerging structures of the Web, are far from being only a technical discussion; the formation of the Web is a complex set of processes which involve interactions of human and technologies, both online and offline. Embedded within the Web graphs are the temporal processes and interactions which a static representation struggles to convey.

However, this growth of the Web can be seen as the emergence of and adoption of processes and practices, which needs to be understood beyond exploring the structural changes of the Web graph. As the analysis revealed, the Web graph could be considered as the digital traces of the activities, interactions, and outcomes of individuals and technologies; in this respect, the Web graph becomes a means to an ends, one of many tools to understand Web growth.

For instance, the Web graph may be able to illustrate the de-centralised structure of UK OGD and represent it in terms of graph metrics, however it fails to capture the socio-technical processes that led to this state. Indeed, the OGD Web activity was driven by multiple diverse networks, which formed a 'top-down, middle-out, and bottom-up' structure [Andrew Stott: para. 170]. The emergence of the UK OGD Web activity was a result of different networks, containing different (and shared) individuals and technologies, each with their own agenda and set of interests. The OGD consisted of individuals within government, academia, civic society communities, developers, and citizens. Some of these networks were already established, some had emerged from the alignment of networks. Although their interests with OGD were for different purposes, all networks shared the common interest of government data, which enabled them to contribute to the eventual translation of UK OGD.

Understanding this requires more than considering the Web graph as analytical tool to understand the structures that emerge from the growth of the Web, it needs an understanding that draws upon the social and technical interactions that are responsible for the outcomes observed. By shifting between perspectives and analytical tools, the analysis has shown how the distributed structure is connected together by individuals who bridge different networks together, becoming gateways for the different networks to communicate and interact. Drawing upon the concept of weak ties (Granovetter 1973), these actors were able to align and form new shared agendas using their position within the network. These different networks were all part of the OGD agenda, and were connected by individuals who participated in activities in multiple networks.

In terms of the successful emergence and growth of the UK OGD Web activity, its distributed and de-centralised structure of networks and individuals helped form and add substance the different parts of the agenda that led to the successful translation of the UK OGD Web activity. Interviews showed that the structure of UK OGD was unlike other government technology projects, it was being led by many individuals rather than being centrally controlled, which often result in project failures due to lack of commitment, technological shortfalls and social barriers to adoption (Heeks 2006). Conversely, the UK OGD Web activity has demonstrated how a de-centralised approach has many advantages: they reduce single points of failure and enable a wider network of individuals to become involved and help achieve the intended goals (Kumar & Best 2006). However, this structure was not just by 'social design', it was also facilitated by the technical architecture and capabilities of the ubiquitous, platform-independent, geographically neutral, Web.

Another issue with the Web graph is that it only provides a static snapshot of the current or end-state of observations; however as the analysis of the UK OGD Web activity has shown, the growth of a Web activity is a dynamic and fluid process. The analysis exposed the layering of different networks and agendas during the growth and involvement of new individuals participating. Initially, not all networks (thus not all individuals) were part of the OGD Web activity; however, overtime, as the agenda developed and more activity occurred, additional networks could be aligned, and new networks and agendas were formed.

The layered structure of the OGD network of activity was a result of the addition of networks during its emergence and growth. OGD began from a small set of networks, and as more networks (and individuals) were aligned, OGD accumulated more goals and outcomes were produced; both technologically and socially. This layering of different networks supported the distributed structure of activity and different agendas, and as more networks became aligned and enrolled, shared agendas were formed, network outcomes were produced, and UK OGD translated towards a stabilised Web activity.

The formation of these layers of agendas and networks directly affected the outcomes of OGD. As a result of the increasing number of individuals involved, the evolving agendas and the production of outcomes, the OGD Web activity led to the emergence of new social and technical practices. This was part of a co-constructive relationship where the inscriptions of a social agenda were simultaneously supported by the development of a technology, of which its inscriptions modified the agenda of the network. Data.gov.uk and the various other datastores are examples of this, and through their stability, new networks emerged, which enabled more OGD activity to occur, and new actors to enrol.

Similarly, as the agenda of OGD was realised through the publication of new datasets and commitment of data providers, new social practices such as data journalism and citizen 'armchair accounting' emerged. The stability of the initial OGD activity acted as building blocks for these new forms of Web activity to translate, and embedded within these activities were emerging technical practices concerning how to use and interpret the Open Government Data. The growth and translation of a Web activity forms over multiple phases, where each phase acts is the building block to enable new actor-networks (or groups of individuals related by a common interest) to become aligned.

From the perspective of the Web graph, these just represent another network of communications, hyperlinks, URIs and Web documents; yet this thesis has shown that these new phases in Web growth are the outcomes of various layers of socio-technical activities, human commitment, and technological advancements.

9.2.3 Serendipity and Competition

A characteristic that weaves its way through the analysis of the OGD Web activity translation process is the contingent, or serendipitous nature of the interactions and associations between different individuals. Distilled from the analysis, the process by which actors become part of the development of the OGD community was the result of events and activities, not orchestrated by a specific agenda or set of actors. Not all network outcomes were planned or expected, and consequently, the translation process of OGD was unpredictable and temporary. Serendipity is the result of the dynamic and ever changing growth of a Web activity, and it affected both the social and technical outcomes of UK OGD. Take for example the emergence of new types of Web activities, 'Data Journalism' or 'armchair accounting', these unanticipated activities led to new social practices, but also new uses of the data and technology that it was not originally designed or developed for. In response to this, existing technologies had to be updated (CKAN, for instance), and policies had to be drafted in order to help maintain control on the published data. As illustrated, these outcomes were unexpected, and caused a resonating effect on the other activities within the UK OGD network. Looking at this from a broader perspective, these findings have implications for understanding the relationship between technical design and social interaction, especially critical for understanding the transition of a technology from the controlled lab environment, to the unknown, uncontrolled, and unmonitored environment of the Web (Berners-Lee 2007).

Taking a critical look at the identification of the serendipitous occurrence of meetings and network outcomes, it is also important to consider the actors, agency and activities that may be hidden from the analysis due to the methods used. As discussed in Chapter 5, methodological decisions have implications of the ability to capture a complete set of activities occurring within the Web activity of OGD; even with a mixed methods, mixed data source approach. Consequently, activities or network outcomes which may appear serendipitous or happen by chance may be the result of unknown or hidden activities. The interactions, processes and outcomes which these hidden actors are involved with may not be captured by the methodology employed within this study. Take for instance the activities which could not be captured, such as the *i*nner workings of Government and the *c*losed-door ministerial meetings, the discussions made between actors during these activities may have a direct impact on the agendas of the UK OGD Web activity, yet they are only witnessed through network outcomes rather than as identifiable activities of the network. In some cases, these type of activities cannot be discovered by the methods used and become events which are hidden from the analysis. Despite this, they have a direct impact on the outcomes of the development of the Web activity and thus must still be recongised. Whilst this is not something that reduces the capabilities of the analysis conducted, as the intentions were to examine the socio-technical processes that led to the emergence and growth of a Web activity rather than the impact of specific events, it is worth noting this as a limitation of the methodology used with in this thesis. This concern also applies to the methods used to select the interviewees, despite using multiple data sources as an approach to identify actors in order to reduce the chances of selecting a specific group of actors, the processes that led to the serendipitous events described by interviewees may have been explained by hidden or inaccessible actors (or actants).

To some degree, the discovery of serendipitous events can be associated to the level of granularity used to observe a Web activity; on the one hand too little granularity reduces the analytical capabilities and increasing serendipity, on the other, too fine a granularity may unpack and explain serendipitous events, yet saturate the study with too much data. It is therefore important to choose a level of granularity that provides sufficient analytical insight and fits the purpose of the study. In the case of the UK

OGD analysis, the granularity provided enough detail to expose the processes that led to the emergence of a Web activity, without focusing on specific events or actors.

Serendipitous processes and outcomes may be hard to articulate and identify, however competition and threats to translation are not. Both were common during the translation of the OGD Web activity; internally between aligned networks, and externally by networks of similar agendas. Threats to translation often result in translation stalling or failing (Callon 1986*a*) (Latour 2005), yet within the UK OGD network, the internal threats encountered (competing datastores, technological barriers) helped define new agendas, and external threats (international OGD initiatives) were responsible for decreasing the development time of data.gov.uk, and increasing the number of published datasets, and contributing government departments. Competition between networks came in the form of technologies and actions of individuals producing outcomes. As the analysis revealed, competition between networks as well as the distributed structure of individuals led to the development of alternative data technologies (such as Socrata, the US OGD equivalent to CKAN, or the London Datastore); however rather than stalling progress, it provided diversity, enabling the diverse collection of networks to continue. As interviewees suggested, the competition between the networks publishing data, the different technologies that supported this, and the different uses of the data helped reduce the timeframe for UK OGD to achieve such a diverse collection of outcomes.

Ultimately, the emergence and growth of the UK OGD was the combination of aligning and mobilising particular individuals, led by a well-developed yet plastic agenda, social process and technical innovation, competition (internally and externally), and the serendipitous nature of human interaction and technological use. Although the translation of this Web activity could not have been predicted, it is possible to distil a number of characteristics that helped it become an active and functioning Web activity. Based upon these analytical findings, in an attempt to reintegrate the Web at the micro level of a single Web activity to the Web at the macro level, the characteristics will be examined in terms of what affordances they provide for understanding the growth of the Web.

9.3 Implications for Understanding the Web and its Growth

In exploring the growth of the Web, this thesis has argued that the Web can be conceptualised as a network of interconnecting and interacting activities which develop via their associations. Given this conceptualisation, questions have been asked about the limitations of current theories and methods in computer science and social science used to understand the growth of the Web. The thesis examined how a socio-technical lens can provide an understanding of the Web in terms of its social and technical growth, and applied this to the analysis of the OGD Web activity.

Through detailed analysis, the Web activity of UK Open Government Data was unpacked, and by using a socio-technical observational and explanatory lens, a complex network of actor-networks, human and technologies, and network artefacts were revealed. By de-punctualising the Web activity, it was possible to see stabilised layers (phases) in this activity, and how new layers of activity emerged from the formation of network outcomes and the development of an agenda.

The activities captured within the study of the UK OGD Web activity were presented as self-contained, and did not take into consideration other external Web activities that may have influenced translation. The observational lens was focused directly at the micro and was ignoring what was happening at the macro level. However, this thesis now wishes to take the lessons learnt from the OGD Web activity, and consider what they might mean for the Web at a macro level. If the aim is the understand how to conceptualise the Web as socio-technical, then at some point a shift in focus, from a single Web activity, to the Web, as a collection of Web activities, is required.

Whilst the current understanding the Web tends to consider it as a quantifiable entity, and its growth can be understood in quantifiable values based on Web graph growth and number of hyperlinks, the analysis of OGD has revealed that Web growth is much more than a quantifiable effect. Through description and explanation, the argument has been made that quantification of a socio-technical Web does not take into consideration the socio-technical processes that occur.

Underpinning the Web's network structures (Web graph) are a complex and dynamic network of human and technological interactions. The Web graph's collection of nodes, edges, hubs and authorities are underpinned by socio-technical interactions. The diagrams that are drawn to represent Web networks hide these underlying processes, as if to suggest that these networks are physical entities that somehow exist in 'cyberspace'. However, these networks are quite the opposite, they are only temporary representations of the associations between actors, the representation of data passing between them, offering only a snapshot of the dynamic nature of the Web (or Web activity) at a point in time. Through a socio-technical lens, the formation of these networks can be likened to the vapour trails or exhaust of the digital traces of actors, and by the time observations are made, they have changed shape, and reconfigured (Latour et al. 2011).

By considering the Web as a network of different Web activities and applying a conceptualisation of the Web's growth based upon these analytical findings, what emerges is a Web which has evolved via the stabilisation and interaction of diverse socio-technical Web activities. In one sense, the Web is an abstract concept, which develops as a result of the continuous stability of previous Web activities, and its structure is the consequence of the associations of actor-networks. The interactions of actor-networks create network artefacts, represented by websites, Web pages, Web content, and Web data. But these networks represent much more than this; the growth of these networks are the result of humans and technologies associated via common interest and goals, which manifests overtime to translate into different Web activities.

The Web does not just 'support' these independent human activities, but the Web's technical capabilities and integration into sociality are extended by humans who are trying to achieve their interests and goals using the Web. For instance, individuals do not just 'do shopping' using a platform called the Web, rather, the Web starts to become an integral part of the way that people shop, the Web adopts new capabilities (e.g. authentication and security protocols and online payment technologies, and design principles such as the 'shopping basket') that enable people to shop. Consequently, the Web challenges previously established 'offline' shopping facilities, which itself react and re-configure. The growth of the Web is a dynamic process that is adopted and adapted by humans (e.g. shoppers, retailers, manufacturers and logistics) to do the activities they want to do and in the process it changes and grows along with the individuals using it.

New Web activities try to establish themselves as part of 'the Web' and in the process of doing so, interact, affect, and re-configure other 'established' Web activities. However this is a dynamic and temporary stabilised process, Web activities emerge, grow and remain operational as long as humans and technologies stay committed, and these activities are likely to change once new Web activities are introduced and established. The 'Web' is defined by the Web activities that become and are becoming temporary stabilised, and as new activities emerge and establish themselves, the Web becomes and is defined by these activities as well.

9.4 Constructing a Theoretical Model of the Web

Based on the empirical findings of the UK OGD analysis and the theoretical position driven by the lens of Actor-Network Theory, this chapter will now describe a theoretical model to understand the growth of the Web. Using the concepts of heterogeneous networks and translation in the analysis provided the basis for understanding the Web in terms of human and technological interaction, and building upon this, what emerged was the socio-technical processes that enabled a Web activity to grow. The processes that the analysis exposed extended beyond the explanatory capabilities of translation, and revealed a development process which involved multiple networks of translation, with different actors and during different points in time.

The following sections will describe the three principles which underpin the theoretical model of understanding the Web and its growth: heterogeneous networks, translation, and phases. These concepts draw on ANT and help explain the complexity and dynamics of the growth of a Web activity and indeed, the Web.

9.4.1 Heterogeneous Networks

The Web is a heterogeneous network, but also the outcome of multiple heterogeneous networks that consist of both humans and technologies associated and interacting with each other. These interactions, which are shaped by an agenda and common interests, produce network outputs. The different heterogeneous networks represent different Web activities, i.e. 'Online Shopping', 'Social Networking', and collectively, these activities are labelled, 'The Web'.

As shown in Figure 9.1, the Web can be represented as a punctualised heterogeneous network of Web activities, which are identifiable by the agenda and types of activities and outputs that the human and technologies are producing. The heterogeneous networks that form are through the associations and interactions of actors, and the associations that form are dynamic and are able to change freely. The associations between actors are formed around shared interests and agendas, and the addition of new actors may cause the agenda to change, and the network to reconfigure.

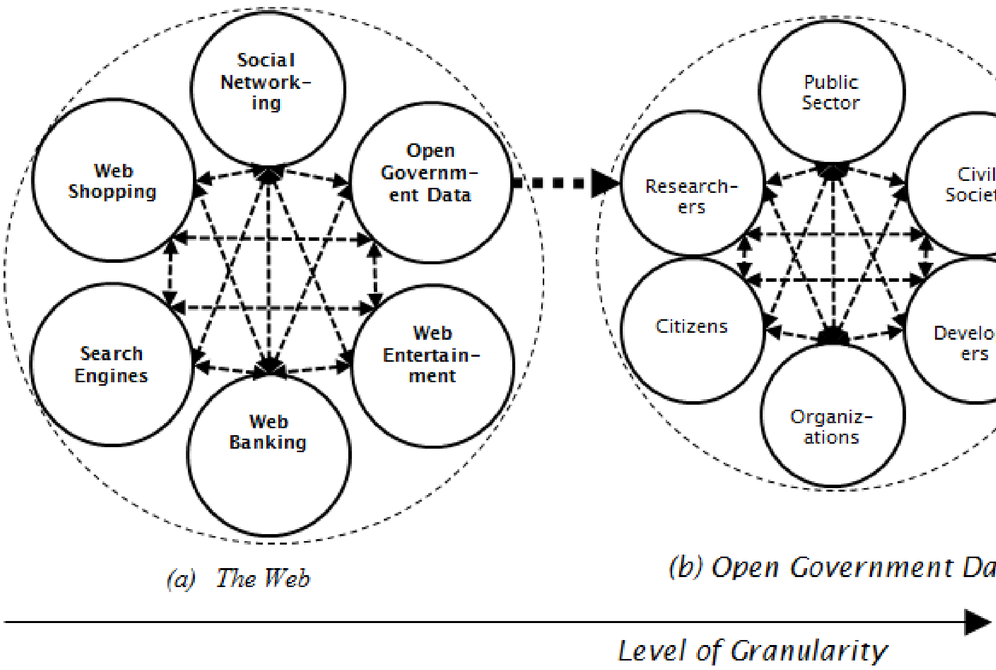


FIGURE 9.1: Illustrating the heterogeneous networks of Web activity '(a)' which collectively can be labelled as 'The Web'. By de-punctualising an actor-network in (a), we are able to examine the actors within specific Web activities; (b) represent the Heterogeneous Networks associated with the Open Government Data Web activities, consisting of the actor-networks associated together via shared actors, actants, and goals.

By de-punctualising a Web activity, as in the case of Open Government Data, the underlying actor-networks are exposed. Choosing the level of granularity helps expose the different networks involved, and their translation is independent of the level of granularity chosen. This addresses the concern of distinguishing between the micro and the macro, where shifting or zooming between the micro and macro are providing two different perspectives of the same phenomena (Latour et al. 2011). To this end, the level of granularity chosen depends on the analysis and insight that is to be achieved.

Theoretically, the Web as a macro phenomenon represents a network of heterogeneous network of Web activities that are interacting with each other. Therefore, observing the Web or observing a Web activity is a matter of re-punctualising the observational lens; the point of focus and perspective changes, the socio-technical understanding does not.

9.4.2 Translation

Translation involves the continuous commitment of the actors already enrolled within the network, and also requires the alignment and enrolment of new actors to successfully achieve the agenda, and for network outcomes to be produced. The heterogeneous networks which represent Web activities all undergo a process of translation towards an agenda which is defined by the actors that form the network. Focal actors are essential in the translation process; these are the actors that initially have identified a problem to solve or an idea to achieve. These actors begin by problematising the actors needed to achieve their agenda, setting obligatory passage points in order for the network to translate, gaining the support of actors; and through these activities, they become indispensable and essential for future translation.

Over time, the addition of new actors affect the original agenda and obligatory passage points, which in-turn change the agency that the focal actor has on the agenda. As a consequence of this, the initial agenda of a network's translation is reconfigured based on the involvement of new actors, reshaping the network and increasing the number of actors responsible for forming the agenda. During the process of aligning new actors, the network agenda adapts, and new passage points are set in order for the new actors to successfully enrol and become engaged with the network activities. As Figure 9.2 illustrates, translation is a process which occurs over time, and it is a fluid and iterative process. Thus, it is possible for a mobilised network to translate to another state depending on the commitment or threats to stability received.

When a network experiences the alignment or loss of actors, its structure changes shape, which affects both the online and offline associations between actors and actants. If new communications and associations are made online, then these may affect the offline interactions between actors; and similarly, establishing offline associations may change the formation of the online network. These associations are not necessarily only between humans, as the impact of technological artefacts (e.g. data as an actor) may have a direct impact on the translation of a network and the structures that emerge.

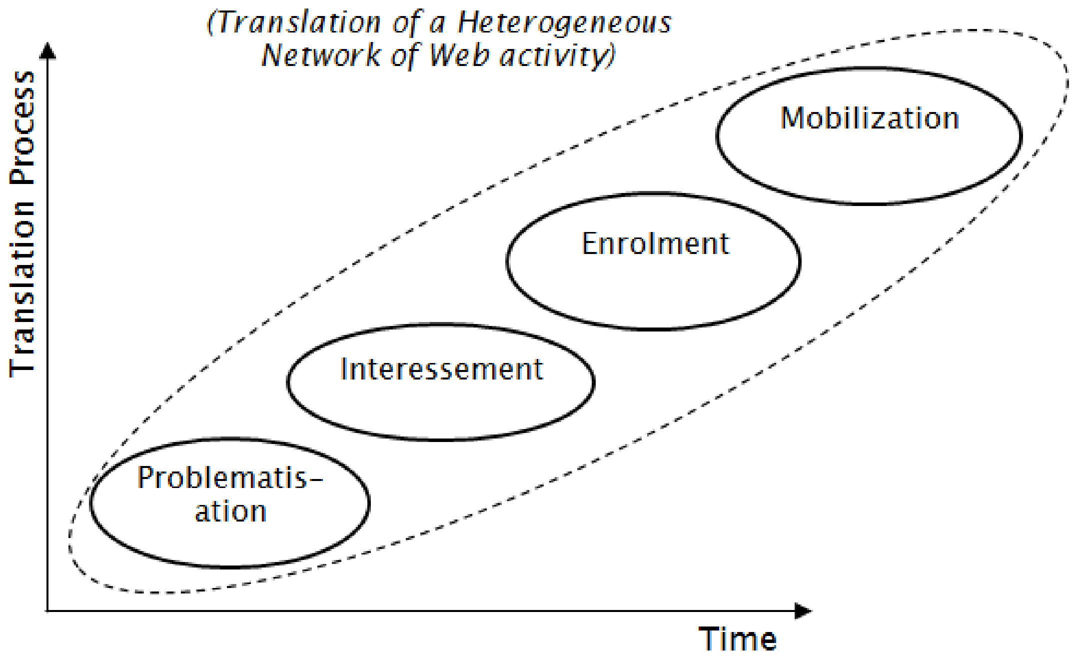


Figure 9.2: Illustrating the second perspective of the conceptualisation of the Web, using the concepts of Heterogeneous networks and Translation

Given that a Web activity is dynamic and constantly changing, its stability (sustainability) is therefore dependent on the commitment of the actors, and threats to the network occur by actors becoming misaligned with the network agenda, or by the translation and enrolment of other Web activities. Whilst threats and competition may affect the success of a networks translation process, as illustrated by the UK OGD Web activity, these activities may also help strengthen the associations between actors within the network; yet this depends on the strength of the inscriptions within the obligatory passage points and the commitment of actors to achieve them.

9.4.3 Phases

The third principle is drawn from the findings of the UK OGD analysis and concerns how the translation and growth of a heterogeneous network (Web activity) is understood. The concept of phases introduces a new theoretical perspective on the translation of a Web activity in regards to the stability and emergence of agendas and networks.

As the analysis has shown, a Web activity is the product of a number of translating networks, and through the associations of shared actors and the mediation of network artefacts, these different networks become aware of each other and aligned towards a common agenda and set of goals. The re-problematisation of an agenda gives way to the emergence of a new actor-network, and the activities of the newly aligned actors set out to achieve the desired agenda, and thus produce network outcomes and artefacts. As these networks translate, and achieve the newly defined agenda, their stability enables additional actor-networks to enrol, alter and re-problematise the agenda and thus the iterative process of forming a new network can occur again.

Different phases in the development of a Web activity represent the different layers of a Web activities history. New actor-networks are constructed from the actors which were part of the previous networks and the emergence of these new actor-networks. In the example of UK OGD, the formation of the Public Transparency Board or the data.gov.uk actor-network represented a new phase in the Web activity. Before the emergence of this network, a number of separate networks were simultaneously undergoing translation towards their own agenda. However, these networks all shared a common interest, and also all of the networks shared common actors (and actants). This, in combination of the previous associations and interactions between networks enabled the PTB to emerge and form a new phase of translation.

The transition from one phase to the next is the result of a number of different network characteristics (which are the result of the actions of the actors); which includes the current state of translation, the commitments and mobilisation of the actors (and the common goals which stem from this), the activity and outputs of the network, the interessement and enrolment of new actors, and also, the threat or competition from other networks. Each of these properties contributes towards to a shift from one phase to another, which can be seen in the analysis of UK OGD. As Figure 9.3(a) illustrates, the common entities that draw together different actor-networks do not necessarily need to be a human; they may be technologies, immutable mobiles, network artefacts, or acts of inscription. Take for example the emergence of the data journalism actor-network, the catalyst within this (amongst others) was the published government data and the technological infrastructure that supported it.

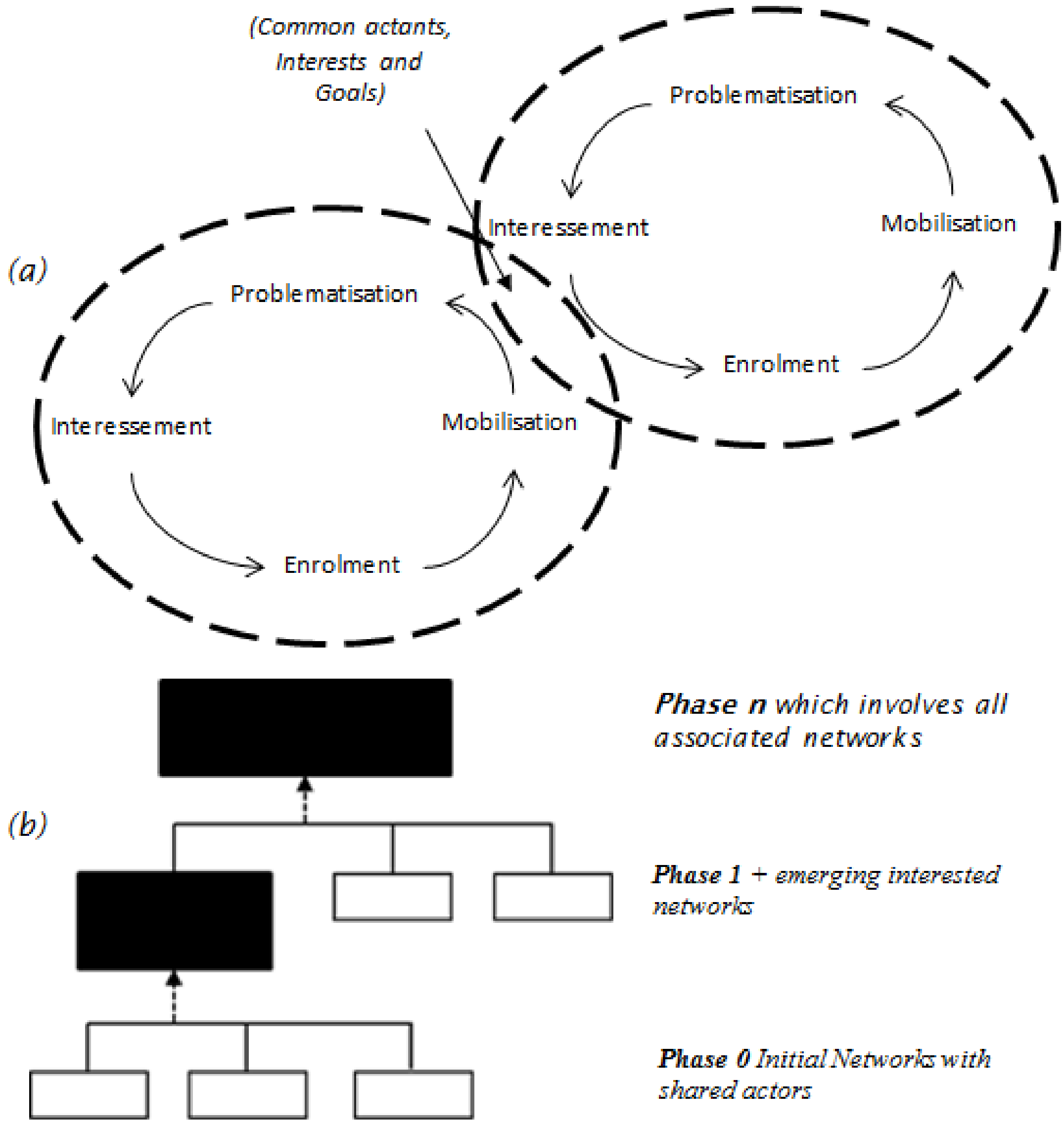


FIGURE 9.3: (a) Illustrating the translation of multiple heterogeneous networks which share common actors and/or actants. (b) Illustrating the perspective of depth in terms of the phases and layers that exist in the translation of a Web activity

Another characteristic of the emergence of a new phase is the continuous stability of existing networks. Within each phase there are numerous heterogeneous networks which are individually translating; however, if these are to act as the building blocks for the emergence of a new phase, they must persist in a stabilised state in order for the subsequent phases to function. This not only requires the actors within the networks to remain committed, but also (depending on the agenda) requires the actors to produce network outputs. Take for example data journalism, the translation of this activity relies on the publication of data from actors within the data.gov.uk and local government network, who themselves rely on the commitment of activities from the PTB network.

The conceptualisation of a Web activity developing in phases also has implications for the translation of a network. Whilst the individual networks within a phase are undergoing translation, there is also an emerging overall translation that is shared across all phases. At each phase, there may be multiple networks which may be at various stages of translation, and the shift from one phase to another may not include all if the networks (and their goals) from the previous phase. However, whilst a network might not be part of a phase change (e.g. the London network when data.gov.uk emerged), their activities may be important for the transition between phases (e.g. the competitive activities of the London network), or may become part of the agenda and activities of the later phases.

The dynamic characteristics of a heterogeneous network mean that translation is a fluid process. From the problematisation and interessement of actors, to their enrolment and mobilisation, the translation of a network is defined by the actors within in, and their commitment towards the agenda. However, a new phase relies on the translation of previous actor-networks, and their ability to remain in a state where they are stable enough to enable new agendas to be problematised.

By adopting the concept of phases within the development of a Web activity, then as Figure 9.3(b) illustrates, the translation of a Web activity can be conceptualised as more than a single heterogeneous network; instead, it is a series of layers which form from new actor-networks emerging from the outcomes of previous activities. These layers represent the underlying social and technical processes and the stabilised networks that enable the translation to a Web activity to occur. From an analytical perspective, observing the translation of the Web at the macro level is the same as understanding the individual networks at the micro level; by taking this perspective, the whole is always smaller than its parts (Latour et al. 2011).

However, translation in its original incarnation (Callon 1986*a*) does not consider the multiple layers or phases that have been achieved during the translation process; instead it is considered as an iterative process of a single heterogeneous network. This thesis refines translation further, and as the analysis has shown, in order to consider the translation of a Web activity, the dimension of depth (layers in Figure 9.3(b)) and phase is essential to expose the various network of actors that have been involved and remain stable.

A theoretical distinction is made here, which challenges the position that ANT champions. Ontologically, symmetry still exists between human and technologies, and so does the notion of agency. However, a new emphasis is placed on the ontological boundaries of an actor-network; rather than taking the position that knowledge, action, and agency all reside from within the network (i.e. there is no a priori assumption of knowledge), the position is taken that the activities of previous networks (and the associations of actors within previous networks) have influence and can be part of new emerging network, as illustrated in Figure 9.4. This relaxes the ontological boundaries of a network, yet keeps intact the epistemological beliefs that ANT adheres to. Despite relaxing the ontological boundaries, this conceptualisation still considers the formation and existence of social structures (i.e. power, society) as a result of the associations and actions of the actors, and that agency of an actor in a network is only through the associations that are made and held in place.

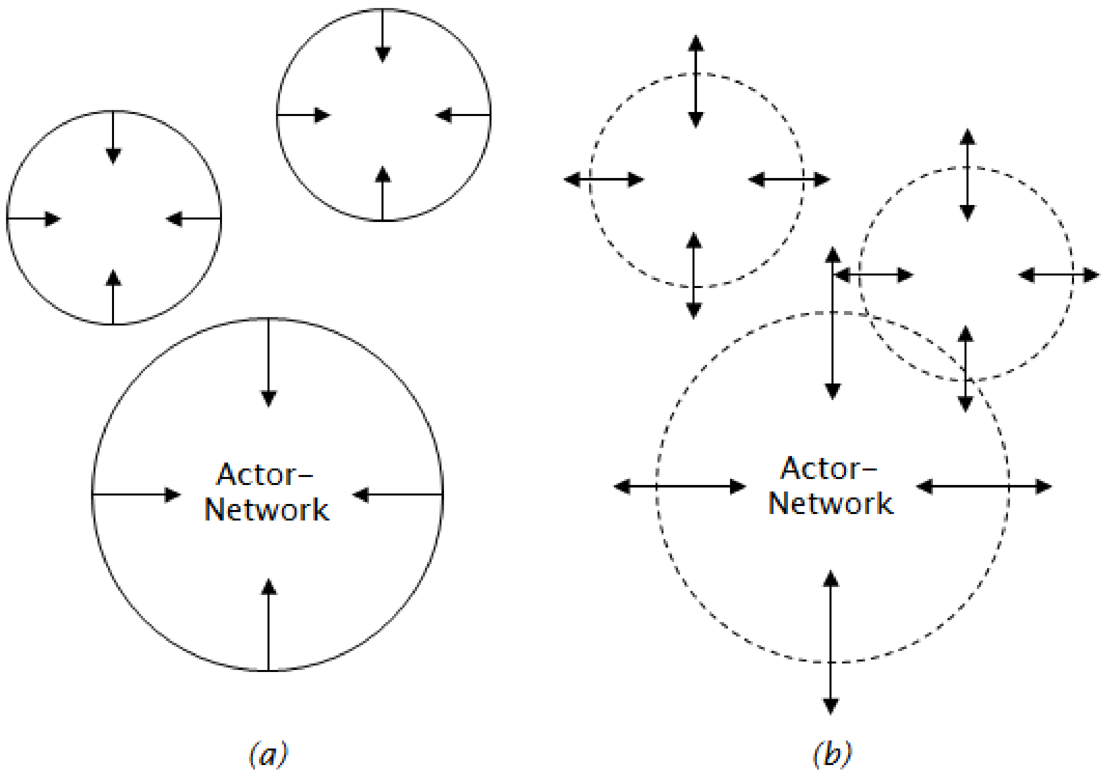


Figure 9.4: Illustrating the difference in ontological boundaries between (a) the traditional ANT theoretical position of no a priori knowledge, and (b) the position taken for within the concept of phases during the development of a Web activity. The arrows represent the flow and of knowledge and agency, the dashed boundary of (b) illustrates the ability to transfer between networks

Without the use of an observational lens that appreciates the multiple-phases within the translation of a Web activity, the emergence of a new network hides the punctualisation of the previous layers activities. The stability of these networks enable a phase to appear black boxed, hiding the actor-networks, activities and commitments which they represent. It is only when a network stops functioning (by the loss of actors, or change of agenda) that these punctualised networks are exposed (Latour 2005). Whilst it may be practical to punctualise these networks for observational purposes, it is valuable to work with a conceptualisation which can unpack the phases when required.

9.4.4 The HTP Model

Based upon the three principles explained above, what has been described is an analytical framework, or more formally; the HTP model Heterogeneous networks, Translation, Phases. HTP is framed around a socio-technical conceptualisation of the growth of the Web. In summary, the three principles argue:

1. The Web is made up of multiple heterogeneous networks that consist of both humans and technologies associated together via common interests and outputs. These networks represent different Web activities, which are collectively and popularly labelled as “The Web”.

2. The heterogeneous networks are all undergoing a process of translation, involving the continuous alignment of the actors already enrolled, and requiring the support and enrolment of new actors for its growth. The networks are dynamic and constantly changing shape as a result of the transient properties of the associations between actors. The stability (sustainability) of a Web activity (and by implication, the Web) is dependent on the actors’ commitment towards the network goals.

3. A Web activity is the product of a number of translating phases, which involves the (re)alignment of different heterogeneous networks through the association of shared actors and interests. Within each of the phases, there exist numerous heterogeneous networks, which individually are translating, and must persist in a stabilised state in order for the subsequent phases to function. The translation

of one network or phase causes changes to other, which in turn re-configures and re-problematises the Web activity.

Individually, the three principles of HTP provide their own analytical insight into the formation, development and structure of a Web activity, and by working with them in a single model, we are able to describe and explain the growth and structure of Web activity, and more generally: 'The Web'.

The conceptualisation that the HTP model provides a way to examine different Web activities, independent of context or domain; whether studying 'Social Networking', 'Human Computation', or more generally, the Web as a network of 'Web activities'. HTP offers not only a way to describe how the activity formed and grew, but also reveal the processes, layers, and structures that are hidden within.

Expanding on a well-established theoretical position which is grounded in the understanding of socio-technical phenomena and the co-constructive relationship between humans and technologies (Sismondo 2010*a*), HTP reconceptualises the current understanding of the Web's growth, both at the theoretical and practical level.

In isolation, HTP offers a theoretical understanding of how the Web has developed, and by pairing it up with a suitable methodology which combines the use of quantitative and qualitative data into a single framework, it becomes a powerful tool in describing, exploring, and understanding how Web activities, and thus the Web, functions. As the analysis has shown, combining quantitative and qualitative sources, it was possible to achieve an understanding of a Web activity's translation, which exposed the interlinking structures made by digital traces, and the ephemeral associations between humans and technologies.

9.5 A Contribution for Web Science

This chapter has moved from examining the UK OGD Web activity towards an understanding of the Web. By distilling and thinking critically about the findings of the analysis of the growth and structure of the Open Government Data Web activity performed in the previous chapters, it has been possible to construct a theoretical model which conceptualises the Web and the underlying social and technical processes that are embedded in it.

The HTP model pushes the current theoretical and practical understanding of a socio-technical Web, and also pushes the underlying theory of ANT that it is grounded in. HTP offers a way to understand the Web beyond the current approaches and their limitations; it pushes beyond the deterministic perspectives that conceive the Web as merely a social or technical entity, and provides a theoretical position that represents what it means for the Web to be socio-technical. HTP offers a way to understand the processes involved in Web growth at different levels of granularity; as a single Web activity, or the 'entire' Web, and by doing so reintegrates the micro and the macro. It also provides a handle on understanding the emergence of the fluid structure beyond the static Web graphs; providing a way to conceptualise the layering and stability of Web activities.

HTP also raises questions about Actor-Network Theory's approach to understanding the translation of complex agendas and multiple actor-networks, and through empirical evidence argues that translation of heterogeneous networks occurs in phases, which form layers of different temporarily stabilised networks. Each layer provides the social and technical building blocks for future networks to form and translate. The Web is an example of this, where different layers represent the social activities and technological advancements to support different kinds of Web activities, which collectively become known as the Web.

With that said, the next chapter will revisit the research aims of this thesis, and discuss the implications of the HTP model, examining its limitations, its application with a mixed methods framework, and what it provides in terms of readjusting the current understanding of the Web, and its ability to help describe the future developments of the Web.

Chapter 10

(Re) Conceptualising the Web

10.1 Re-addressing the Research Aims and Questions

This thesis has asked and answered a number of questions which challenge the current understanding of what is the Web and how it grows. Central to this challenge was a conceptualisation of the Web as socio-technical, drawing together both computational and sociological approaches to understand the Web. Paradoxically, whilst previous research has aimed to understand the social and technical elements and activities of the Web, studies have either focused solely at the micro or macro level, or limited themselves to the quantitative or qualitative paradigm. As a consequence of this, the methods, techniques, and theories that these approaches used did not provide an understanding of the Web in terms of the context of socio-technical.

This thesis has shown that by working with an interdisciplinary mix of computational methods and sociological theory it is possible to understand the growth of a Web activity and the socio-technical processes of human and technological interaction. By distilling these findings into a number of characteristics which represent the co-constructive processes responsible for Web activity growth, a new conceptualisation and understanding of the Web has been developed. The HTP model, as described in the previous chapter, re-conceptualises the Web, helping unpack the socio-technical processes that contribute to the growth of the Web in terms of a co-constructive process between human and technological networks. HTP has application at different levels of granularity, studying individual Web activities, or the Web as a whole; helping reintegrate the concept of the Web at the micro and macro level. Ultimately, the development of the HTP model provides a novel perspective for understanding the Web as socio-technical, and its application in combination with a mixed methods approach as used within this thesis, provides a powerful analytical framework to examine the implications for future Web growth.

Based upon the findings and development of this thesis, the chapter will now discuss the implications of this research on the use of methods, theory, and its application to future Web Science research.

10.2 Application of the HTP Model

The development of the HTP model provides a way to conceptualise and understand the growth of Web activities, and by inference, the Web itself. Grounded in the empirical analysis and findings of the Open Government Data Web activity, the model helps describe how online-offline human and technological activities lead to the emergence, growth and development of different forms of Web activities. The analysis of the UK OGD Web activity has also helped HTP develop a novel way to understand the development process of a Web activity, as a multi-layered and multi-phased translation process, which is applicable for understanding the growth of a single Web activity, or the Web in general.

Although HTP was developed based upon the analytical findings of a single Web activity, albeit one of the most dominant and influential activities in terms of the Web's development, in order to reflect on how it can be generalised as a model for understanding other forms of Web activities, as well as its application for understanding the current and future growth of the Web.

10.2.1 A Theory of Web Activity

HTP is a scalable model which can be used to conceptualise at varying levels of granularity and scales of Web activity. Although this thesis has not explored Web activity at different levels of granularity, based upon the underlying theory of heterogeneous networks and the punctualisation of actors, HTP can be applied independent of the size or scope of the Web activity being observed. For example, whilst this thesis has studied translation of the UK Open Government Data Web activity, which represents a large, complex network of networks with multiple agendas; HTP can also be used to understand the translation of smaller, more simplistic Web activity, such as the development of a new Web service, or, a new 'social machine' (Berners-Lee 1999).

HTP appreciates the distributed growth and emergent structure of the Web and provides the middle layer to reintegrate the micro and the macro; it offers a way to study the Web as individual networks of Web activities, or examine them at a more abstract level, as a collection of Web activities, or more formally, the Web.

By itself, HTP is a theoretical model to conceptualise the growth and structure of translating heterogeneous actor-networks, and paired with a mixed methods approach as developed within this thesis, it is possible to expose and provide context of the socio-technical processes and co-constructive relationship between the human and technologies that underpin the growth of Web activities. By drawing upon both large-scale quantitative data sources, along with the variety of qualitative data sources available, and using the appropriate methods and techniques, it is possible to do away with the boundaries of the micro and macro, which are often associated with qualitative or quantitative findings.

10.2.2 A Theory of the Web

The transition from a theory of Web Activity to a theory of the Web is a step in terms of refocusing the observational lens from the micro to the macro. It is also an ontological step; the weaving of the Web is the Web. As an abstract model, HTP offers a way to describe and understand the interactions of actor-networks, whether these represent the activities of a network of individuals, or less granular, a Web activity; this all becomes a matter of punctualisation and focus of the observation.

The Web, conceptualised via the observational lens of HTP, is a collection of Web activities which are constantly interacting, competing and translating. The development of the Web is the result of these interactions, and the emergence of new Web activities can be conceptualised in terms of the phases of Web growth. For example, observed through the lens of HTP, what has become collectively known as Web 2.0 (Murugesan

2007), and the development towards a Web 3.0 (Barassi & Trere 2012) is no more than the emergence of new Web activities that are building upon the success and stability of Web activities in previous phases. The implications of this are beyond just providing a new narrative for understanding and describing the growth of the Web; it argues that without the stability of the previous networks of activities, the growth of the 'next Web' (which through the lens of HTP is simply the new collection of Web activities which become collectively known as Web 4.0), will not be possible.

As the analysis of UK OGD demonstrated, the stability of existing Web activities becomes the platform for future activities to take place. HTP provides a way to conceptualise this independent of scale; thus by exploring what Web activities exist and have remained stable during the Web's various phases of development, it becomes possible to understand what specific Web activities are required for continued stability. By understanding the co-constructive processes that have enabled these activities to remain stable, it may be possible to help support the future stability and growth of the Web by ensuring that these resources (humans, technologies, infrastructures, policies, laws, etc.) continue to remain available.

10.2.3 A Theory to Predict Web Growth

The growth of the Web is a socio-technical process which involves the activities, interactions and importantly, the commitment of human and technologies. Consequently, the prediction or extrapolation of the current and future trajectory of the Web becomes problematic for a number of reasons: from uncontrolled and unexpected environments and external conditions, to the multi-variant behaviour and decisions of humans (Dhar 2012). However, this is not to say that HTP cannot provide some insight and understanding to the potential pathways of Web growth. Whilst HTP does not claim to be a predictive model for understanding the growth of the Web, through its application it provides a way to distil the socio-technical processes and characteristics responsible for the growth of Web activities.

Reflecting upon the extensibility of HTP, whilst the aim of this thesis was not to provide a solution for modeling or predictive analysis, through the discovery of the socio-technical processes that led to HTP, what has been provided is the groundwork for future studies to take and develop further this insight in terms of modeling and simulating Web growth for anticipatory analysis and understanding of other kinds of Web activities. There may be potential utility in using the principles described in HTP to improve current approaches of Web analytics.

Although prediction in the traditional sense is not possible, by using HTP to examine, compare, and contrast the characteristics of a variety of Web activities, there is the potential to understand how specific characteristics (contextual and structural) are responsible for specific translation pathways. By comparing a range of different kinds of Web activities which have varying agendas and are large and small scale, the identification of commonalities and traits in the socio-technical processes may help form basic guidelines to supporting a stabilised and growing Web.

Take the example of the Open Government Data Web activity, by distilling the socio-technical processes that supported its emergence and growth, a number of characteristics were revealed such as the de-centralised structure of networks, the role of competition, and the adoption and manipulation of technologies to support communications. By taking these characteristics and comparing them to a collection of other Web activities, it may be possible to infer common traits which enabled them to translate. By taking these characteristics and processes it may be possible to help improve the adoption and success of other international OGD Web activities, overcoming barriers and challenges which may have been identified and tackled within the UK OGD Web activity. In this sense, HTP not only offers utility to help describe and analyse the Web, but it also becomes a tool to pro-actively and re-actively support the growth of Web activities.

Furthermore, given that HTP conceptualises the Web as a number of layers of stabilised activities, by examining the layers of new Web activities, it may be possible to 'map' certain types of characteristics as the enablers for specific kinds of Web activities. For instance, whilst the stability of UK OGD enabled Data Journalism to emerge, the stability of a social networking site such as LinkedIn might become the enabler for new types of professional networking Web activities to emerge.

Understanding the Web is not just about looking at success, it is important to understand failures in order to increase the chances of success in the future. By using HTP to identify the characteristics which led to the failure of Web activities, it is possible to construct guidelines that represent the potential barriers and challenges that may affect the future development of the Web. At a larger scale, understanding the subsequent outcomes from the failure of Web activities will also be beneficial to understand the consequences of not maintaining the stability of critical Web activities.

Ultimately, moving from modelling to predicting the translation of the interplay of human and technological activity - which essentially is what the Web represents - has limitations. However, learning from what has already happened, and observing the Web in real-time through appropriate methods may provide some benefit to understanding the Web's growth.

10.3 Implications for Actor-Network Theory

From theory through to application, Actor-Network Theory has been fundamental to this thesis. It has provided the basis for understanding the co-constructive relationship between society and technology, and offered a theoretical position that takes into consideration the agency of both humans and technologies. The concepts that ANT offer were central to the research aims and questions of this thesis; underpinning the theoretical perspective of the HTP model, and used as an analytical lens within the mixed methods framework to understand the processes that led to the translation of a Web activity

Whilst ANT offers concepts and theory to understand socio-technical processes that enable a heterogeneous network to form and translate, the empirical analysis has raised questions regarding ANT's position and adaptability for understanding complex socio-phenomena such as the Web (and Web activity), and also its application with a methodology which draws upon the quantitative and qualitative paradigm.

By working critically with Actor-Network Theory, this thesis has taken a step forward in terms of its integration with a mixed methods approach to understanding the co-constructive process between humans and technologies. By challenging ANT position on how networks are constructed and represented within the translation process, and by demonstrating that in practice, the boundaries limit analytical capabilities; this thesis has developed ANT in order to understand the translation of Web activity and the Web as part of a multi-layered and multi-network process. In reflection, this thesis has provided a contribution to ANT, both in application and theory.

10.3.1 Rethinking Methods

A critical approach within this thesis was the combination of ANT with a mixed methods methodology in order to capture the human and technological interactions that are associated with the micro and macro scale, both online and offline. In order to create a methodology, and by extension, an analysis framework which could capture Web activity, both quantitative and qualitative data sources and analytical techniques were drawn upon, providing a way to add context to the quantifiable patterns, and associate patterns with the hidden and tacit knowledge found.

A challenge within the development of the methodology was pairing up the theory and concepts provided by ANT with a mixed methods approach to understanding socio-technical phenomena. This presented both a theoretical and practical challenge. In previous ANT studies insight was generated from the use of rich qualitative sources and focused far less on the use of the large scale or real-time sources of data. However, with the ever growing pool of big data sources (Manovich 2011), which offer insights into the social at a scale which was previously not possible (Savage & Burrows 2007) (Tinati et al. 2013), this thesis pushed the analysis to engage with methods across both the qualitative and quantitative paradigms, and with data sources which were highly dynamic, linkable, and real-time. The use of these types of 'big data' sources reflected and catered for the Web's dynamic and fluid properties and the transient boundaries between the online and offline.

By working critically with analytical methods and techniques, this thesis pushed ANT to engage with both the quantitative and qualitative paradigm, following the digital traces of actors at the micro (Latour et al. 2011), whilst also working at the macro level in order to consider the structures that were emerging from the data. This thesis has developed ANT to enable the observational lens to be shifted between different levels of granularity, re-integrating the micro and the macro, to provide an understanding of the socio-technical processes and the co-constructive relationship and agency of human and non-human actors, independent of scale.

This thesis has also shown that it is possible to successfully engage with mixed methods to exploit both the richness of quantitative and qualitative sources; providing an analytical framework with both descriptive and explanatory capabilities. By using these sources as complementary, an analysis rich in context and structure can be achieved, essential to understand human and technological interaction. Furthermore, engaging with ANT across a mix of data sources helps expose the interactions between actors, and how the shared agency between humans and technologies is responsible for a network's translation.

10.3.2 Rethinking Theory

An essential part of this thesis was engaging with ANT at a theoretical level in order to expose the socio-technical processes and the translation that underpinned the growth of a Web activity. Harnessing the concepts of ANT helped expose specific characteristics between the interplay of humans and technologies, and also helped reveal the interaction between the online and offline activities.

The process of translation was a critical tool for understanding the growth of a Web activity, and is an essential property within the HTP model developed. Whilst it provides a framing to exploring the growth of socio-technical activities and the mobilisation of an agenda, it also has a more fundamental role in terms of how networks are transient, dynamic and only temporarily stabilised. By using the empirical study of the UK OGD Web activity and thinking critically about the role of actor-networks and how translation occurs, the growth of a Web activity is much more complex than the translation of a single heterogeneous network; instead it consists of multiple actor-networks, operating with their own agendas, and through various different social and technical processes and interactions, networks become temporarily stabilised, and new networks emerged and begin to translate.

Based on these findings the formation of the HTP model has pushed ANT towards understanding the development of a Web activity as a series of actor-networks which translate in phases, forming layers of activities, which remain stable in order to achieve a new phase. This conceptualisation of how an agenda is shaped, modified, and mobilised, not only applied to the growth of a Web activity, but more generally, the growth of the Web. The concept of phases also reflects the fractal nature of the Web graph, in contrast to traditional computational approaches which expose 'communities, or 'hubs and authorities', HTP exposes translating actor-networks with specific agendas. The evolution of the graph, and the growth, increasing density and strength of ties is the formation of new phases; aligning existing networks and the emergence of new networks.

This thesis has pushed ANT beyond a descriptive framework, pairing it up with a mixed methods framework to offer explanatory capabilities to help understand and inform the development, phases, and translation of the current and future layers of the Web. By pushing ANT in terms of the theory, methods and data sources, HTP offers a conceptualisation of the Web which reflects its dynamic and evolving structure, and co-constructive growth.

10.4 Implications for Web Science Research

The re-conceptualisation of the socio-technical Web offered by the HTP model has implications for future Web Science research in terms of the theoretical perspective of understanding the Web and the methods used to explore its growth and structure. This thesis has used an empirical-driven study to construct a novel conceptualisation for how Web activities emerge, evolve and mobilise, and how these Web activities are the result of the associations and socio-technical interactions between humans and technologies.

At the practical level, contributing a new model and analytical framework provides Web Science with a set of tools to continue research into the growth and structure of the Web; however, at a more theoretical level, by engaging across disciplinary boundaries, this thesis has tackled some of the fundamental epistemological and ontological challenges of interdisciplinary research.

10.4.1 Distilling a Socio-Technical Understanding

Driving the research aims at the start of this thesis was the application and use of the term socio-technical to describe the Web as a technological phenomenon. Whilst this described the formation and evolution of the Web graph, the identification of communities and clusters, and the relationship of the Web and society; it did not adequately describe the co-constructed relationship of humans and technologies. However this thesis has developed a perspective to integrate human and technological action; by engaging with social theory embedded in understanding socio-technical phenomena, a theoretical perspective was developed to expose the underlying social and technical processes which represent the 'socio-technical' phenomena often described. Ultimately, a socio-technical Web is one which grows through a co-constructive relationship, shaped by the configuration of humans and technology.

Drawing upon this theoretical position, the empirical analysis of Web activities demonstrated the underlying interactions between humans and technologies and the social and technical processes that emerge. Through this observational lens, the Web graph no longer represents only a network of Web documents connected by hyperlinks, but as a network formed through the interactions and associations between technologies and humans. The exchange of data and the formation of Web pages are the network artefacts and outcomes of socio-technical processes, and the network structure (or the edges) represents a snapshot of the temporarily stabilised state of Web activity. Whilst this thesis is not suggesting that future Web Science research should disassociate itself with understanding the Web as a technical structure of hyperlinks and documents; what this thesis is providing is an additional lens to understand the Web beyond just its technical architecture.

This understanding of socio-technical is a theoretical distinction as much as a practical one, and has important implications for Web Science research. By exposing the agency and co-construction of humans and technologies within the Web's development, Web Science is now in the position to think critically about the relationship between the Web and society; and through the development of models, methods and analytical techniques, Web Science is now armed with the necessary practical tools in order to achieve this.

10.4.2 Interdisciplinary Thinking

Web Science is interdisciplinary by nature, and requires a balance of theory, methods and analytical techniques in order to answer the kinds of questions Web Science research asks. However, as part of this process of interdisciplinary research, challenges are not only faced at the theoretical and practical level, but at the communication layer as well; especially as different disciplinary practices play an essential role in being able to convey research questions, ideas, and findings. Akin to the differences between the quantitative and qualitative paradigms, the underlying epistemological and ontological beliefs that disciplines side with, makes achieving true interdisciplinary research a challenging task.

Nuances in language, formatting, and even presentation play an important role in how disciplines can communicate with each, even the simplest of words can create confusion and miscommunication. For example, take the word 'Ontology', not only does it convey entirely different meaning between the social sciences and computational sciences, but even the way it is written becomes an interdisciplinary issue ('Ontology' or 'ontology'). Although a crude example, this is one of the many issues faced when researching in the crossroads between disciplines. Facing these challenges, this thesis has demonstrated that by drawing upon methods, theories, and techniques from the computational and social sciences, it is possible to conduct interdisciplinary research and still adhere to and appreciate the underlying perspectives taken by such distinctly different disciplines. Moreover, it is these differences that make interdisciplinary research so powerful; rich in terms of analytical capabilities, and grounded in terms of the theoretical underpinning. The different perspectives offered complement each other, highlighting areas and findings that a single discipline may not consider. The disciplines that such research draws on can also learn from the techniques used, helping breakdown the academic borders associated with methods and theory (Savage 2010).

To conduct Web Science research, interdisciplinarity is essential, and it is the combination of both theory and practice from multiple disciplines that will enable research to continue to develop insights about the Web.

10.5 The Future of the Web

The final section of this chapter revisits the original questions posed that helped problematise and establish Web Science as a discipline (Berners-Lee et al. 2006), and essentially, the questions that provided the inspiration and footing of this thesis.

The Web, as described throughout this thesis is a growing complex network of humans and technologies, data, services, protocols, standards, and policies. It is a changing network of socio-technical processes which lead to the formation of Web activities which collectively represent the Web in its current state. As a consequence of its temporary stability, possibilities to make sense of its future growth are at best, limited to a real-time snapshot of associations and state of activity. Despite this, attempting to understand the future of the Web in terms of its growth, structure, capabilities and barriers, should not be discouraged, and the more research conducted, the more likely that society, researchers, and individuals will be able to make informed decisions about the development of new services, standards and technologies, and their potential impact within the wider context of the Web.

10.5.1 Understanding Web Growth

The driving research aim of this thesis was to explore and develop a conceptualisation of the growth and structure of the Web, and as part of this journey, Web activities were explored, theories were examined, and models were created. Looking through the observational lens of HTP offers a novel contribution to the conceptualisation of how Web activities - and by inference, the Web - grows, and how the structure of the Web emerges.

Arguably, to be able to understand how something may develop in the future, it is essential to understand how it has developed in the past. The lessons learnt from previous successes and failures are important properties of how the Web, over the past two decades has grown. Being able to understand the structures that coincide and form due to the growth of the Web is also critical for being able to identify similarities in the future. Similarly, at a theoretical level, understanding the growth and structure of the Web requires the ability to unpack the socio-technical interactions that underpin the structure (and Web activities) that emerge.

To be able to explore the co-constructive relationship between the Web and society, politics, economies and governance, it is essential to be able to understand the growth and structure of the Web, and the development of HTP and the analytical framework provides the capabilities to do so.

10.5.2 A Final Thought: 'Web Engineering'

"The Web needs to be studied and understood, and it needs to be engineered" (Berners-Lee et al. 2006)

This thesis has demonstrated that the Web is a socio-technical system co-constructed by the activities of both humans and technologies, and its success (and failure) is reliant on their interactions. This thesis has shown that embedded within these interactions are complex social and technical processes that influence the translation of the Web.

However, whilst studying and understanding the Web is possible, engineering it presents a number of challenges. As the analysis within this thesis has shown, the development and acceptance of a technology is part of a complex socio-technical process at the micro and macro level. It involves the interplay between humans, policies, governance, politics, and economies, and what is typically engineered in the micro (controlled lab environment) becomes something wildly different at the macro (uncontrolled network of humans and other technologies).

Further engineering of the Web in some sense is inevitable; developers will continue to develop, and users will continue to reinvent, break, adapt and 'misuse' technologies. Web activities will continue to flourish and fail, and Web Science research will learn from these successes and failures through models and analytical frameworks from computer science and social science, and interdisciplinary approaches such as the one used in this thesis. But it should not just stop there; the lessons learnt need to be re-applied in design methodologies and frameworks which govern and guide Web engineering practices and principles. Traditional approaches of design which view the Web as a technology and fail to grasp its co-constructed nature are no longer adequate or sufficient. The Web is neither social nor technical; it is a co-constructed heterogeneous and dynamic network, which is constantly in translation.

Appendix A

Interview Questions

A.1 Understanding the Open Data Community

A.1.1 Interview Topic Guide

This study includes interviews with a number of different people associated with the open data community. The interview is like a conversation as the researcher wants to find out your views and ideas. The questions below are a guide to the kinds of things the researcher is interested in but you may introduce other issues and the researcher may prompt you for further discussion.

A.1.2 Questions

The order of questions and structure of the interview is not fixed and will vary for each interview.

- Can you tell me a bit about your involvement with UK open government data? Has it changed over time (describe how)?
- I've been thinking about the people involved in open government data as a community would you describe it as a community? (or do you have another term you'd use to describe it)
- What do you think were the beginnings of UK open government data

- Were there any memorable events around UK open government data that you recall? Is there anything that you see as a milestone or a key turning point?
- Who do see as the key influencers/stakeholders/ around UK Open Government Data ?
- What do you think helped the formation /development of UK Open Government Data community? And what hindered/limited it? What sorts of things did people do to overcome these barriers?
- What do you see as important technologies within the Open Government Data Community?
- What do you think about the relationship between the development of technologies within the Open Government Data Community and the impact they have?
- What do you think about the Semantic Web and Linked Data technologies in relation to the Open Government Data Community?
- Thinking about the future where do you see UK Open Government Data heading? What are the next steps?
- Is there anyone else that you would recommend I talk to about these ideas?

Appendix B

Coding Schema

B.1 Initial Coding Schema (Seeded List)

Code	**Description**
Actor [Individual, Networks of Actors]	Content that identifies an actor, can be human or non-human, but also can represent a network of actors, a community, or organisation
Actor Attributes [Interest, Power, Role, OPP, Focal Actor, Influence]	Content relating to the relational effect that an actor has within a network; this can include their interest in the activity and their power/authority, their role within a network, which also includes specific ANT identifiable roles within a network including the Focal Actor, and Obligatory Pass Point [OPP]. Also another actors attribute is defined by the influence on other actors or artefacts that an actor has within the network
Actor Interaction [Strengthening Ties, Negotiations, Partnership, Disruption, Curating Networks, Commitment]	Content relating to the interaction of identifiable actors, this could be either for positive or negative effect on the network. Content relating to the strengthening of ties between actors, the negotiations, the formation of partnerships [individuals and actor-networks], disruption of actors [and actor-networks], and interactions to curate networks [join actor-networks together]

Network Artefacts [Policies, Government, Citizen, Law, Legal Issues, Artefacts, Interessement Device]	Content related to different kind of social artefacts, this can include policies, which can be Government or citizen driven. Also content can relate to new and existing laws or legal frameworks being created, this is closely related to legal issues, which is content related to the implications of the Laws. There are also entities within the network which act as Interessement devices, connecting actors and strengthening ties.
Network Interactions [Events, Meetings, Serendipitous, Encounters]	Content relating to different kinds of social interactions rather than just actor interactions. This can include content relating to an event notification of event, review of event and the event could be public or private. Meetings are for specific events, within a specific actor-network or identified community. There are also interactions in the network that occur serendipitously, and through the interactions of actors and artefacts.
Network Development [Structure, Milestones, Drivers, Barriers, Outcomes, Competing Networks]	Content related to different developments in the network of activity. This includes its structure, milestones [related to different events and meetings], specific drivers within the process of translation, including the actions and outcomes of actor interactions, barriers inhibiting the process of translation of the actor-network, and the outcomes of the network. Content may also relate to the competition between networks, either for positive or negative outcomes.
Network Goal [Individual, Network]	Content related to either an individual's goal, or the goal of an entire network of activity. Goals relate to future achievements or milestones set
Network Outcomes [Physical Artefacts, Network Artefacts]	Content relating to the outcomes of the actor interaction within the network, this can be in the form of physical artefacts such as documents, data, or network artefacts, such as knowledge, agreements, formations of communities
Network Activity [Online, Offline]	Content related to specific kinds of activity, either, by not exclusive to each other, online or offline activity. This is related to content regarding network interactions [events, meetings], or actor interactions

Incentives [Financial Reward, Natural, Coercive/Punishment]	Content relating to direct and indirect financial reward, and also threats to a network to coax it into performing certain activity. There are also natural incentives due to curiosity and interest into the network activity
Technology [Agency, Development, Change, Adaption, Failure]	Content relating to technology [or non-human actors] within an actor network. This could involve their agency, and how they exist within the network, their development and changes made to the, how they have adapted to changes within the network, and also their failings and problems
Translation [Problematisation, Interessement, Enrolment, Mobilisation]	A data source which contains evidence of actors [not necessarily the focal actor] assessing the problems or the goals of the network, what actors are required to be included, and what barriers exist. A data source which contains evidence of actors working together on the problems that have been identified. A data source which contains evidence of actors working together based on their interessement. The evidence will show progress or the beginnings of change. A data source which contains evidence of new outcomes as a result of actor interaction and successful enrolment and cooperation between actors. A data source which contains evidence of outcomes which are important for the networks stability, including new policies and laws being created and acted upon, or the introduction or formation of new actors based on the outcomes of the actor-network

Table B.1: Initial Coding Schema

B.2 Final Coding Schema

Code	**Sources**	**References**
Open Government Partnership	1	1
TFL	1	1

Translation	2	2
Passage Point	2	2
Stability	8	9
Iteration	14	20
Instability	13	24
Mobilisation	14	24
Enrolment	18	39
Problematisation	23	57
Interessement	21	62
Negotiations	2	2
Blurred Networks	3	3
Activism	3	3
Open Data Institute	2	3
Barrier	3	4
Location Barrier	2	4
Legal Barriers	7	11
Financial Barrier	10	17
Technical Barrier	16	23
Organisation Barrier	8	28
Social Barrier	20	45
Socio-Technical Barrier	20	65
Event	4	5
Conference	5	5
Hackathon	6	9
Meeting	8	9
World Bank	1	5
Council	4	5
Council Development	5	11
Council Structure	4	12
Common Goal	5	6
Legal Issues	5	6
London Datastore	5	6

Social Media	4	6
Network of Actors	5	8
Freedom of Information	5	8
Online Activity	7	9
Milestone	9	10
Serendipitous Events	8	11
Financial Reward	8	12
Network Failure	9	12
Incentive	9	13
Financial Incentive	11	17
Cost	8	13
Process Change	8	13
Example	10	14
Example - Social Media	1	1
Example - Social Barrier	2	2
Example - Application Dev	2	2
Example - Data Request	4	4
Example - Data problems	4	6
Example - Council	4	6
Technology	9	17
Socrata	1	1
PDF	1	1
Web Chat	1	2
Technology Failure	4	4
Technology Change	6	6
WordPress	2	6
CKAN	4	8
Application	7	12
Technology Development	11	14
Data.gov.uk	9	14
Twitter	13	29
Linked Data	16	41

National-Level Network	7	17
National-Level Engagement	3	5
Interessement Device	10	18
Network Benefits	10	18
Previous Agenda	13	20
Network Artefact	11	20
Organisational Structure	14	22
Data	14	22
Data Journalism	1	1
Corporate Data	2	3
Personal Data	3	4
Stars of Data	4	4
Data Users	10	17
Data Problems	13	18
Data Publishing	8	21
Network Agenda	14	23
Fast Timeframe	13	23
Threat	14	24
Time	16	27
Policy	14	27
Strengthening Ties	18	31
Network Goal	15	31
Previous Networks	14	37
Politics	16	38
Individual Agenda	16	41
Competing Networks	20	47
Healthy Competition	4	6
Network Outcomes	20	50
Curating Networks	19	56
Future Adoption	22	57
Drivers	21	66
Drivers-Actors	9	13

Network of Networks	18	68
Network Structure	23	97
Network Development	23	99
Actor	23	157
Local Council	1	1
Armchair Auditors	2	2
Citizens	2	3
Academic	3	7
Government	7	12
Actor Commitment	14	25
Actor Influence	17	31
Developers	16	33
Actor Role	18	51
Actor Interaction	19	57

Table B.2: Final Coding Schema

Appendix C

Timeline of UK Open Government Data Events

C.1 Timeline of Events

C.1.1 Source Key

Source	Name
OKFN	Bar Camp UKGovWeb 2009 Hackathon
Academic	Academic Publications
G	The Guardian Newspaper Online
GDB	The Guardian Datablog
data.gov.uk	Data.gov.uk Online Blog
PTB	Public Transparency Board Meeting Notes
CO	Cabinet Office Documents and Blog
PM Letter	UK Prime Minister Letters
HM	UK Government Official Documents

Table C.1: Source Legend

C.2 Activities and Events

Source	**Description**	**Date**
OKFN	Bar Camp UKGovWeb 2009 Hackathon	6th February 2009
GDB	The Guardian begins their Data Blog section	10th March 2009
OKFN	CKAN updated to 0.8	12th May 2009
G	Gordon Brown on constitutional reform prime minister's statement in full	10th June 2009
G	Web inventor to help Downing Street open up government data	10th June 2009
OKFN	OpenTech 2009 Conference discussing Open Government Data with government	3rd July 2009
OKFN	Openly Local Released	4th August 2009
OKFN	CKAN updated to 0.9	13th August 2009
OKFN	Bus Stop, Post Box and Open Street Map Data released	28th August 2009
data.gov.uk	Sneak peak of data.gov.uk released to developers	30th September 2009
GDB	Article urging developers to get involved with government open data	2nd October 2009
OKFN	data.gov.uk Beta released	8th October 2009
data.gov.uk	150 datasets added to data.gov.uk	8th October 2009
Academic	Government Linked Data: A Tipping Point for the Semantic Web	8th November 2009
OKFN	Open Data and Semantic Web Workshop	13th November 2009
OKFN	Ordnance Survey releases statement about open map data	19th November 2009
OKFN	Where Does My Money Go Released	11th December 2009

OKFN	London Data store added	11th January 2010
OKFN	Draft of Open Data Commons Attributions Licence produced	12th January 2010
OKFN	data.gov.uk publically released	21st January 2010
GDB	Tim Berners-Lee and Nigel Shadbolt first post on data blog talking about importance of government data	21st January 2010
GDB	Video presentation by Tim Berners-Lee and Nigel Shadbolt discussing making data publi	21st January 2010
GDB	Video presentation by Stephan Timms (MP) talking about the launch of data.gov.uk	21st January 2010
Academic	Here Comes Everything: Preparing for the Lined Web of Data	25th January 2010
data.gov.uk	User accounts added to data.gov.uk to allow for user interaction	29th January 2010
OKFN	How to make open data catalogues meeting	11th February 2010
Academic	data.gov.uk the linked data revolution	17th March 2010
GDB	Article on Ordnance Survey and their open map data	2th April 2010
OKFN	World Bank makes their data public	20th April 2010
OKFN	Warwickshire creates their own data site	30th April 2010
OKFN	CKAN updated to v1.0	18th May 2010
GDB	London Authority spending application made available	18th May 2010
data.gov.uk	Seminar on COINS datasets	18th May 2010
data.gov.uk	Finance Data (COINS) is made available as datasets on data.gov.uk	19th May 2010
PM Letter	Letter from Prime Minsters regarding his plans to open up government data	31st May 2010

GDB	Article reporting the release of the COINS datasets	31st May 2010
PTB	1st Public Transparency Board Meeting Nigel Shadbolt's letter of OGD commitment	24th June 2010
data.gov.uk	Development of the UK Government Licensing Framework underway	29th June 2010
OKFN	UK Transparency board set up	28th June 2010
data.gov.uk	Open Spending data expanded	2nd July 2010
GDB	Article on how much does government websites cost to run	5th July 2010
Academic	Open Government Data: A Case Study in Web Science	8th July 2010
data.gov.uk	data.gov.uk team forms partnership with Open Tech	29th July 2010
data.gov.uk	The source code for data.gov.uk is made open	20th August 2010
OKFN	CKAN releases Drupal module for developers	21st August 2010
OKFN	Public Report on access to information and open government data	10th September 2010
data.gov.uk	data.gov.uk takes part in Open Tech Event	11th September 2010
PTB	2nd Public Transparency Board Meeting PM Letter of Opening Data discussed	15th September 2010
CO	Transparency frequently asked questions	5th October 2010
data.gov.uk	Nigel Shadbolt speaks at Activate 2010	15th October 2010
CO	Send us your comments says new Transparency Board	25th October 2010
OKFN	WhereAreTheCuts.org made publically available	26th October 2010

GDB	Article on the release of the UK Transport data	27th October 2010
CO	Ministerial gifts, hospitality, travel and meetings with external organisations	29th October 2010
GDB	Fixed Penalty notices application made available	7th November 2010
Academic	Why Open Government Data? Lessons from data.gov.uk	8th November 2010
GDB	Article Listing all the business reports by government departments	8th November 2010
data.gov.uk	data.gov.uk forms partnership with Where-CampUK	9th November 2010
OKFN	More Data added to Where Does My Money Go	19th November 2010
OKFN	Open Government Data Camp Event	19th November 2010
PTB	3rd Public Transparency Board Meeting new OGD policies	19th November 2010
data.gov.uk	API made available for legislation.gov.uk portal	25th November 2010
data.gov.uk	Nigel Shadbolt discusses with the National Archives data.gov.uk and transparency	4th January 2011
PTB	4th Public Transparency Board Meeting Another Letter from Nigel, Kieran, O'Hara reviews crime data	11th January 2011
CO	Public Data Corporation to free up public data and drive innovation	12th January 2011
data.gov.uk	Open Data Master Classes are happening around the country	13th January 2011
GDB	Article by Nigel Shadbolt, A year of data.gov.uk	21st January 2011

GDB	Article about how successful has government data in the UK been	21st January 2011
GDB	Article by Paul Clarke on how he would fix data.gov.uk	22nd January 2011
GDB	Article listing all local council spending over £500	1st February 2011
PTB	5th Public Transparency Board Meeting commitments of department of transport	8th February 2011
data.gov.uk	Linked Data Workshop	10th February 2011
PTB	6th Public Transparency Board Meetings Right to Data policy discussed	2nd March 2011
GDB	Article about funding cuts to data.gov and the open data movement	5th April 2011
data.gov.uk	Online Tax Calculator application developed	7th April 2011
data.gov.uk	School Scout application developed	18th April 2011
data.gov.uk	UK Fuel Prices application developed	5th May 2011
PTB	7th Public Transparency Board Meeting police.uk showcased	19th May 2011
PM Letter	Letter from Prime Minister regarding his commitment to open data	8th June 2011
PTB	8th Public Transparency Board Meeting second PM letter and health sector reviewed	8th June 2011
GDB	Article listing all Whitehall civil servants, their role and their employer	16th June 2011
data.gov.uk	Prime minister discusses his commitments to releasing more data on health, schools, courts and transport	7th July 2011
CO	Governments to publish new data on health, schools, courts and transport	7th July 2011

GDB	Article reporting the new government transparency initiative	7th July 2011
PTB	9th Public Transparency Board Meeting Gaining interest from new data providers	19th July 2011
HM Gov.	Release of Making Open Data Real, corresponding to the Open Data consultation	1st August 2011
GDB	Article regarding the benefits and dangers of open government data	4th August 2011
OKFN	County of Kent releases Open Data Portal	24th August 2011
OKFN	City of Manchester releases Open Data	25th August 2011
OKFN	FixMyTransport.com publically released	30th August 2011
PTB	10th Public Transparency Board Meeting reviewing DEFRA's datasets	6th September 2011
data.gov.uk	A Local Public Data panel was ran discussing what data to be released	30th September 2011
GDB	Article Looking at the Future of Open Data	11th October 2011
PTB	11th Public Transparency Board meeting Discussing the Open Government Partnership (OGP), licensing and new data	11th October 2011
OKFN	New Head of data.gov.uk announced Antonio Acuna	12th October 2011
CO	Top Companies commend Francis Maude for Open Data	12th October 2011
Academic	Open Government Data: Making it Real	21st October 2011
OKFN	Transparency board asks for public response to UK Data consultation	21st October 2011
OKFN	Open Data: Wishlist for the Next Year	23rd October 2011
CO	How should Government become even more open?	4th November 2011

OKFN	CKAN updated to v1.5	14th November 2011
OKFN	Data Journalism Handbook: Hackers and Developers Come together	14th November 2011
OKFN	European Commission to adopt an Open Data Strategy	24th November 2011
GDB	Article demoing a new application, FactLab	24th November 2011
GDB	Article discussing the Ministry of Justice opening up court data	25th November 2011
CO	Making Open Data Real: A Public Consultation	27th November 2011
OKFN	Seize the (Data) Day for Open Data Day	28th November 2011
GDB	Article Discussing the release of Whitehall statistics	28th November 2011
CO	Open Data measures in the Autumn Statement	29th November 2011
OKFN	International Open Data Hackathon	30th November 2011
OKFN	Article on the benefits of Open Data for Science	9th December 2011
GDB	Article discussing European Open Data	9th December 2011
PTB	Autumn statement reviewed, also discusses new finding for data.gov.uk. New focus will be the Open Data White Paper.	13th December 2011
GDB	Article on the publication of primary school data	15th December 2011
GDB	Article on the Book: the Power of Data	6th January 2012
OKFN	Article on what Civic Society actors are doing with datasets	12th January 2012

OKFN	Article on Linked Open Data and Low Carbon Development	17th January 2012
OKFN	Open Economics Hack Day Saturday January 28th 2012	18th January 2012
GDB	Announcement of the international data journalism award	19th January 2012
PTB	New departments data are discussed Department of Works and Pensions and also Crime and Justice Department. Departments to start to produce data, and also TB need to consider EU Directive	24th January 2012
CO	Making Open Data Real Consultation - A Summary of Responses	30th January 2012
CO	Francis Maude's speech to the World Bank Transparency and Open Data	30th January 2012
OKFN	Press release that CKAN will Power the new European Commission Data Portal	31st January 2012
GDB	Article discussing the Open Data Consultation	31st January 2012
GDB	Article discussing problems with Open Government Data	31st January 2012
GDB	Article showcasing Search and Recuse application	7th February 2012
GDB	Article showcasing roadwork applications	16th February 2012
OKFN	Press release of the Open Definition Licenses Service	16th February 2012
OKFN	Article requesting public to submit their votes for the Data Journalism Awards	20th February 2012
OKFN	Press release of the Open Data Handbook version 1.0	22nd February 2012

CO	Second release of GOV.UK beta	28th February 2012
GDB	Article on what is a Data Scientist	2nd March 2012
CO	Information Commissioner's Conference Francis Maude keynote speech	6th March 2012
GDB	Article describing the Role of Data Journalism in the Guardian	7th March 2012
GDB	Article on Big Data and the loss of theory	9th March 2012
GDB	Radio interview: Hans Rosling, Martin Rosenbaum and Simon Rogers on open data	13th March 2012
CO	Public data means the UK is open for business	14th March 2012
CO	Open Data Innovation Community - Francis Maude speech	14th March 2012
GDB	Announcements of release of Tax receipts since	15th March 2012
GDB	Announcements of publication of every open spending data site in the US ranked and listed	16th March 2012
GDB	Article discussing the top 10 datasets being used	21st March 2012
OKFN	Article discussing the School of Data Journalism	23rd March 2012
OKFN	ePSI Platform Conference 2012	23rd March 2012
GDB	Announcement of Guardian Open Day: working with fusion tables	24th March 2012
OKFN	Article examining Linked Open Data and Europeana	26th March 2012

PTB	Discussed the Rights to Data White Paper, Francis Maude says there should be an ambitious wish list. Also discussed the Local Government Association, how they can release data etc.	27th March 2012
OKFN	Press Release to introduce the DataStore, its new features etc.	27th March 2012
GDB	Article Discussing Open Street Map vs. Google Maps	28th March 2012
GDB	Article discussing new Open Data train timetable application	2nd April 2012
GDB	Announcement that Google using London Underground live Data (TFL)	3rd April 2012
GDB	Article highlighting the benefits of Open Data Journalism	13th April 2012
GDB	Article discussing the Open Government Partnership	16th April 2012
OKFN	Article informing about the launch of the Open Data Census 2012!	17th April 2012
GDB	Article providing overview of UK Open Government Data official audit	18th April 2012
OKFN	Article providing an overview of the UK Open Standards Consultation policy	18th April 2012
GDB	Article on outcomes of Open Cities Conference	20th April 2012
GDB	Announcement of the Data Journalism Handbook	20th April 2012
OKFN	Article providing an introduction the Data Journalism Handbook	21st April 2012

PTB	Department for international development looking at ways to create transparency. Releasing their data etc. ODI discussion Nigel update board on current progress	24th April 2012
OKFN	Article providing an overview of The Open Data Cities Conference (20th April 2012)	25th April 2012
OKFN	Article examining the Data Journalism handbook, and why it is important	25th April 2012
CO	70 billion of potential government business published to boost UK growth	26th April 2012
GDB	Announcement of the Data Journalism Awards	27th April 2012
GDB	Article discussing Data Journalism handbook, how to use it (School of Data Journalism)	27th April 2012
GDB	Announcement of the City Dashboard for major UK cities	27th April 2012
CO	Chair of the Open Data User Group appointed	21st May 2012
PTB	HM Revenue and Customs releasing new Tax calculator application. Department for Culture Media and Sport reviewing their priorities for the Open Data Strategy. Sport England and Fitness Industry Association presented new application spogo to find fitness centres direction is towards the Olympics 2012. Open Data white paper reviewed again	22nd May 2012
CO	Plans to establish Open Data Institute published	22nd May 2012
GDB	Article discussing how anyone can be a Data Journalist	24th May 2012

OKFN	Article discussing why Open Data isn't enough.	28th May 2012
GDB	Article about the limitations of Open Data	31st May 2012
HM	Open Data White Paper Unleashing the power of Open Data	1st June 2012
OKFN	Article requesting support for the Open Science Hackday (7th July 2012)	15th June 2012
CO	New data releases mark an era of open data	28th June 2012
GDB	Article providing an overview of the Open Data White Paper	28th June 2012
OKFN	UK Government Releases Open Data White Paper and new Data.Gov.UK	28th June 2012
OKFN	Press release of the Linked Open Vocabularies (LOV), enabling the vocabulary commons	10th July 2012
CO	Open Data User Group to help unleash the potential of open data	10th July 2012
GDB	Article on the Role of Open Data and Science	20th July 2012
CO	First publication of government services data paves way for efficiency savings	24th July 2012
OKFN	Recap of the Open Science Hackday (7th July 2012)	24th July 2012
GDB	Article discussing Open Data and London Olympics 2012	3rd August 2012
GDB	Announcement of release of London 2012 Olympic data	3rd August 2012
GDB	Article on what a data scientist is, and what is it worth	7th August 2012
OKFN	Article examining the benefits of global transparency	24th August 2012

GDB	Article on the outcomes of the Development Data Hackday	29th August 2012
OKFN	Article requesting support for the Development Data Challenge	31st August 2012
OKFN	Article describing the first commercial open data curation project!	4th September 2012
GDB	Announcement of MP Expenses data for 2011/2012	7th September 2012
CO	Open Data Institute appoints CEO	12th September 2012
GDB	Announcement of 20 million cultural data records in Europe	12th September 2012
GDB	Article of Data Journalism	20th September 2012
OKFN	Article providing research on the economic benefits of Open Data	5th October 2012
OKFN	Press Release, US Congress opened up data	9th October 2012
GDB	Article on how Government Changes are killing Open Government Data	9th October 2012

Table C.2: Events and Activities

C.3 Additional Documents and Web Sources

Source	**Description**	**Date**
OKFN-mailing Lists	Various Mailing lists co-ordinated and maintained by the OKFN, available at http://lists.okfn.org	January 2008 - December 2012

Office of Public Sector Information Reports	Reports by the Office of Public Sector Information regarding their activities and minutes from their meetings	January 2007 - December 2012
Mayor of London Reports	Monthly report of the London Mayor's Assembly	August 2008 - May 2012
OKFN Blog	Additional articles collected from the Open Knowledge Foundation Blog	July 2007 - December 2008
Academic Publications	Various publications by actors associated with the UK OGD community	January 2008 - December 2012

TABLE C.3: Additional Documents and Web Sources

Note: documents in Table C.3 were identified by references within other sources of data. Each entry may represent more than one document.

Appendix D

Twitter Analytics and Algorithms

D.1 Overview

This appendix contains the algorithms developed and used within the analysis process of this thesis. The appendix will draw upon related work and algorithms published in (Tinati, Carr & Hall 2012) as well as a further algorithms and software tool known as 'Flow140', developed alongside Edelman Ltd which are subject to patent pending (Submitted 2013).

D.2 Identifying information diffusion in Twitter networks

One aspect of the analysis used in this thesis concentrates on how the diffusion (also referred to as the flow) of information or content can help identify different users within a network, which in the case of the analysis undertaken, helps identify different individuals based on their Twitter user account. This technique is used for the purposes of data extraction and user identification, which then helps informs the process of selecting individuals as interviewees.

The follow section will describe the processing algorithm used to identify user roles within a Twitter communications network, which is performed in the pre-processing engine as show in Figure D.1.

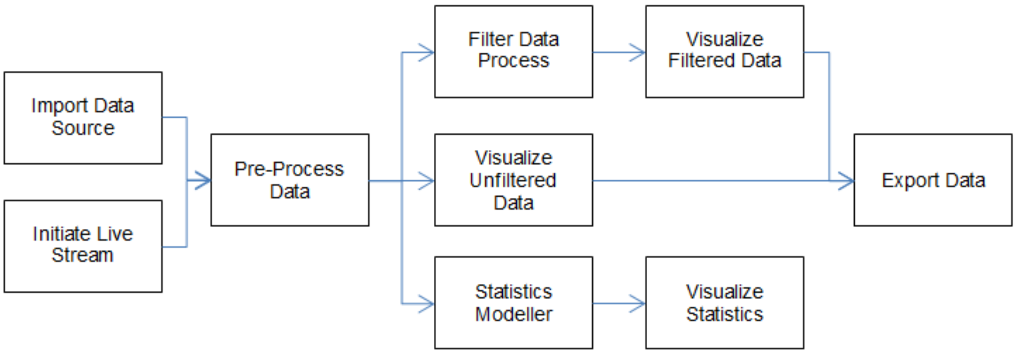


Figure D.1: Flow 140 Twitter processing and engine

D.2.1 Classification of Twitter User Roles

Flow140's pre-processing engine Figure D.1 uses a set of procedures to parse the imported or live stream data to the underlying data structure used within filtering and visualisation engine. The procedures taken to transform the data into Flow140s data structure require several unique processes including:

1. Extracting specific fields from each tweet message text
2. Pattern matching and verification of similarities between tweet message texts
3. Constructing information flow pathways for matched tweet message text records
4. Constructing communication networks from extracted fields from tweet message text

D.2.2 Filtering Model: Classification of Different Twitter Users

Flow140's filtering model provides a way to classify Twitter users within the constructed communication network based upon their unique network characteristics. The classification provides a categorization of users, which are not mutually exclusive, as often the categorization of users behaviour falls under multiple categories. The categories defined are used as a base measurement to classify Twitter users in terms of their basic characteristics within the communication network. Working with this, Flow140 provides an additional set of metrics to extract the sub-set of Twitter users that are classified as more than one type of user. As a result of this there are a total of 9 categories of user type, which are all visually represented by a colour scheme by the visualisation engine:

1. Idea Starter
2. Amplifier
3. Curator
4. Commentator
5. Viewer (Not visible)

D.2.2.1 Idea Starter

These can be considered as an individual who starts a conversational meme. They tend to be highly engaged with the media, in the on- and off-line environment. They utilize multiple sources of social media, but have an intricate network of trusted relationships, especially online. As a result of this, their network of connections is usually limited, but this ensures that the connections are of high quality Although Idea Starters may not be the one with the bright idea, they are the ones which start the conversation, and due to its trusted connections, are in a fertile environment for the idea to grow.

D.2.2.2 Amplifier

These can be considered as an individual who collates multiple thoughts and shares ideas and opinions. Amplifiers thrive off sharing opinions of others; enjoy being the first to do so. They have a large network of connections and are trusted within their community. Although they do not synthesis the information being shared, they tend to be the firehose of knowledge. Amplifiers tend to be the individuals that are part of small trusted network of certain idea starters, taking their original ideas and sharing them to a larger, more visible audience. Due to this process, there is the risk that idea starters will slowly become amplifiers over time due to increased exposure.

D.2.2.3 Curator

These can be considered as an individual who use a broader context to define ideas. Curators tend to offer a level of transparency beyond that of Idea Starters and Amplifiers. By following the conversation path, they have an impact on the way the conversation is shaped and spread. They take the ideas of the idea starters and the amplifiers and either validate, question, challenge, or dismiss them. They are the ties that form between the Idea Starters and Amplifiers, aggregating the ideas together to help clarify and steer the topic of conversation. Curators are connected to a large audience, and often pick up information outside their primary community of interest tailoring the information to suit their networks circle of interest.

D.2.2.4 Commentator

These can be considered as an individual who detail and refine ideas. Commentators add to or adapt the flow of conversation, adding in their own opinions, insights, but without becoming too immersed in the conversation. Unlike the other categories described so far, Commentators are do not seek recognition of their leadership, or want to increase their status, they are taking part in something to which they strongly feel about. They want to share the conversation not for self-benefit.

D.2.2.5 Viewer

These can be considered as an individual who detail and refine ideas. Commentators add to or adapt the flow of conversation, adding in their own opinions, insights, but without becoming too immersed in the conversation. Unlike the other categories described so far, Commentators are do not seek recognition of their leadership, or want to increase their status, they are taking part in something to which they strongly feel about. They want to share the conversation not for self-benefit.

D.2.3 Implementation of Filtering Model

Based on these principles, the classification of the different user types can be made clearer. Figure D.2 a, b, and c provide the underlying model for the filtering algorithmic process.

Within this model, an idea starter is categorized as a user who has their tweets retweeted by a large number of people thus suggesting that their ideas are of value to share. An idea starter is calculated by finding the sum of all the retweets of a user divided by the minimum retweet number, if this is greater than one, then they are classified as an idea starter.

An amplifier is a user who is the initial person to retweet a tweet one which is part of a retweet chain of length n. We first use the pre-processing engine to calculate retweet chains and then apply the model to calculate if a specific user is an amplifier. Furthermore, as an additional metric to further distinguish between the users categorized as amplifiers, the model calculates the number of times that they are the n+1 user in the retweet chain. An amplifier is calculated by finding the sum of all the users tweets divided by the number of tweets which are retweets. This is then multiplied by the number of tweets which are first in a chain of retweets.

A curator is a user who is part of two or more retweet chains, where their position in the chain is n. The model calculates a curator as a user who retweet two or more idea starters, thus acting as a point of aggregation for their own network. The pre-processing engine is first used to calculate the retweet chains, curators are then calculated by finding the sum of all the users tweets which are retweets divided by the number of users (which are idea starters themselves) that the user has retweeted from. To further distinguish between the users categorized as curators, the range of curators is calculated based upon the number of retweet chains they feature in.

(a). Idea Starter

$$\frac{\sum U^{rt}}{RT^{min}} > 1$$

Where U^{rt} is number of retweets of a user and RT^{min} is minimum retweet number

(b). Amplifier

$$\frac{\sum U^{t}}{\sum RT^{u} \times \sum rt^{orig}} > 1$$

Where U^{t} is number of user's tweets, RT^{u} is number of user's retweets, and RT^{orig} is number of retweets which were first in retweet chain

(c). Curator

$$\frac{\sum RT^{u}}{\sum U^{uniqRT}} > 2$$

Where RT^{u} is number of a user retweets and U^{uniqRT} is number of unique number of users that a user has retweeted.

Figure D.2: Formal solution for Twitter user roles identification based on diffusion chain construction

Finally commentators have been modelled as users who do not meet the above requirements, but have actively retweeted within the harvested data.

As this model only considers active users within the harvested data, viewers will not be represented by this model. The classification of the different users within the model will be based on a single factor which is controlled by the operator the number of retweets a user needs to be classified as an idea starter. This in turn affects the users represented as amplifiers and curators. The aim of having a variable which enables the overall structure of the graph to change offers a way to explore the classification ratio of users based on the minimum number of retweets needed to be an idea starter.

D.2.4 Constructing Information Diffusion Chains

In order for Flow140 to restructure the data ready for analysis, a sub-progress within the pre-processing stage requires the construction of retweet chains which consists of matching tweets of similar content in order to examine how information (which may include, URLs, Images, Videos, other media content, Twitter users, or words [single or multiple strings]) is diffused within the network. The construction of the retweet uses a fuzzy string matching algorithm which involves two components:

- Matching of the Tweet's text this involves matching the string patterns between each of the tweets within the inputted dataset. In order to achieve this, the algorithm developed uses an approximate string matching technique as well as incorporating the Levenshtein Distance calculation; both are required to ensure that the tweets are similar, containing the same content. The algorithm developed takes into consideration small variations within the tweets text, which may include difference such as: extra 'white spaces', grammatical errors, spurious ASCII characters. Given specific accuracy (which can be set by the user), If the tweets being compared fall within the boundaries set, they are identified as part of the same retweet chain
- Constructing the Retweet Chain once the tweets that are statistically similar have been identified, the construction of the retweet chain requires a temporal analysis of the tweets within the chain to firstly ensure that they are actually part of the same chain, and secondly to restructure the retweet chain to represent the

> diffusion pattern in chronological order. Two parameters are considered during the validation and reordering process, time between retweets, and volume of retweets. Based upon the combination of these two parameters the algorithm statistically validate if the tweets identified as part of the same retweet chain, and if validation is passed, then the chain is reordered to represent the diffusion of information overtime.

Once the retweet chains have been constructed, Flow140 is able to continue to process the data and apply the filtering and identification mechanisms described next.

```
Data: Twitter Dataset
Result: Content Diffusion Chains
initialization;
for Tweet t in Dataset* do
    Construct Sub-Dataset without the current tweet;
    for Tweet t1 in sub-dataset do
        Calculate Levenshtein distance (LD) between t and t1 ;
        if if LD(t1,t0) > minimum similarity value** then
            Add a new diffusion tree or add to end of existing tree;
        end
    end
end
```

Algorithm 1: *The Dataset needs to be already sorted (ascending) by tweet timestamp, **the minimum similarity value can determine how similar the tweet text string needs to be in order to match. 95% tends to match tweet chains that have one or two missing/additional characters

D.3 Community Detection in Twitter

The detection of Twitter communication networks used in this thesis draws upon methods identified by Stienhaeuser and Chawla Steinhaeuser & Chawla (2008) on examining edge weights, the use of traditional social network analysis clustering techniques such as centrality and degree (Girvan & Newman 2002), and also research which identify communities based on network modularity (Newman & Girvan 2004). The community detection techniques used in this analysis also draw upon windowing and sampling technique techniques in literature surrounding the emerging field of Dynamic Network Analysis (Carley 2003).

The techniques developed in the analysis to expose the different Twitter communities that emerge overtime were based on a number of principles, as described below:

- A community is identified by using a 24-hour windowing technique applied across the entire dataset collected. The time period of a window is identified by using the first identified timestamp within the dataset, which is then advanced by +24:00 hours for each window. This +24:00 hours is labelled the 'tick' size.
- At each window, two algorithms are applied to the communications (retweets and mention network graph calculated separately) that fall within the time period.
- 1. The identification of nodes (Twitter users), and the edges (tweets or retweets) made between each other), this is collected alongside the node's metrics of in- and out-degree. Thus at each windowing period each node contains the following metrics show in Table D.1
- 2. Based upon the calculation of each nodes metrics, the association matrix for each node is calculated to identify the connected sub-networks within the mention and retweet network
- Each window is calculated individually. For each window calculated, the results of the node metrics and the association matrixes are stored in a list which then provide a 'static snapshot' of the network advanced by a 24:00 hour time period.

These 24-hour snapshots then can be stitched together to calculate the temperal dynamics of the network, thus the become essential to identify the existence of communities within the retweet and mention network.

```
Data: The Window List
Result: Retweet or Mention Communication Communities
initialization;
// Begin by splitting the dataset into 24 hour windows without overlap
for Window: w do
    for Node: n in w do
        Construct node graph w(n(g)) and metrics* for node n;
    end
end
// Compare and match association (community) matrix with previous windows calculated
for Window: w do
    for Node: n in w do
        for Node: n1 in w do
            if n(g) matches* n1(g) then
                if Association matrix w(am) then
                    Add n to association matrix w(am) if exists and append score value ;
                else
                    Create new association matrix w(am) and add n;
                end
            end
        end
    end
end
// Compare Windows to identify dynamic community structures
for Window: w do
    for Association matrix: am in w do
        for Window: w1 do
            if(w(am) matches** w1(am) if dynamic community: dm then
                Add w(am) to dynamic community dm and append score value;
            else
                Create new dynamic community dm and add w(am);
            end
        end
    end
end
```

Algorithm 2: *The Metrics are those discussed in Table D.1 **The match between association matrix can return True if the comparison matches a sub-set of the retweet-/mention matrix

As Algorithm 2 describes, the algorithm compares the graphs (and the subgraphs that exist) identified in each 24 hour window to see if there are any similarities between edges of users, e.g. do users in t0 communicate (retweet or mention) with the same users in t1 . . . tn. Each time a match is identified, an arbitrary community score value is increased. The comparison between communities within a window uses a function which enables

Table D.1: Metrics calculated for both the Mention and retweet graph between, *These metrics are used to construct the association matrices which represent the community that a specific user belongs to within a given window

Metric Name	**Metric Description**
Node in-degree	Number of mentions or retweet received
Node out-Degree	Number of mentions or retweets made
Retweet_by_list*	List of users that a user has been retweeted by
Retweed_list*	List of users that a user has retweeted
Mention_by_List*	List of users that a user has been mentioned by
Mentioned_list*	List of users that a user has mentioned

subsets of communities to match in order to identify communities which are potentially growing or shrinking in users. This is reflected in addition community score values. This is used to help identify the growth or shrinking of specific communities.

Note: the detection of these communities relies on consistency within the data, specifically with regards to the twitter username or twitter userID (both are used as methods of user identification).

The labeling of these communities is then achieved by using external data sources which can help validate or act as a point of reference. For example, within the study of the Open Government Data community, communities (or actor-networks) the 'academics' were identified by actors associated with material which related to either academic publications, other communities such as the London network were identified by interview with members within the community.

The visualisation of these networks is achieved by extracting the processed data using Flow140 and then importing the Node and Edge list into network visualisation software packages such as Gephi, NodeXL, or Network Workbench. These software packages provide a solution to display the network with various visual techniques and graph layouts (for this analysis, Fruchterman-Reingold layout was used). In order to use these visualisations as part of the mixed methods approach used within the analysis of OGD, overlays were also applied (manually) to the graphs to help identify the different communities that were identified during the initial analysis of the data.

Bibliography

Adamic, L., Buyukkokten, O. & Adar, E. (2003), 'A social network caught in the Web', *First Monday* **8**(6).

Ahmed, N. & Berchmans, F. (2010), 'Time-based sampling of social network activity graphs', *Machine Learning with Graphs* .
URL: *http://dl.acm.org/citation.cfm?id=1830253*

Ahn, Y., Han, S., Kwak, H., Moon, S. & Jeong, H. (2007), Analysis of topological characteristics of huge online social networking services, *in* 'Proceedings of the 16th International ACM World Wide Web Conference', pp. 835–844.
URL: *http://dl.acm.org/citation.cfm?id=1242685*

Akrich, M. (1992), The de-scription of technical objects, *in* W. E. Bijker & J. Law, eds, 'Shaping technology building society', Vol. pp of *Inside Technology*, MIT Press, chapter 7, pp. 205–224.

Akrich, M. & Latour, B. (1992), A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies, *in* W. E. Bijker & J. Law, eds, 'Shaping Technology/Building Society', MIT Press, Cambridge, pp. 259–264.

Alani, H., Dupplaw, D., Sheridan, J., Hara, K. O., Darlington, J., Shadbolt, N. & Tullo, C. (2008), Unlocking the Potential of Public Sector Information with Semantic Web Technology, *in* 'The 6th International Semantic Web Conference (ISWC)', Busan, Korea.

Albert, R. & Barabasi, A.-L. (2000), 'Topology of evolving networks: local events and universality', *Physical Review Letters* **85**(24), 13.
URL: *http://arxiv.org/abs/cond-mat/0005085*

Albert, R., Jeong, H. & Barabasi, A.-L. (1999), 'Diameter of the World-Wide Web', *Nature* **401**(6749), 130–131.
URL: *http://dx.doi.org/10.1038/43601*

Alexa Internet (2013), 'World Internet Statistics'.
URL: *http://www.internetworldstats.com/stats.htm, Accessed 1st May 2013*

Altheide, D. L. & Johnson, J. M. (1994), Criteria for assessing interpretive validity in qualitative research, *in* N. K. Denzin & Y. S. Lincoln, eds, 'Handbook of Qualitative Research', Vol. 1, Sage, pp. 485–499.

Anger, I. & Kittl, C. (2011), Measuring Influence on Twitter, *in* 'Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies'.

Ayyad, M. (2009), 'Using the Actor-Network Theory to interpret e-government implementation barriers', *Proceedings of the 3rd International Conference on Theory and Practice of Electronic Governance - ICEGOV '09* p. 183.
URL: *http://portal.acm.org/citation.cfm?doid=1693042.1693080*

Bagrow, J. P. & Bollt, E. M. (2004), 'A Local Method for Detecting Communities', *Physical Review E - Statistical, Nonlinear and Soft Matter Physics* **72**, 046108.

Bakshy, E., Hofman, J. M., Watts, D. J. & Mason, W. A. (2011), Everyones an Influencer: Quantifying Influence on Twitter , *in* 'WSDM'11'.

Ball, B. & Newman, M. (2013), 'Friendship networks and social status', *Network Science* **1**(01), 16–30.
URL: *http://www.journals.cambridge.org/abstract_S2050124212000045*

Barabási, A.-L. (2009), 'Scale-free networks: a decade and beyond.', *Science (New York, N.Y.)* **325**(5939), 412–3.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19628854*

Barabasi, A.-L. & Albert, R. (1999), 'Emergence of Scaling in Random Networks', *Science* **286**(October), 509–512.

Barassi, V. & Trere, E. (2012), 'Does Web 3.0 come after Web 2.0? Deconstructing theoretical assumptions through practice', *New Media & Society* **0**(0), 1–17.

Barnes, J. A. (1954), 'Class and Committees in a Norwegian Island Parish', *Human Relations* **7**(1), 39–58.
URL: *http://hum.sagepub.com/cgi/doi/10.1177/001872675400700102*

Baxter, G. & Sommerville, I. (2011), 'Socio-technical systems: From design methods to systems engineering', *Interacting with Computers* **23**(1), 4–17.

Baygeldi, M. & Smithson, S. (2004), Ability of the Actor Network Theory (ANT) to Model and Interpret an Electronic Market, *in* J. N. D. Gupta & S. Sharma, eds, 'Creating Knowledge Based Organizations', IGI Global, pp. 109–126.
URL: *http://www.igi-global.com/chapter/ability-actor-network-theory-ant/7211/*

Beauvisage, T., Beaudouin, V. & Assadi, H. (2007), 'Internet 1.0: early users, early uses', *Annals of Telecommunications* **62**(5).

Becker, H. S. (1958), 'Problems of inference and proof in participant observation', *American Sociological Review* **23**(6), 652–660.
URL: *http://www.jstor.org/stable/2089053*

Berger, P. L. & Luckmann, T. (1967), *The Social Construction of Reality*, Vol. 51, Doubleday.
URL: *http://www.jstor.org/stable/323448?origin=crossref*

Berners-Lee, T. (1988), Information Management: A Proposal, Technical report, CERN.
URL: *http://www.w3.org/History/1989/proposal.html*

Berners-Lee, T. (1999), *Weaving the Web*, Harper Collins.

Berners-Lee, T. (2007), The Two Magics of Web Science, *in* 'Keynote Speech at 16th International World Wide Web Conference (WWW2007)', Banff: Canada.

Berners-Lee, T., Hall, W., Hendler, J. & Shadbolt, N. (2005), 'Creating a Science of the Web', *Artificial Intelligence* **2005**(September).

Berners-Lee, T., Weitzner, D. J., Hall, W., O'Hara, K., Shadbolt, N. & Hendler, J. a. (2006), 'A Framework for Web Science', *Foundations and Trends in Web Science* **1**(1), 1–130.
URL: *http://www.nowpublishers.com/product.aspx?product=WEB&doi=1800000001*

Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A. & Bar-yam, Y. (2013), 'Sentiment in New York City : A High Resolution Spatial and Temporal View', *NECSI Report* pp. 1–12.

Beyer, M. A. & Laney, D. (2012), The Importance of 'Big Data': A Definition, Technical report, Gartner.

Bijker, W., Hughes, T. P. & Pinch, T. (1989), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, The MIT Press.
URL: *http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0262521377*

Bird, C. (2011), Sociotechnical coordination and collaboration in open source software, *in* 'Proceedings of the 27th IEEE International Conference on Software Maintenance', IEEE Comput. Soc.
URL: *http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6080832*

Bloor, D. (1976), *Knowledge and Social Imagery*, second edi edn, University of Chicago Press.

Bloor, D. (1981), 'The Strengths of the Strong Programme', **11**, 199–213.

Bourdieu, P. (1984), *Distinction: A Social Critique of the Judgement of Taste*, Vol. 37, Harvard University Press.
URL: *http://www.jstor.org/stable/590650?origin=crossref*

Boyd, d. & Crawford, K. (2011), 'Six Provocations for Big Data', *Computer* pp. 1–17.

Boyd, D., Golder, S. & Lotan, G. (2010), Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter, *in* 'HICSS-43', IEEE Comput. Soc, Kauai, HI.

Brandes, U., Robins, G., McCraine, A. & Wasserman, S. (2013), 'What is network science?', *Network Science* **1**(01), 1–15.
URL: *http://www.journals.cambridge.org/abstract_S2050124213000027*

Broder, A., Kumar, R., Maghoul, F. & Raghavan, P. (2000), 'Graph structure in the Web', *Computer Networks* **33**, 309–320.

Broder, A. Z., Glassman, S. C., Manasse, M. S. & Zweig, G. (1997), 'Syntactic Clustering of the Web', *6th International Conference on the World Wide Web* pp. 1–13.

Brooks, L., Fitzgerald, G. & Atkinson, C. (2008), 'Using Actor Network Theory to Interpret the Introduction of Information Systems within the London Ambulance Service Using Actor Network Theory to Interpret the Introduction of Information Systems within the London Ambulance Service', *Information Systems* .

Burt, R. S. (1978), 'Applied Network Analysis: An Overview', *Sociological Methods Research* **7**(2), 123–130.
URL: *http://smr.sagepub.com/cgi/doi/10.1177/004912417800700201*

Callon, M. (1986*a*), Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay, *in* J. Law, ed., 'Action and Belief: A New Sociology of Knowledge', Routledge, pp. 196–223.

Callon, M. (1986*b*), 'The sociology of an actor-network: The case of the electric vehicle', *Mapping the dynamics of science and technology* .
URL: *http://epl.scu.edu:16080/ stsvalues/readings/Callon.pdf*

Callon, M. (1992), Techno-economic networks and irreversibility, *in* J. Law, ed., 'A Sociology of Monsters: Essays on Power, Technology and Domination', Routledge, London, UK, pp. 132–161.

Capocci, A., Servedio, V. D. P., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S. & Caldarelli, G. (2006), 'Preferential attachment in the growth of social networks: the case of Wikipedia', *Physical Review E* **74**(3), 4.
URL: *http://arxiv.org/abs/physics/0602026*

Carey, J. W. (1993), 'Linking Qualitative and Quantitative Methods: Integrating Cultural Factors into Public Health', *Qualitative Health Research* **3**(3), 298–318.
URL: *http://qhr.sagepub.com/cgi/doi/10.1177/104973239300300303*

Carley, K. M. (2003), 'Dynamic Network Analysis'.

Carrington, P. J., Scott, J. & Wasserman, S. (2005), *Models and methods in social network analysis*, Vol. 10 of *Structural analysis in the social sciences*, Cambridge University Press.
URL: *http://www.amazon.com/dp/0521600979*

Casebeer, A. L. & Verhoef, M. J. (1997), 'Combining qualitative and quantitative research methods: considering the possibilities for enhancing the study of chronic diseases.', *Chronic Diseases in Canada* **18**(3), 130–135.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/9375260*

Chakrabarti, D. & Faloutsos, C. (2012), *Graph Mining: Laws, Tools, and Case Studies*, Morgan & Claypool Publishers.

Chakrabarti, S. (2002), *Mining the Web: Discovering Knowledge from Hypertext Data*, Vol. 55, Morgan Kaufmann.
URL: *http://doi.wiley.com/10.1002/asi.10382*

Chakrabarti, S., Dom, B. E., Gibson, D., Kleinberg, J. & Kumar, R. (1999), 'Mining the Link Structure of the World Wide Web', *World Wide Web Internet And Web Information Systems* pp. 1–12.

Chakrabarti, S., Dom, B., Raghavan, P., Rajagopalan, S., Gibson, D. & Kleinberg, J. (1998), 'Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text', *World Wide Web Conference - 7th Preceeding of* .

Chen, H.-l. (2009), 'The use and sharing of information from Wikipedia by high-tech professionals for work purposes', *The Electronic Library* **27**(6), 893–905.
URL: *http://www.emeraldinsight.com/10.1108/02640470911004011*

Christensen, C. (2011), 'Discourses of Technology and Liberation: State Aid to Net Activists in an Era of Twitter Revolutions', *The Communication Review* **14**(3), 233–253.
URL: *http://www.tandfonline.com/doi/abs/10.1080/10714421.2011.597263*

Chu, Z., Gianvecchio, S. & Wang, H. (2010), Who is Tweeting on Twitter: Human, Bot, or Cyborg?, *in* 'Proceedings of the 26th Annual Computer Security Applications Conference', pp. 21–30.

Clark, A. (2007), 'Understanding Community: A review of networks, ties and contacts', *Methods* **44**(May), 39.
URL: *http://eprints.ncrm.ac.uk/469/*

Clarke, P. N. & Yaros, P. S. (1988), 'Research Blenders: Commentary and response', *Nursing Science Quarterly* **1**, 147–150.

Clauset, A. (2005), 'Finding local community structure in networks.', *Physical Review E - Statistical, Nonlinear and Soft Matter Physics* **72**, 7.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonc, B., Flammini, A. & Menczer, F. (2011), Political Polarization on Twitter, *in* 'Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media', AAAI, pp. 89–96.

Cordella, A. (2003), 'Actor Network Theory and After: What's New for IS Research?', *Information Systems* .

Cordella, A. (2011), 'Emerging Standardization', *International Journal of Actor-Network Theory and Technological Innovation* **3**(3), 49–64.
URL: *http://eprints.lse.ac.uk/43664/*

Creswell, J. (2003), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Sage Publications.

De Choudhury, M., Counts, S. & Horvitz, E. (2013), 'Social media as a measurement tool of depression in populations', *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13* pp. 47–56.
URL: *http://dl.acm.org/citation.cfm?doid=2464464.2464480*

Denzin, N. K. (1978), *The Research Act: A Theoretical Introduction to Sociological Methods*, McGraw-Hill, New York, NY, USA.

Denzin, N. K. & Lincoln, Y. S. (1994), Introduction: Entering the Field of Qualitative Research, *in* N. K. Denzin & Y. S. Lincoln, eds, 'Handbook of Qualitative Research', Vol. 1, Sage Publications, chapter 1, pp. 1–17.

Depauw, J. (2008), Web2.0 under the Actor-Network Theory point-of-view: Conceptualization and Definition Analysis, *in* 'Politics: Web 2.0: an International Conference', number April.

Dhar, V. I. (2012), 'Data Science and Prediction', *SSRN Electronic Journal* (May).

Dill, S., Kumar, R., Mccurley, K. S., Rajagopalan, S., Sivakumar, D. & Tomkins, A. (2002), 'Self-similarity in the web', *ACM Transactions on Internet Technology* **2**(3), 205–223.
URL: *http://portal.acm.org/citation.cfm?doid=572326.572328*

Dourisboure, Y., Geraci, F. & Pellegrini, M. (2009), 'Extraction and classification of dense implicit communities in the Web graph', *ACM Transactions on the Web* **3**(2), 1–36.
URL: *http://portal.acm.org/citation.cfm?doid=1513876.1513879*

Edwards, G. (2010), 'Mixed-Method Approaches to Social Network Analysis', *Research Methods* (January).

Edwards, G. & Crossley, N. (2009), 'Measures and Meanings: Exploring the Ego-Net of Helen Kirkpatrick Watts, Militant Suffragette', *Methodological Innovations Online* **4**, 37–61.

Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M. & Apkarian, A. V. (2003), 'Scale-free brain functional networks.', *Physical Review Letters* **94**(1), 4.
URL: *http://arxiv.org/abs/cond-mat/0309092*

Emirbayer, M. & Goodwin, J. (1994), 'Network Analysis, Culture, and the Problem of Agency', *American Journal of Sociology* **99**(6), 1411.
URL: *http://www.journals.uchicago.edu/doi/abs/10.1086/230450*

Emmel, N. (2009), 'Participatory Mapping : An innovative sociological method', *ESRC National Centre for Research Methods* pp. 1–8.

Erdös, P. & Rényi, A. (1959), 'On random graphs', *Publ Math Debrecen* **6**(290-297), 290–297.
URL: *http://www.citeulike.org/group/3509/article/4012374*

Erzverger, C. & Kelle, U. (2003), Making inferences in mixed methods: The rules of integration, *in* A. Tashakkori & C. Teddlie, eds, 'Handbook of mixed methods in social and behavioral research', Sage Publications, pp. 457–488.

Eschenfelder, K. & Chase, L. (2002), 'Socio-technical networks of large, post-implementation Web information systems: tracing effects and influences', *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* **00**(c), 1112–1121.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=994032*

Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999), On power-law relationships of the internet topology, *in* 'ACM SIGCOMM Computer Communication Review', Vol. 29,

ACM, pp. 251–262.
URL: *http://dl.acm.org/citation.cfm?id=316229*

Faraj, S., Kwon, D. & Watts, S. (2004), 'Contested artifact: technology sensemaking, actor networks, and the shaping of the Web browser', *Information Technology & People* **17**(2), 186–209.
URL: *http://www.emeraldinsight.com/10.1108/09593840410542501*

Fiedler, M. (1973), 'Algebraic connectivity of graphs', *Czechoslovak Mathematical Journal* **23**(98), 298–305.
URL: *http://dspace.dml.cz/handle/10338.dmlcz/101168*

Flake, G., Lawrence, S. & Giles, C. (2002), 'Self-organization and identification of web communities', *Computer* .
URL: *http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=989932*

Flake, G., Lawrence, S. & Giles, C. L. (2000), 'Efficient Identification of Web Communities', *Communities* pp. 150–160.

Fuchs, C. (2003), 'The Internet as a Self-Organizing Socio-Technological System', *SSRN Electronic Journal* pp. 1–19.
URL: *http://www.ssrn.com/abstract=458680*

Fuchs, C. (2009), 'Information and Communication Technologies and Society', *European Journal of Communication* **24**(1), 69–87.

Gao, Q., Abel, F., Houben, G.-j. & Yu, Y. (2012), 'A Comparative Study of Users Microblogging Behavior on Sina Weibo and Twitter', pp. 88–101.

Gargi, U., Lu, W., Mirrokni, V. & Yoon, S. (2011), Large-Scale Community Detection on YouTube for Topic Discovery and Exploration, *in* 'Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media'.
URL: *http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/*

Georgas, J. & Sarma, A. (2011), 'STCML: an extensible XML-based language for socio-technical modeling', *Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering* pp. 61–64.
URL: *http://dl.acm.org/citation.cfm?id=1984657*

Gibson, D., Kleinberg, J. & Raghavan, P. (1998), Inferring Web Communities from Link Topology, *in* 'HyperText 98', pp. 225–234.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. & Brilliant, L. (2009), 'Detecting influenza epidemics using search engine query data.', *Nature* **457**(7232), 1012–4.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/19020500*

Girvan, M. & Newman, M. E. J. (2002), 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences of the United States of America* **99**(12), 7821–6.
URL: *http://www.pnas.org/content/99/12/7821.full.pdf*

Glaser, B. G. & Strauss, A. L. (1967), *The Discovery of Grounded Theory*, Aldine.
URL: *http://books.google.com/books?id=tSi7KiOHkpYC*

Golafshani, N. (2003), 'Understanding reliability and validity in qualitative research', *The qualitative report* **8**(4), 597–606.
URL: *http://www.nova.edu/ssss/QR/QR8-4/golafshani.pdf*

Gong, A. (2011), 'An Automated Snowball Census of the Political Web', *SSRN Electronic Journal* pp. 1–34.
URL: *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1832024*

Granovetter, M. S. (1973), 'The strength of weak ties', *American Journal of Sociology* **78**(6), 1360–1380.
URL: *http://www.journals.uchicago.edu/doi/abs/10.1086/225469*

Greene, J. C., Caracelli, V. J. & Graham, W. F. (1989), 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational Evaluation & Policy Analysis* **11**(3), 255–274.
URL: *http://epa.sagepub.com/cgi/doi/10.3102/01623737011003255*

Guba, E. G. (1987), 'What have we learned about naturalistic evaluation?', *Evaluation Practice* **8**(1), 23–43.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0886163387800375*

Guba, E. G. & Lincoln, Y. S. (1994), Competing paradigms in qualitative research, *in* N. K. Denzin & Y. S. Lincoln, eds, 'Handbook of qualitative research', Vol. 2 of

Handbook of Qualitative Research, Sage, chapter 6, pp. 105–117.
URL: *http://create.alt.ed.nyu.edu/courses/3311/reading/10-guba_lincoln_94.pdf*

Gunawong, P. (2010), 'Understanding eGovernment Failure: An Actor- Network Analysis of Thailand's Smart ID Card Project', *Information Systems* .

Haase, J. E. & Myers, S. T. (1988), 'Reconciling paradigm assumptions of qualitative and quantitative research.', *Western Journal of Nursing Research* **10**(2), 128–137.

Halford, S., Pope, C. & Carr, L. (2009), 'A Manifesto for Web Science?', *Proceedings of the Web Science 2009* pp. 1–6.

Hall, W. (2011), 'The Ever Evolving Web: The Power of Networks', *Journal of Communication* **5**, 651–664.

Hall, W., Shadbolt, N., Tiropanis, T., Hara, K. O. & Davies, T. (2012), 'Open data and charities', *NominetTrust* .

Hall, W. & Tiropanis, T. (2012), 'Web evolution and Web Science', *Computer Networks* **56**(18), 3859–3865.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S1389128612003581*

Han, I. & Lee, S. (2007), 'Graph structure of the Korea web', *Advances in Databases: Concepts, Systems and Applications* pp. 930–935.
URL: *http://www.springerlink.com/index/H32M46828H181326.pdf*

Han, J. & Kamber, M. (2000), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers Inc.

Hannemyr, G. (2003), 'The Internet as Hyperbole: A Critical Examination of Adoption Rates', *The Information Society* **19**(2), 111–121.
URL: *http://www.tandfonline.com/doi/abs/10.1080/01972240309459*

Hanseth, O. & Monteiro, E. (1997), 'Inscribing Behaviour In Information Infrastructure Standards', *Accounting, Management and Information Technnology* **7**(4), 183–211.

Hanseth, O. & Monteiro, E. (1998), CHAPTER 6 - Socio-technical webs and actor network theory, *in* 'Understanding Information Infrastructure'.

Heath, S., Fuller, A. & Johnston, B. (2010), 'Chasing shadows: defining network boundaries in qualitative social network analysis', *Qualitative Research* **9**(5), 645–661.
URL: *http://dx.doi.org/10.1177/1468794109343631*

Hedrick, T. E., Bickman, L. & Rog, D. J. (1993), *Applied Research Design: A Practical Guide*, Sage Publications Ltd.

Heeks, R. (2006), *Implementing and Managing eGovernment*, Sage Publications.

Heidegger, M. (1977), *The Question Concerning Technology, and Other Essays*, Harper & Row, New York, New York, USA.

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T. & Weitzner, D. J. (2008), 'Web science : an interdisciplinary approach to understanding', *Communications of the ACM - Web Science* **51**(7).

Hesse-Biber, S. & Griffin, a. J. (2012), 'Internet-Mediated Technologies and Mixed Methods Research: Problems and Prospects', *Journal of Mixed Methods Research* .
URL: *http://mmr.sagepub.com/cgi/doi/10.1177/1558689812451791*

Hester, A. (2012), 'Measuring alignment within relationships among socio-technical system components: a study of Wiki technology use', *Proceedings of the 50th annual conference on SIGMIS-CPR* pp. 147–154.
URL: *http://dl.acm.org/citation.cfm?id=2214091.2214132*

HM Government (1985), Local Government (Access to Information) Act 1985, Technical report, HM Government.
URL: *http://www.legislation.gov.uk/ukpga/1985/43/pdfs/ukpga_19850043_en.pdf*

HM Government (2012), Open Data White Paper, Unleashing the Potential #opendata, Technical report, HM Government, London.
URL: *http://www.cabinetoffice.gov.uk/sites/default/files/resources/CM8353_acc.pdf*

Hogg, T. & Szabo, G. (2009), 'Diversity of User Activity and Content Quality in Online Communities', *Artificial Intelligence* pp. 58–65.

Honeycutt, C. & Herring, S. C. (2009), Beyond Microblogging: Conversation and Collaboration via Twitter, *in* '42nd Hawaii International Conference on System Sciences', Vol. 0 of *42nd Hawaii International Conference on System Sciences*, IEEE Computer

Society, pp. 1–10.
URL: *http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4755499*

Howe, K. R. (1988), 'Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard', *Educational Researcher* **17**(8), 10–16.
URL: *http://edr.sagepub.com/cgi/doi/10.3102/0013189X017008010*

Huang, X. & Lai, W. (2003), 'Identification of clusters in the Web graph based on link topology', *Seventh International Database Engineering and Applications Symposium, 2003. Proceedings.* pp. 123–128.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1214919*

Hurlock, J. & Wilson, M. L. (2011), Searching Twitter: Separating the Tweet from the Chaff, *in* 'Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media', AAAI, pp. 161–168.

Hyunwoo, C., Haewoon, K., Young-Ho, E., Yong-Yeol, A., Moon, S. & Jeong, H. (2008), 'Comparison of online social relations in volume vs interaction: a case study of cyworld', *Proceedings of the 6th ACM/Usenix Internet Measurement Conference (IMC08)* **v**.
URL: *http://dl.acm.org/citation.cfm?id=1452528*

Imafuji, N. (2003), 'Finding a Web Community by Maximum Flow Algorithm', *Database* .

Jari, S. & Kaski, K. (2004), 'Scale-free networks generated by random walkers', *Physica A: Statistical Mechanics and its Applications* **341**(80).

Jo, Y., Hopcroft, J. & Lagoze, C. (2011), 'The Web of Topics: Discovering the Topology of Topic Evolution in a Corpus', *International Conference on The World Wide Web* pp. 257–266.
URL: *http://dl.acm.org/citation.cfm?id=1963444*

Johnson, R. B. & Onwuegbuzie, a. J. (2004), 'Mixed Methods Research: A Research Paradigm Whose Time Has Come', *Educational Researcher* **33**(7), 14–26.
URL: *http://edr.sagepub.com/cgi/doi/10.3102/0013189X033007014*

Kang, J., Lee, D. & Mitra, P. (2005), 'Identifying value mappings for data integration: An unsupervised approach', *Web Information Systems EngineeringWISE ...* pp. 544–551.

URL: *http://link.springer.com/chapter/10.1007/11581062_46*

Kernighan, B. W. & Lin, S. (1970), 'An efficient heuristic procedure for partitioning graphs', *Bell System Technical Journal* **49**(2), 291–307.

URL: *http://www.cs.utexas.edu/ pingali/CS395T/2009fa/papers/kl.pdf*

King, G., Keohane, R. O. & Verba, S. (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Vol. 24, Princeton University Press.

URL: *http://www.loc.gov/catdir/toc/prin031/93039283.html*

Kleinberg, J. (1997), 'Authoritative sources in a hyperlinked environment', *Journal of the ACM* **46**(5), 604–632.

URL: *http://portal.acm.org/citation.cfm?doid=324133.324140*

Kleinberg, J. (1999), 'Hubs, authorities, and communities', *ACM Computing Surveys (CSUR)* **31**(4es), 5.

URL: *http://dl.acm.org/citation.cfm?id=345982*

Kleinberg, J. & Lawrence, S. (2001), 'Network analysis. The structure of the Web.', *Science (New York, N.Y.)* **294**(5548), 1849–50.

URL: *http://www.ncbi.nlm.nih.gov/pubmed/11729296*

Klemm, K. & Eguíluz, V. (2002), 'Highly clustered scale-free networks', *Physical Review E* **65**, 1–5.

Knox, H., Savage, M. & Harvey, P. (2005), 'Social networks and spatial relations: networks as method, metaphor and form', *Social Networks* **44**(1), 0–24.

Kumar, R. & Best, M. L. (2006), 'Impact and Sustainability of E-Government Services in Developing Countries: Lessons Learned from Tamil Nadu, India', *The Information Society* **22**(1), 1–12.

URL: *http://www.tandfonline.com/doi/abs/10.1080/01972240500388149*

Kunder, M. (2013), 'The Size of the World Wide Web'.

Kuzel, A. J., Like, R. C., Norton, P. G., Stewart, M., Tudiver, F., Bass, M. J. & Dunn, E. V. (1991), Standards of trustworthiness for qualitative studies in primary care, *in*

P. Norton, M. Stewart, F. Tudiver, M. Bass & E. Dunn, eds, 'Primary care research traditional and innovative approaches', number 10, Sage Publications, pp. 138–158.
URL: *http://www.loc.gov/catdir/enhancements/fy0655/90021954-t.html*

Kwak, H., Lee, C., Park, H. & Moon, S. (2010), 'What is Twitter, a social network or a news media?', *19th international conference on the World Wide Web* pp. 591–600.
URL: *http://dl.acm.org/citation.cfm?id=1772751*

Lafaille, R. & Wildeboer, H. (1995), *Validity and Reliability of Observation and Data Collection in Biographical Research*, Antwerp, Belgium.
URL: *http://eurotas.org/uploads/pdf/valideng_full.pdf*

Lankshear, C. & Knobel, M. (2008), Peter Lang Publishing Inc.

Lansdall-welfare, T. & Lansdall, T. (2012), 'Effects of the Recession on Public Mood in the UK', *Time* pp. 1221–1226.

Latour, B. (1987), *Science in Action*, Harvard University Press.
URL: *http://books.google.com/books?id=sC4bk4DZXTQC*

Latour, B. (1990), The Force and the Reason of Experiment, *in* H. LeGrand, ed., 'The Force and the Reason for Experiment', Kluwer, pp. 49–80.

Latour, B. (1991), Technology is society made durable, *in* J. Law, ed., 'A sociology of monsters essays on power technology and domination', Vol. 38 of *Sociological review monograph ;38*, Routledge, pp. 103–131.
URL: *http://books.google.com/books?hl=en&lr=&id=4Q54hSwGJvAC&pgis=1*

Latour, B. (1999), *Pandora's Hope: Essays on the Reality of Science Studies*, number 5032, Harvard University Press.
URL: *http://www.amazon.com/dp/067465336X*

Latour, B. (2005), *Reassembling the Social: An Introduction to Actor-Network-Theory by Bruno Latour*, Vol. 10, Oxford University Press.

Latour, B. (2010), Tardes idea of quantification, *in* M. Candea, ed., 'The Social After Gabriel Tarde: Debates and Assessments, Routledge', London, UK, pp. 146–162.

Latour, B. (2011), 'Networks, Societies, Spheres : Reflections of an Actor-Network Theorist', *Journal of Communication* **5**, 796–810.

Latour, B., Jensen, P., Venturini, T., Grauwin, S. & Boullier, D. (2011), 'The Whole is Always Smaller Than Its Parts - A Digital Test of Gabriel Tardes Monads', *British Journal of Sociology* pp. 1–21.

Law, J. (1987), Technology and Heterogeneous Engineering: The Case of Portuguese Expansion, *in* W. E. Bijker, T. P. Hughes & T. J. Pinch, eds, 'The Social Construction of Technological Systems New Directions in the Sociology and History of Technology', MIT Press, chapter 111-134, pp. 111–134.

Law, J. (1992), 'Notes on the Theory of the Actor Network : Ordering , Strategy and Heterogeneity', *Strategy* (1), 1–11.

Law, J. (2008), 'On sociology and STS', *The Sociological Review* **56**(4), 623–649.
URL: *http://doi.wiley.com/10.1111/j.1467-954X.2008.00808.x*

Law, J. & Singleton, V. (2005), 'Object lessons', *Organization* **12**(3), 1–17.
URL: *http://org.sagepub.com/content/12/3/331.short*

Lee, H. (2005), 'How Technology Shapes the Actor-Network of Convergence Services : A Case of Mobile Banking', *Information Systems* .

Lee, N. & Brown, S. (1994), 'Otherness and the Actor Network', *American Behavioural Scientist* **37**(6), 772–790.

Leskovec, J., Backstrom, L. & Kleinberg, J. (2009), 'Meme-tracking and the dynamics of the news cycle', *erg, J. Meme-tracking and the Dynamics of the News Cycle. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.* pp. 497–506.
URL: *http://dl.acm.org/citation.cfm?id=1557077*

Leskovec, J., Lang, K. J. & Mahoney, M. (2010), 'Empirical comparison of algorithms for network community detection', *Proceedings of the 19th international conference on World wide web - WWW '10* p. 631.
URL: *http://portal.acm.org/citation.cfm?doid=1772690.1772755*

Liu, G., Yu, Y., Han, J. & Xue, G. (2005), 'China web graph measurements and evolution', *Web Technologies Research and Development-APWeb 2005* (60473122), 668–679.
URL: *http://www.springerlink.com/index/rhhen2q0vkx7w9wv.pdf*

Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. (2011), 'Controllability of complex networks.', *Nature* **473**(7346), 167–73.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/21562557*

Luo, F., Wang, J. & Promislow, E. (2006), 'Exploring Local Community Structures in Large Networks', *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)* pp. 233–239.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4061371*

Luu, M.-d., Lim, E.-p., Hoang, T.-a., Chong, F. & Chua, T. (2011), 'Modeling Diffusion in Social Networks Using Network Properties', pp. 218–225.

MacKenzie, D. & Wajcman, J. (1985), Introduction: the Social Shaping of Technology, *in* MacKenzie, D & J. Wajcman, eds, 'The Social Shaping of Technology', Open University Press.

Macskassy, S. A. & Michelson, M. (2011), Why Do People Retweet? Anti-Homophily Wins the Day!, *in* 'Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media', pp. 209–216.

Mähring, M., Holmström, J., Keil, M. & Montealegre, R. (2004), 'Trojan actor-networks and swift translation: Bringing actor-network theory to IT project escalation studies', *Information Technology & People* **17**(2), 210–238.
URL: *http://www.emeraldinsight.com/10.1108/09593840410542510*

Manovich, L. (2011), 'Trending: The Promises and the Challenges of Big Social Data', *Debates in the Digital Humanities* pp. 1–17.

Mays, N. & Pope, C. (1995), 'Qualitative Research: Rigour and qualitative research', *British Medical Journal* **311**, 109–112.

McLean, P. (2007), *The art of the network : strategic interaction and patronage in Renaissance Florence*, Duke University Press.
URL: *http://www.worldcat.org/title/art-of-the-network-strategic-interaction-and-patronage-in-renaissance-florence/oclc/421491207&referer=brief_results*

Meeder, B., Karrer, B., Borgs, C., Ravi, R. & Chayes, J. (2011), 'We Know Who You Followed Last Summer: Inferring Social Link Creation Times In Twitter', *WWW '11 Proceedings of the 20th international conference on World wide Web* .

Merton, R. K. (1973), *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press, Chicago.

Metaxas, P. & Mustafaraj, E. (2013), 'The rise and the fall of a citizen reporter', *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13* pp. 248–257.
URL: *http://dl.acm.org/citation.cfm?doid=2464464.2464520*

Miettinen, R. (1998), 'Object Construction and Networks in Research Work: The Case of Research on Cullulose-Degrading Enzymes', *Social Studies of Science* **28**, 423–464.

Milgram, S. (1967), 'The small world problem', *Psychology Today* **2**(1), 60–67.

Mische, A. (2003), Cross-talk in movements: Reconceiving the culture-network link, *in* M. Diani & D. McAdam, eds, 'Social Movements and Networks Relational Approaches to Collective Action', Oxford University Press, pp. 258–80.

Mische, A. & White, C. H. (1998), 'Between Conversation and Situation: Public Switching Dynamics across Network Domains', *Social Research* **65**(3), 695–724.

Mislove, A., Marcon, M. & Gummadi, K. (2007), Measurement and analysis of online social networks, *in* 'Proceedings of the 7th ACM SIGCOMM conference on Internet measurement', pp. 29–42.
URL: *http://dl.acm.org/citation.cfm?id=1298311*

Mitchell, J. C. (1969), The Concept and Use of Social Networks, *in* J. C. Mitchell, ed., 'Social networks in urban situations', Manchester University Press, chapter 1, pp. 1–50.
URL: *http://books.google.com/books?hl=de&lr=&id=8RrpAAAAIAAJ&pgis=1*

Monteiro, E. & Hanseth, O. (1995), Social shaping of information infrastructure: on being specific about the technology, *in* W. Orlikowski, G. Walsham, J. M.R & D. J.I, eds, 'Information technology and changes in organisational work', Chapman & Hall, pp. 325–343.

Moreno, J. L. (1934), *Who shall survive?: A new approach to the problem of human interrelations*, Vol. 58 of *Nervous and mental disease monograph series*, Nervous and Mental Disease Publishing Co.
URL: *http://www.asgpp.org/docs/WSS/WSS.html*

Morgan, D. L. (1998), 'Practical strategies for combining qualitative and quantitative methods: applications to health research.', *Qualitative Health Research* **8**(3), 362–376.
URL: *http://qhr.sagepub.com/cgi/doi/10.1177/104973239800800307*

Morse, J. (1994), Designing funded qualitative research, *in* N. Denzin & Y. Lincoln, eds, 'In Handbook of Qualitative Research', pp. 220–235.

Mumford, L. (1967), *The Myth of the Machine (Vol. 1)*, Harcourt Brace Jovanovich, New York, New York, USA.

Murata, T. (2007), 'Discovery of User Communities Based on Terms of Web Log Data', *New Generation Computing* **25**, 293–303.

Murugesan, S. (2007), 'Understanding Web 2.0', *IT Professional* **9**(4), 34.

Newbery, D., Bently, L. & Pollock, R. (2008), 'Models of public sector information provision via trading funds', *Cambridge University* .
URL: *www.berr.gov.uk/files/file45136.pdf*

Newman, M. E. & Girvan, M. (2004), 'Finding and evaluating community structure in networks', *Physical Review E2* **69**.

O'Cathain, A., Murphy, E. & Nicholl, J. (2007), 'Integration and Publications as Indicators of "Yield" From Mixed Methods Studies', *Journal of Mixed Methods Research* **1**(2), 147–163.

Office of Public Sector Information (2007), The United Kingdom Implementation of the European Directive on the Re-use of Public Sector Information, Technical Report July, HM Government, London, UK.

O'Hara, K. & Hall, W. (2012), Web Science, *in* 'The Oxford Handbook of Internet Studies', Oxford Univ Press.

Orlikowski, W. J. (2000), 'Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations', *Organization Science* **11**(4), 404–428.
URL: *http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.11.4.404.14600*

Orlikowski, W. J. & Baroudi, J. J. (1991), 'Studying Information Technology in Organizations: Research Approaches and Assumptions', *Information Systems Research* **2**(1), 1–28.
URL: *http://isr.journal.informs.org/cgi/doi/10.1287/isre.2.1.1*

Ortiz, J. R., Zhou, H., Shay, D. K., Neuzil, K. M., Fowlkes, A. L. & Goss, C. H. (2011), 'Monitoring Influenza Activity in the United States: A Comparison of Traditional Surveillance Systems with Google Flu Trends', *PLoS ONE* **6**(4), 9.
URL: *http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3083406*

Page, L., Brin, S., Motwani, R. & Winograd, T. (1998), 'The PageRank Citation Ranking: Bringing Order to the Web', *World Wide Web Internet And Web Information Systems* **54**, 1–17.

Pahl, R. & Spencer, L. (2004), Capturing Social Communities, *in* C. Phillipson & G. Allan, eds, 'Social Networks and Socila Exlcusion: Sociological And Policy Perspectives'.

Pant, G. & Srinivasan, P. (2009), 'Predicting Web Page Status', *Information Systems Research* **21**(2), 345–364.
URL: *http://isr.journal.informs.org/cgi/doi/10.1287/isre.1080.0231*

Park, J. & Lee, W. (2011), Revolution 2.0 in Tunisia and Egypt: Reactions and sentiments in the online world, *in* 'Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media', Vol. 1, AAAI.

Paul, M. J. & Dredze, M. (2011), 'You Are What You Tweet: Analyzing Twitter for Public Health', *Artificial Intelligence* pp. 265–272.

Pear Analytics (2009), Twitter Study August 2009, Technical Report August, Pear Analytics.

Peay, E. R. (1980), 'Connectedness in a general model for valued networks', *Social Networks* **2**, 385–410.

Petricek, V., Cox, I. J. & Margetts, H. (2006), 'The Web Structure of E-Government - Developing a Methodology for Quantitative Evaluation'.

Pinch, T. & Bijker, W. (1984), 'The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other', *Social Studies of Science* **14**, 399–411.

Popper, K. (1963), 'Conjectures and Refutations: The Growth of Scientific Knowledge'.

Pothen, A., Simon, H. D. & Liou, K.-P. (1990), 'Partitioning Sparse Matrices with Eigenvectors of Graphs', *SIAM Journal on Matrix Analysis and Applications* **11**(3), 430–452.
URL: *http://hdl.handle.net/2060/19970011963*

Rahimi, B. (2011), 'The Agonistic Social Media: Cyberspace in the Formation of Dissent and Consolidation of State Power in Postelection Iran', *The Communication Review* **14**(3), 158–178.
URL: *http://www.tandfonline.com/doi/abs/10.1080/10714421.2011.597240*

Rao, D., Yarowsky, D., Shreevats, A. & Gupta, M. (2010), Classifying Latent User Attributes in Twitter, *in* 'Proceedings of the 2nd international workshop on Search and mining user-generated', ACM Press, New York, New York, USA.

Reichardt, C. S. & Rallis, S. F. (1994), 'Qualitative and quantitative inquiries are not incompatible: A call for a new partnership', *New Directions for Program Evaluation* **1994**(61), 85–91.
URL: *http://www3.interscience.wiley.com/journal/109751601/abstract*

Reid, A. J. (1996), 'What we want: qualitative research. Promising frontier for family medicine.', *Canadian Family Physician* **42**, 387–389, 397–400.

Riles, A. (2000), *The Network Inside Out*, University of Michigan Press.
URL: *http://www.amazon.com/Network-Inside-Out-Annelise-Riles/dp/0472088327*

Ritchie, J. & Lewis, J. (2003), *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, Vol. 134, SAGE.
URL: *http://books.google.co.uk/books?id=e6EO83ZKGYoC*

Ritter, R. S. & Hernandez, I. (2013), 'Happy Tweets: Christians Are Happier, More Socially Connected, and Less Analytical Than Atheists on Twitter', *Social Psychological and Personality Science* **June**.

Rossi, C. (2008), 'Web 2.0 and Music: the case of Myspace', *stsitalia.org* pp. 19–21.
URL: *http://www.stsitalia.org/conferences/papers2008/7_2_rossi.pdf*

Rossman, G. & Wilson, B. (1994), 'Numbers and words revisited: Being shamelessly eclectic', *Quality and Quantity* **28**(3), 315–327.

Russell, S. (1986), 'The Social Construction of Artefacts: a Response to Pinch and Bijker', *Social Studies Of Science* **16**(2), 331–346.
URL: *http://sss.sagepub.com/cgi/doi/10.1177/030631286016002008*

Sadilek, A. & Kautz, H. (2012), Modeling spread of disease from social interactions, *in* 'Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media', pp. 322–329.

Sale, J. E. M. & Brazil, K. (2002), 'Revisiting the Quantitative-Qualitative Debate : Implications for Mixed-Methods Research', *Community Health* pp. 43–53.

Sarker, S., Sarker, S. & Sidorova, A. (2006), 'Understanding Business Process Change Failure: An Actor-Network Perspective', *Journal of Management Information Systems* **23**(1), 51–86.

Savage, M. (2010), *Identities and Social Change in Britain since 1940: The Politics of Method*, Vol. 117, Oxford University Press.
URL: *http://books.google.com/books?hl=en&lr=&id=FdpGLicT3-EC&pgis=1*

Savage, M. & Burrows, R. (2007), 'The Coming Crisis of Empirical Sociology', *Sociology The Journal Of The British Sociological Association* **41**(5), 885–899.
URL: *http://dx.doi.org/10.1177/0038038507080443*

Savage, M. & Burrows, R. (2009), 'Some Further Reflections on the Coming Crisis of Empirical Sociology', *Sociology The Journal Of The British Sociological Association* **43**(4), 762–772.
URL: *http://dx.doi.org/10.1177/0038038509105420*

Schell, C. (1992), 'The Value of the Case Study as a Research Strategy', *Manchester Business School* .
URL: *http://www.finance-mba.com/Case Method.pdf*

Scott, J. (2000), *Social Network Analysis: A Handbook*, Vol. 3, Sage Publications.
URL: *http://www.amazon.com/dp/0761963391*

Scott, J. (2008), Network Analysis, *in* A. Darity, William, ed., 'International Encyclopaedia of the Social Sciences', 2nd editio edn, Gale Virtual Reference Library, Detroit, pp. 475–477.

Scott, J. (2010), 'Social network analysis: developments, advances, and prospects', *Social Network Analysis and Mining* **1**(1), 21–26.
URL: *http://www.springerlink.com/index/10.1007/s13278-010-0012-6*

Secker, J., Wimbush, E., Watson, J. & Milburn, K. (1995), 'Qualitative methods in health promotion research: some criteria for quality', *Health Education Journal* **54**, 74–87.

Shadbolt, N. (2009), Government Linked Data: A Tipping Point for the Semantic Web, number December.

Shadbolt, N. (2010*a*), Open Government Data: A Case Study In Web Science, Technical Report July, University of Southampton; HM Government.

Shadbolt, N. (2010*b*), 'Transparency and Open Data: Next Steps', *Letter to Rt Hon Francis Maude MP dated 14 June 2010* .

Simmons, M., Adamic, L. A. & Adar, E. (2011), 'Memes Online: Extracted, Subtracted, Injected, and Recollected', *ICWSM 2011* .
URL: *http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/*

Sismondo, S. (2010*a*), Actor-Network Theory, *in* 'Introduction to Science and Technology Studies', 2nd edn, Vol. 10, Wiley-Blackwell, pp. 94–105.

Sismondo, S. (2010*b*), *An Introduction to Science and Technology Studies*, 1st editio edn, Wiley-Blackwell.

Slattery, S. (2009), edit this page: the socio-technological infrastructure of a wikipedia article, *in* 'Proceedings of the 27th ACM international conference on Design of communication', pp. 289–295.
URL: *http://dl.acm.org/citation.cfm?id=1622052*

Smith, J. K. (1983), 'Quantitative versus qualitative research: An attempt to clarify the issue', *Educational Researcher* **12**(3), 6–13.
URL: *http://links.jstor.org/sici?sici=0013-189X(198303)12:3¡6:QVQRAA¿2.0.CO;2-L*

Smith, J. K. & Heshusius, L. (1986), 'Closing Down the Conversation: The End of the Quantitative-Qualitative Debate Among Educational Inquirers', *Educational Researcher* **15**(1), 4–12.
URL: *http://edr.sagepub.com/cgi/doi/10.3102/0013189X015001004*

Somboonviwat, K., Suzuki, S. & Kitsuregawa, M. (2008), 'Connectivity of the Thai web graph', *Progress in WWW Research and Development* pp. 613–624.
URL: *http://www.springerlink.com/index/f1143221406066p2.pdf*

Steinhaeuser, K. & Chawla, N. (2008), Community detection in a large real-world social network, *in* 'Social Computing, Behavioral Modeling, and Prediction', Springer US, pp. 168–175.
URL: *http://link.springer.com/chapter/10.1007/978-0-387-77672-9_19*

Strauss, A. L. & Corbin, J. (1990), *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage.

Tashakkori, A. & Teddlie, C. (1998), *Mixed Methodology: Combining Qualitative and Quantitative Approaches (Applied Social Research Methods)*, Sage Publications.

Tatnall, A. (1999), 'Actor-Network Theory and Information Systems Research', *Innovation* pp. 955–966.

Tatnall, A. & Burgess, S. (2004), Using Actor-Network Theory to Identify Factors Affecting the Adoption of E- Commerce in SMEs, *in* 'E-Business Innovation and Change Management', pp. 152–169.

Tatnall, A. D. (2010), 'Using actor-network theory to understand the process of information systems curriculum innovation', *Education and Information Technologies* **15**(4), 239–254.
URL: *http://www.springerlink.com/index/10.1007/s10639-010-9137-5*

Taylor, P. (1995), 'Co-Construction and Process : A Response to Sismondo's Classification of Constructivisms', *Social Studies Of Science* **25**(2), 348–359.

Tellis, W. (1997), 'Introduction to Case Study', *The Qualitative Report* **3**(2), 1–11.
URL: *http://www.nova.edu/ssss/QR/QR3-2/tellis1.html*

Thomas, G. (2011), 'A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure', *Qualitative Inquiry* **17**, 511–521.

Tinati, R. (2012), 'UK Open Government Data Community Timeline'.
URL: *http://users.ecs.soton.ac.uk/rt506/UKOGDTimeline*

Tinati, R., Carr, L. & Hall, W. (2012), 'Identifying communicator roles in twitter', *Proceedings of the 21st international conference companion on World Wide Web* pp. 1161–1168.
URL: *http://dl.acm.org/citation.cfm?id=2188256*

Tinati, R., Halford, S., Carr, L. & Pope, C. (2012), Mixing Methods and Theory to Explore Web Activity, *in* 'Proceedings of the 3rd Annual ACM Web Science Conference (WebSci '12)', ACM, New York, NY, USA.

Tinati, R., Halford, S., Carr, L. & Pope, C. (2013), 'Big Data: Methodological Challenges and Approaches for Sociological Analysis', *Sociology* .

Trist, E. L. & Bamforth, K. W. (1951), 'Some social and psychological consequences of the Longwall method of coal-getting', *Human Relations* **4**(1), 3–38.
URL: *http://doi.apa.org/psycinfo/1951-08304-001*

Trotter, R. T. (2007), Friends, relatives and relevant others: Conducting ethnographic network studies, *in* J. J. Schensul, M. D. LeCompte, R. T. Trotter, E. K. Cromley & M. Singer, eds, 'Mapping social networks spatial data hidden populations', Alta Mira.

Twitter (2013), 'The Twitter Glossary'.
URL: *https://support.twitter.com/articles/166337-the-twitter-glossary*

Tyler, J. R. & Wilkinson, Dennis M. Huberman, B. A. (2003), Email as spectroscopy: Automated discovery of community structure within organizations, *in* 'Proceedings of the First International Conference on Communities and Technologies (ICCT03)', Dordrecht, Netherlands.

Vazquez, A. (2002), 'Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations', *Physical Review E* **67**, 17.

Velasquez, A. (2012), 'Social media and online political discussion: The effect of cues and informational cascades on participation in online political communities', *New Media & Society* **14**(4), 1–18.

Wakita, K. & Tsurumi, T. (2007), Finding Community Structure in Mega-scale Social Networks [Extended Abstract], *in* 'Proceedings of the 16th international conference on World Wide Web', pp. 1275–1276.

Walsham, G. (1995), 'Interpretive case studies in IS research: nature and method', *European Journal of Information Systems* **4**(2), 74–81.
URL: *http://www.palgrave-journals.com/doifinder/10.1057/ejis.1995.9*

Wang, X. W. X. & Reiter, M. K. (2010), 'Using Web-Referral Architectures to Mitigate Denial-of-Service Threats'.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4641937*

Ward, M. D., Ahlquist, J. S. & Rozenas, A. (2013), 'Gravity's Rainbow: A dynamic latent space model for the world trade network', *Network Science* **1**(01), 95–118.
URL: *http://www.journals.cambridge.org/abstract_S2050124213000015*

Wasserman, S. & Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Vol. 8 of *Structural analysis in the social sciences, 8*, Cambridge University Press.
URL: *http://www.amazon.com/dp/0521387078*

Watts, D. J. (1999), *Small Worlds*, Vol. 19, Princeton University Press.
URL: *http://doi.wiley.com/10.1002/rsa.1018*

Watts, D. J. & Strogatz, S. H. (1998), 'Collective dynamics of 'small-world' networks.', *Nature* **393**(6684), 440–2.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/9623998*

Weinberger, D. (2012), *Too Big to Know*, Basic Books.

Weinreich, H., Obendorf, H., Herder, E. & Mayer, M. (2008), 'Not quite the average', *ACM Transactions on the Web* **2**(1), 1–31.
URL: *http://portal.acm.org/citation.cfm?doid=1326561.1326566*

Weng, L., Flammini, a., Vespignani, a. & Menczer, F. (2012), 'Competition among memes in a world with limited attention', *Scientific Reports* **2**, 1–8.
URL: *http://www.nature.com/doifinder/10.1038/srep00335*

Whittle, a. & Spicer, a. (2008), 'Is Actor Network Theory Critique?', *Organization Studies* **29**(4), 611–629.
URL: *http://oss.sagepub.com/cgi/doi/10.1177/0170840607082223*

Whitworth, B. & Liu, T. (2009), 'Channel E-mail: A Sociotechnical Response to Spam', *Computer* (July), 63–72.
URL: *http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5165529*

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P. & Zhao, B. Y. (2009), 'User interactions in social networks and their implications', *Proceedings of the fourth ACM european conference on Computer systems - EuroSys '09* p. 205.
URL: *http://portal.acm.org/citation.cfm?doid=1519065.1519089*

Winner, L. (1993), 'Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology', *Science Technology And Human Values* **18**(3), 362–378.
URL: *http://www.jstor.org/stable/689726*

Woolgar, S. (1991), 'The Turn to Technology in Social Studies of Science', *Science, Technology, & Human Values* **16**, 20–50.

Woolgar, S. & Latour, B. (1979), *Laboratory Life : The Construction of Scientific Facts Laboratory Life*.

World Wide Web Consortium (W3C) (2009), 'What is the Difference between the Web and the Internet'.
URL: *http://www.w3.org/Help/*

Yardi, S., Romero, D., Schoenebeck, G. & Boyd, D. (2010), 'Detecting spam in a twitter network', *First Monday* **15**(1).

Yin, R. K. (2003), *Case Study Research: Design and Methods - Third Edition*, Vol. 5 of *Applied Social Research Methods Series*, SAGE Publications.

Yu, H. & Robinson, D. G. (2012), 'The New Ambiguity of Open Government ', *Transit* (September 2011), 1–24.

Zhan, A. (2010), 'Understanding the Standardization Strategies in China from the Perspective of Actor-network Theory', *Intellectual Property* (2006).

Zhou, Z., Bandari, R. & Kong, J. (2010), 'Information resonance on Twitter: watching Iran', *Proceedings of the First Workshop on Social Media Analytics (SOMA '10)* pp. 123–131.
URL: *http://dl.acm.org/citation.cfm?id=1964875*

Zhu, J., Meng, T., Xie, Z., Li, G. & Li, X. (2008), 'A teapot graph and its hierarchical structure of the chinese web', *... conference on World Wide Web* p. 1133.
URL: *http://portal.acm.org/citation.cfm?doid=1367497.1367692 http://dl.acm.org/citation.cfm?id=1367692*

Zimmermann, B. (2008), Uncovering Cultural Issues in the Internet of Things: A Design Method, *in* 'Internet of Things'.
URL: *http://archive-ouverte.unige.ch/*