Mixing Methods and Theory to Explore Web Activity

Ramine Tinati
University of Southampton, UK
rt506@ecs.soton.ac.uk

Susan Halford
University of Southampton, UK
susan.halford@soton.ac.uk

Leslie Carr
University of Southampton, UK
lac@ecs.soton.ac.uk

Catherine Pope
University of Southampton, UK
c.j.pope@soton.ac.uk

ABSTRACT
Web Science is now well recognized as an interdisciplinary field, drawing on research from the computational, natural and social sciences. These disciplines bring diverse theoretical and methodological approaches, providing alternative perspectives and insight into Web activity. Consequently, Web Science faces the challenge of developing research methods that transcend disciplines, not least in dealing with the epistemological tensions between different methodological approaches. As a start, this paper argues that, a mixed methods approach is required. To demonstrate the affordances of this, the activities of the UK Open Government Data community are analyzed by combining quantitative computational science techniques with qualitative social science methods underpinned by social theory. This provides a richer and more detailed analysis than either approach alone could offer and one which enables us to apprehend the Web as a complex socio-technical phenomenon.

Author Keywords
Web Science, Methodology, Open Government Data

ACM Classification Keywords
H.1.1 Systems and Information Theory

General Terms
Design, Human Factors, Theory

INTRODUCTION
The Web is a large-scale and diverse socio-technical phenomenon driven by technical architectures, government policies, business economics and the social interactions of billions of people in everyday life, In their clarion call for Web Science, Berners-Lee *et al.* [5] insisted that robust and useful understanding of the Web would depend on multidisciplinary research, drawing together theory and methods from the social sciences, humanities, and natural sciences.

Responding to this call, *The Manifesto for Web Science* [23] outlined some of the theoretical and philosophical questions that might be involved in building collaborative research about the Web. Halford *et al.* [23] proposed five principles for this: (1) research should be more than the sum of its disciplinary parts, aiming for interdisciplinarity not a smorgasbord of disciplinary perspectives on the Web (2) research must attend to the co-constitutive nature of humans and technologies in the emergent Web (3) we must follow the actors – individuals, collectivities and technologies - involved in the Web, appreciating the extensive nature of the networks (beyond 'the Web itself') involved in producing and reproducing the Web (4) a range of epistemologies and methodologies must be harnessed to understand the Web at both micro and macro level, (5) Web Science should be a critical discipline, addressing moral, political and ethical questions about the growth and direction of Web development.

This drive towards a theoretically grounded, multi-disciplinary Web Science raises a methodological challenge: what methods will enable us to research the Web as a socio-technical phenomenon? The principles outlined by Halford *et al*. [23] suggest that we should transcend the methodological divisions inherent in the current disciplinary settlement across the human, social and natural sciences. Although we would be the first to recognize (and welcome) exceptional cases, these divisions can be characterized in broad brush-strokes as a continuum from positivist epistemology and quantitative method, commonly found in the natural sciences to the interpretivist and qualitative approaches associated more strongly with many of the social sciences. This paper explores the opportunities and challenges involved bringing together approaches from along this continuum to build Web Science methodologies. We begin, below, with a brief exploration of some of the tensions involved in doing this by comparing quantitative modeling approaches to the Web, in Computer Science, with the qualitative methods that dominate Web research in Sociology. Following this, we draw these different methods together to explore the Open Government Data (OGD) movement, currently at the forefront of Web development, driven by a mix of social and technological agendas. We argue that a mixed methods approach, generates more robust and useful insights than individual methodologies and that the persistent epistemological tensions may be used, in a positive way, as 'irritants' to ensure that we maintain a critical perspective on what we can know, and how we can know it.

TWO METHODOLOGICAL APPROACHES
How can we research the Web as a complex socio-technical phenomenon? We begin with a consideration of two very different research paradigms, positivism and interpretivism/constructivism which are often respectively characterized by either quantitative or qualitative approaches to research [40]. Different methodological approaches are often linked with particular disciplines: thus – for instance - studies in computer science tend to use quantitative data, underpinned by positivist assumptions (whether acknowledged or not) to model findings, whereas sociology makes more use of qualitative research drawing on interpretive/constructivist perspectives to unravel meaning and understanding and to provide explanation of the processes producing particular phenomena.

Research investigating the same phenomena from a quantitative and qualitative approach appear to be studying the same thing, but may have very different underpinning beliefs and constructs. To draw out the distinctive nature of the different approaches we

WebSci 2012, June 22–24, 2012, Evanston, Illinois, USA.

explore three questions: (1) what types of research questions are asked? (2) What types of methods are being used? (3) What types of knowledge are produced?

The quantitative paradigm is predicated on empirical research. The positivist ontological position is that there is an objective truth independent of human thought and action. The goal of science in this framework is to measure and analyze causal relationships between variables within a value-free framework [14]. Quantitative studies use large sample sizes necessary to statistical methods used [10] and make predictive, generalizable statements [22]. Alternatively, qualitative methods are commonly informed by interpretivism [2] and constructivism [22], and by perspectives which hold that there are multiple and dynamic realities and truths which arise from both the researcher's and research participants' construction and understanding of reality [5][43]. The emphasis of qualitative research is on process and meaning, and particularly in UK sociology, makes extensive use of surveys, interviews and, observational methods. Samples in qualitative research are not meant to provide results that can be generalized to other or larger populations, instead they are typically purposively chosen to reveal in detail the behaviors and understandings of the groups studied [39].

Quantitative Web research often borrows techniques and concepts from graph theory and network sciences [8]. These methods rely on mathematical proofs and statistically based analytical methods to describe the Web. Studies conducted within the computational sciences explore various aspects of the Web; examining the dynamics of networks [32] and the modeling of the Web graph [42], aimed at understanding the Web's growth and how it evolves over time. There has been an increase in research investigating adaptive networks [21], [35], which examine the Web as a coevolving network of behavior and network structures, using quantitative data and technical analysis to provide an understanding of a network's growth. There also exists research which is concerned with the structure of Web communities based upon the changes in topology [27] [29], how information cascades through traditional social networks [36] and micro-blogging social networks [26], and also how the spread of innovation occurs [37].

The quantitative approach has become increasing popular with the rise of social network analysis (SNA) [16]. These studies are interested in how messages and information are transmitted [26] [33] [13] [34], and in examining the structure and dynamic properties of the networks. SNA also include the study of message propagation within social networking systems such as Twitter and Facebook, replicating node-and-edge networks based on relationships or information passed between users [27]. Supporting this area of research, studies have explored computer mediated approaches to user modeling [38] [1] and the influence of users on message propagation [12] [48] [3]. These studies aim to examine and model individual behavioral characteristics within social networks, to predict future network developments. Consequently, these studies offer mathematical reasoning for the patterns and structures observed, but may be criticized for content stripping and the exclusion of meaning [22], potentially overlooking the social nature of the findings. Although some studies attempt to provide context for their observations [7], they continue to be grounded by positivism, emphasising the verification of an *a priori* hypothesis. This research may provide positivist methodological rigor, but it risks glossing over other possible findings [22].

In contrast, qualitative approaches to researching the Web are driven by different types of research questions which, although related, are distinguished by alternative theoretical and epistemological roots. Especially in the British context, Sociology makes extensive use of qualitative methods, and offers a variety of theoretical approaches. The philosophical positions that underpin qualitative methods result in a different epistemological and ontological understanding compared to the quantitative approach. The type of questions being asked provides a more nuanced and often theoretical understanding of the Web. Examples include: how Web 2.0 in shapes communities and identities [15], the development and impact of Web Services [18], the practices and integration of online banking [4], or the role that Domain Name System has played on shaping the development of the Web [49]. These studies are often performed at a small scale, using qualitative data collection techniques including, interviews observations and focus groups, to develop an understanding of the context, practices and effects of socio-technical networks. These studies may also draw on qualitative methods to position or support their findings, and are often also linked to social theories [46] [28] [44], which provide alternative interpretations of the findings (or, to put it another way alternative claims to 'truth'). Science and Technology Studies (STS) provide the theoretical groundwork for a number of these studies, offering an approach that we might use to analyze the Web as a socio-technical network. Whilst few studies of the Web to date have drawn on this framework, the perspectives developed by Callon [9] and Law [31] in other contexts demonstrate how a qualitative account of historic events that gives equal weight to social and technical actors can be used to understand emergent outcomes or, in this case, the evolution of the Web.

In comparison to a quantitative approach, the qualitative approaches provide a smaller-scale yet often richer understanding of the phenomena, taking into consideration broader social, political and economic factors and influences. Typically these Qualitative methods require a text based, and therefore labor intensive process of data collection and analysis, which limits the size of the studies. Whilst this means that findings are not statistically generalizable to larger populations [39], these methods yield analytical and conceptual findings that may be transferred to other settings. As qualitative studies often require higher levels of researcher engagement with data collection (face to face interviews, participant observation) and seek to build interpretation they are often regarded as less objective than quantitative approaches; potentially allowing multiple perspectives and interpretations of the findings, and as a result have been open to debate and criticism [40].

Within these two research paradigms, the research questions asked and the knowledge produced are the product of the underpinning theoretical, ontological and epistemological assumptions. Both present different ways to examine phenomena, equally important in their own right; "there are no bad methods, just bad research" [22].

To further explore these two different approaches and how using them together enables a complementary and robust methodology for analyzing and understanding the Web, we will now shift our focus towards the analysis of a highly active Web community. By using methods drawn from both approaches described above we aim to demonstrate how the underlying epistemological and ontological assumptions drive the studies research questions and knowledge produced and demonstrates how a mixed methods

approach can provide an analysis that draws upon the benefits of both quantitative and qualitative research.

UNDERSTANDING NETWORKS ON THE WEB – EXAMINING OGD USING MIXED METHODS

The case for mixed methods is not new [40], and has already been applied to social network analysis [16], research where quantitative methods have been used to redress, small scale and unrepresentative qualitative findings [40] and, conversely, qualitative research has been employed to explore social processes and meanings only indicated in descriptive and inferential statistics. Both quantitative and qualitative studies tracing the dramatic rise of social networking on the Web have been conducted in computer science and sociology [19], exploring phenomena involving the dramatic rise in social networking on the Web and in society [16]. Computer Science studies tend to explore the properties and structure of the communications network, whereas sociological studies, focus on why and how different people do (or don't) use tools like Twitter, and the social implications of this activity. However what is not clear is the use of a mixed methods approach towards understanding the socio-technical properties of Web activity.

We will now demonstrate how mixing quantitative and qualitative methods enables us to explore the growth of the Web more fully than either methodology could offer on its own. The UK Open Government Data (OGD) community is used as an exploratory case study, providing a good example of current activities shaping the evolution of the Web. Following Halford *et al*. [23] we trace socio-technical interactions on-line and off-line, at a micro and a macro scale in order to develop a comprehensive understanding of how and why the OGD movement is emerging.

Data Sources

In order to access online and offline activities around OGD, at both macro and micro we access a range of data sources, using mixed methods. (1) We explore the on-line community promoting OGD by analyzing participants' use of Twitter and, in particular the networks that operate to link individuals and groups. Twitter combines elements of social networking and blogging, providing each user with their own profile and timeline which contains messages that they and those that the individuals/groups that they choose to 'follow'. Based on these 'following' relationships, networks can be constructed, where nodes represent the users, and directed edges are the friendship links between them. Twitter users can also send messages containing an explicit link to other user(s), using '@' followed by a user's name (i.e. @johndoe), offering an alternative network graph, of users (nodes) and tweets (directed edges). These types of relationships and messages are shown in Figure 1. Finally, Twitter allows users to copy and 'retweet' other people's tweets to their own timeline, a process known as 'retweeting'. The retweet present in the retweeter's timeline shows the original author that created the tweet, thus displaying the original author to a wider audience, potentially. As Figure 2 demonstrates, the process creates a network of users (nodes) and retweets between users (directed edges).

For this paper, a dataset of tweets relating to the #datagovuk hashtag was collected; containing 3853 tweets from 2209 unique users, during January 2010 till December 2011. The #datagovuk hashtag is used by individuals who are tweeting about the activities of UK Open Government Data and also data.gov.uk – the UK's central Web portal and repository for published government data. Each record in the dataset corresponds to a unique tweet using the #datagovuk hashtag, and contains

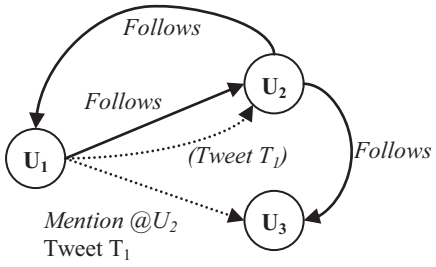

Figure 1. 'Following' and 'Mention' Network Structure

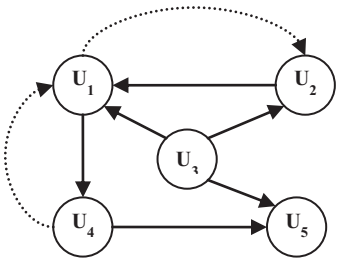

Figure 2. Twitter Retweet Functionality

information regarding the unique identifier of the tweet and the user who made the tweet, a timestamp of tweet creation time, and the tweet text. The records are stored in chronological order which aids parsing the dataset when examining the dynamic properties of the communications.

(2) A second source of evidence on the activities of the OGD community is derived from the dataset deposit records in data.gov.uk (Records were harvested via the data.gov.uk CKAN API) have also been collected; and analyzed using repository techniques [11]. The collection which spans from June 2009 to December 2011 contains 7407 records, grouped by daily intervals. Each record contains information regarding its deposit date, the dataset owner, and government department that it belongs to.

(3) Finally, original qualitative data has been collected by conducting 15 semi-structured interviews with key stakeholders within the Open Government Data community. Participants had varying roles within the community, and included civil servants, developers, activists, and other interested parties. The interviews (which have been anonymised) explored the growth of the UK OGD community, concerning issues such as: the role of technologies and its importance within the community, the drivers and barriers to the adoption of OGD, changes in the community, and the future of UK OGD. Also, interviewees were asked to identify other key actors known to be influential and well regarded in the OGD community. Documentary analysis was also conducted, examining a number of Web resources to help construct a timeline [45] of events, meetings, and interactions of the UK OGD community in order to provide a chronological account of the activities that occurred.

Examining the Activities of the UK Open Government Data Community

Our exploration of the UK OGD community begins with an examination of the quantitative data collected. The Twitter data provides a way to see how individuals tweeting about #datagovuk are communicating with each other which allows us to understand the interactions that both drive and reflect the growth of the community. Using the data collected, two network graphs can be

constructed: a *mention* network graph, which represents the tweets sent between different users, and the *retweet* network graph, which represents the sharing of messages between each other. Both provide a different representation of the activities within the #datagovuk conversation timeline, revealing important structural and dynamic properties that will be then used in combination with the qualitative analysis to provide a more informed understanding of the UK OGD community's growth and functioning.

Figure 3 and 4 illustrate the *mention* and *retweet* network graphs, respectively; the two graphs represent the harvested Twitter #datagovuk tweets, mapping the users (the nodes) and the tweets (directed edges) between each other. The layout of the graphs were produced using the GEM layout algorithm [20]. A BIN pack was then applied to the graph, which provides a tidier view arrangement of the network. The inclusion of Figure 3 and 4 not only serves as a visual aid to the analysis to follow, but also to demonstrate the structural differences that underpin the various communications within a social networking service such as Twitter.

Table 1. #datagovuk Mention and Retweet Graph Metrics

Graph Metrics	Mention	Retweet
Max. in-Degree	19	79
Mean. in-Degree	1	1.3
Max out-Degree	14	17
Mean out-Degree	1	1.3
Strongly Connected Components	159	119
Weakly Connected Components	603	953
Communities	167	138
Modularity	0.91	0.83

As a way into the mixed methods approach taken within this research, the first step will apply a number of SNA and graph theory metrics to the mention and retweet network. Table 1 lists various graph metrics which provide a measure of the network structural properties that are exhibited by the mention and retweet networks.

We begin by examining the degree metric of the networks, which is a representation of the number of edges a user has made to other users. The in-degree represents different structures for the mention and retweet network; the former being simply the number of tweets that a user has had directed to them, the latter being the number of their tweets that have been retweeted. Alternatively, the out-degree represents the number of edges that a user had made. In regards to the mention network, is a representation of the number of mentions a user has made, and in the retweet network, how many retweets a user has made. Considering the in-degree metric, both the retweet and mention network have a fairly low average degree, an indication that communications and sharing of tweets amongst users is limited. However, with regards to the maximum in-degree, the retweet network in comparison to the mention network has a much higher value, indicating that there are users that have a large number of their tweets retweeted, a possible indicator that the tweet contains valuable content [7]. Examining the various out-degree metrics for the mention and retweet network tells the same story, the mean out-degree are both similar, and so are the maximum out-degrees, which could be considered a representation of the overall connectivity within the network, i.e. are users operating in siloed networks of communications rather than conversing to a wide set of individuals? This is an interesting question, which will be explored in more detail later on within the qualitative analysis.

Examining the structure of the networks further, we can consider the number of *weakly* and *strongly* connected components, and also the number of *communities* as a result of the communications between users provides further evidence towards understanding the network structures. A network is strongly connected if there is a path from a node and every other node, in comparison to this, a weakly connected network is one where there are not paths to from one node to all other nodes. Examining the strongly and weakly connected component metrics in Table 1 show that although the number of strongly connected components are similar for the mention and retweet networks, the number of weakly connected components in the retweet network is much higher. Putting this into context, this suggests that there exists a large number of one-way retweeting of content between users, whereas within the mention network there are more two-way conversation's occurring. This again is another finding that will be addressed within the qualitative analysis, offering some explanation as to why there are one-way steams of communication.

Figure 3. #datagovuk Mention Network Graph

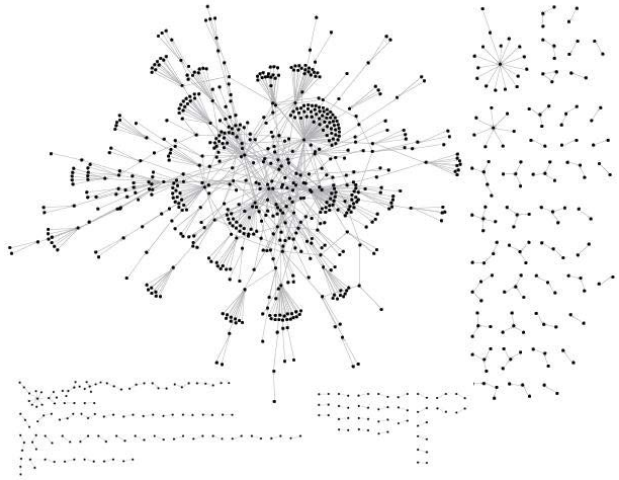

Figure 4. #datagovuk Retweet Network Graph

Further exploring the connectivity and structure of the networks, we will now consider the communities and modularity metric, the former a metric to determine the number of sub-communities within a network, the latter a metric to measure the strength of division of the network into the communities. The values for these metrics given for the mention and retweet network are reflected in the visual representations shown in Figure 3 and 4, the higher level of modularity in the mention network indicates a sparsely connected network of communities in comparison to the more densely connected (between communities) retweet network. Furthermore, the larger number of communities within the mention network is also apparent in Figure 3, strengthening the earlier analytical finding that there exist siloed networks of communications.

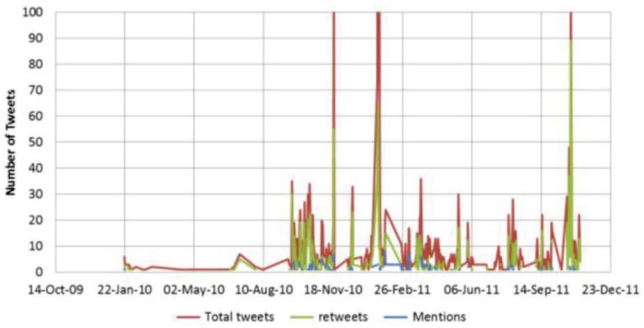

Figure 5. #datagovuk Twitter Activity Graph

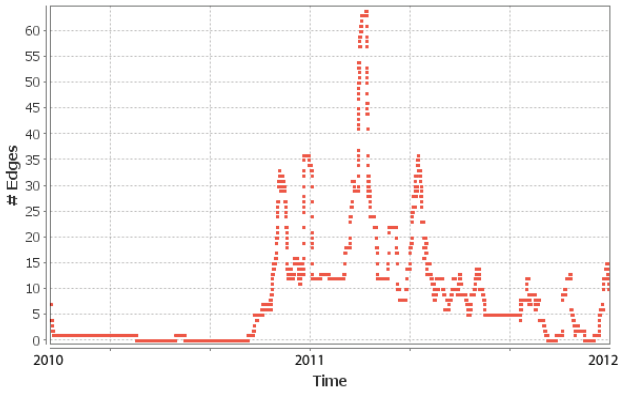

Figure 6. #datagovuk Mention Dynamic Number of Edges

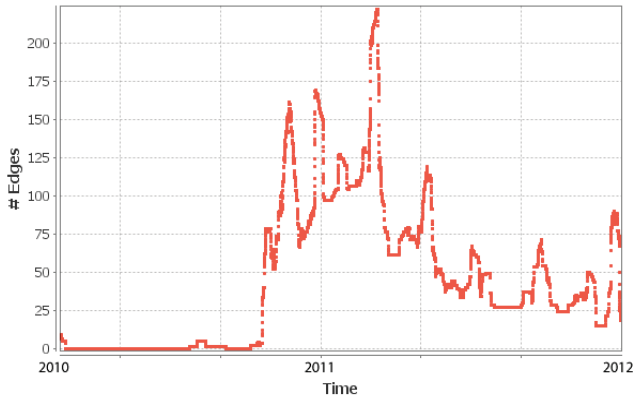

Figure 7. #datagovuk Retweet Dynamic Number of Edges

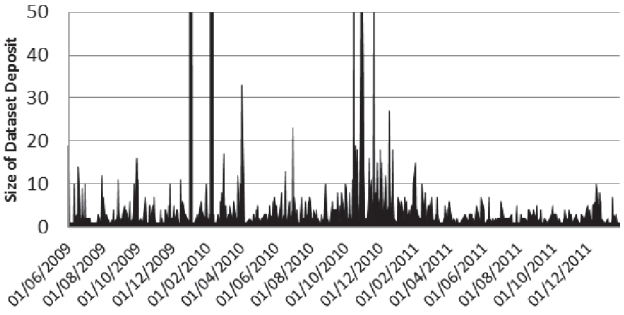

Figure 8. Dataset Deposit Activity – data.gov.uk

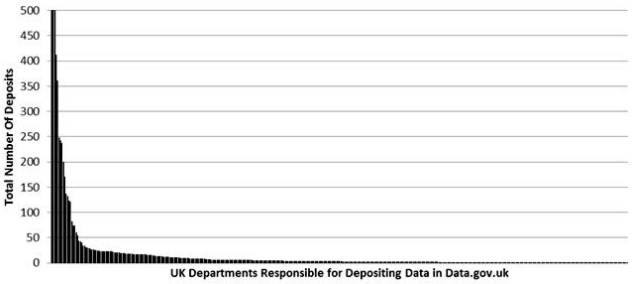

Figure 9 Distributions of Deposits between Departments

Table 2. Top 5 Retweeted Users (Names Anonymised)

Rank	Screenname	Tweets	Retweets
1	U********	11	96
2	N********	10	53
3	T********	1	50
4	D********	21	49
5	H********	7	34

Examining the temporal activities of the #datagovuk tweets shown in Figure 5, the activity levels recorded over the period of January 2010 to October 2011 show that there is continuous levels of tweet activity and at certain points, extremely high number of tweets and retweets are occurring. This again prompts some interesting questions regarding the overall activities of the Twitter network; do certain spikes of activity correspond to important changes within the development of the community? The graphs shown in Figure 3 and 4 represents only a snapshot of the interactions between users; visualizing the activities based over a time period provides a way to examine these activities at multiple snapshots - some which represent little of no activity of the network at all. This can be pushed further, 'stitching together' analytical snapshots of the network over the timeline of collected tweets provides a dynamic perspective of the evolution of the network in regards to the number of users participating and the fluctuating number of communications. This effectively provides another dimension to SNA, which as Scott suggests [41] is an area which may reveal great analytical insight. Supporting Figure 5, the dynamic analysis of the number of edges over time shown in Figure 6 and 7 provide an effective method to 'slice' along the data, showing that the communications were not as constant as the static analysis depicts; instead there is a large amount of variance in communication frequency in both the mention and retweet network. Again, the qualitative research will complement these analytical points raised, unpacking what the different levels of

activities represent, and what effect they have had on the UK OGD community.

Shifting focus from examining the network at a macro level towards a finer micro-level granularity; we can push the analysis towards examining the actual users within the #datagovuk network. By analyzing the individuals who are participating in the #datagovuk communications, they can be sorted by number of tweets, mentions and retweets. Building upon existing research [12,47] retweet can be used as a metric to identify important and influence users within a Twitter network. Table 2 shows the top five users to be retweeted, and reveals that the number of retweets that a user receives does not necessarily depend on the number of tweets made, as user 'T********' tweet to retweet ratio shows. Reexamining the network graph in Figure 4, the nodes that are highly connected (the large hubs) are potentially the actors identified within Table 2.

Our quantitative analysis also draws upon the deposit activities within the data.gov.uk data catalogue. This provides a good way of measuring the 'health' of the repository, providing insights into how it is being ran [11]. Figure 8 represents the datasets grouped by date deposited in the data.gov.uk data catalogue. The large spike in deposit activity at the end of 2009 (just before data.gov.uk was publically launched) followed by the decrease in activity is a common trait of repository activity, usually indicating the use of batch depositing, rather than the steady release of data. However, Figure 8 shows continuous, frequent deposits, an indicator of a healthy repository. Interestingly, the spike in dataset deposits at November 2010 corresponds to a rise in #datagovuk Twitter activity at the same time as shown in Figure 5; potentially as the result of the publication of important or controversial datasets, which spurred on discussion on Twitter. However, the activity levels during Q3 and Q4 of 2011 have decreased; an indicator that the 'health' of the repository needs examining. The datasets can also be examined based on the additional metadata collected; the datasets – according to the metadata provided – has been deposited by 406 different departments. However, as Figure 9 illustrates, the distribution of the deposits is not evenly spread across the departments; in fact only 16 out of the 406 departments were responsible for publishing a substantial proportion (>1%) of the total datasets available.

The quantitative data collected has provided a way to observe the activities of the UK OGD community via graph metrics and other statistical measures. The SNA has shown that there exists a loosely connected network of actors, where some are more influential than others. The analysis will now call upon qualitative data and methods in order to build upon and draw out further issues that the quantitative analysis has already found. At this point, we shift from a positivist perspective, towards an interpretive understanding of the observed activities. By contrast to the quantitative approach, the qualitative analysis will employ a sociological socio-technical perspective on the network of humans and technological function and interact, placing interested on studying the 'lived experiences' [17] in the identified social networks, and the consequences that these interactions have on society [24]. Building upon the quantitative SNA, the qualitative analysis draws upon sociological literature which not only is concerned with socio-technical phenomena, but also social network literature that extends early anthropological studies concerned with research on communities, friendships, kinships and neighborhoods; typically relying on in-depth interviews, observations and narratives to provide a narrative of the implications in a wider context[16].

The interviews conducted in combination with a timeline of events [45] help reconstruct an account of the UK OGD community activities. Interviewees, ranging from various roles within the OGD community, were identified and selected by their profile in many of the timeline events and also their centrality within the Twitter #datagovuk network. We use social theory to frame this analysis, in order to examine the complex socio-technical relationships in play [30].

Recapping the initial quantitative analysis findings, (1) there are different structures between the mention and retweet networks, and within this, (2) the dynamic analysis of the retweets has revealed that the network is always in a fluctuating state and stability cannot be assumed or predicted. This is also reflected in the data.gov.uk data catalogue, with a growing but non-consistent frequency and size of dataset deposits (3) Finally, there exist a number of Twitter users who have a large number of their tweets retweeted; these Twitter users, identified in Table 2, have provided an initial step to identify the actors who potentially played a key role in the activities and development of the UK OGD community.

With an informed insight into the patterns of communications and levels of online activity, the qualitative data will push the quantitative findings further to provide context and potentially hidden processes that underpin what has been found. We are interested in how the OGD community operates, what influences these patterns observed, and how it is growing and evolving. Shifting from a wide to narrow analytical lens will fluently weave in context to the already important findings.

(1) The functionality of the retweet feature, which reflects the sharing of ideas in the real-world has been suggested to be something of convenience, and the higher degree of connectivity reflects the simplicity and speed to be able to share a tweet. This however is not only affecting the topology of the networks but also the attitudes of the users, promoting the ethos as Interviewee 1 (Employee at international organization involved with Open Data) said: "I've got a couple of hundred retweets and my job is done". In comparison to this, the process required to construct and direct a tweet to a particular user is more taxing, but offers a way to "reach specific audiences" (Interviewee 1), thus also providing other individuals with a public channel to observe communications within the community. These points were widely repeated in other interviews, and may help explain why apparent community-like structures that were identified within the initial network analysis of the mention communications? Examining this further, the use of the #datagovuk hashtag may infer a unified community of discussion, a 'tag' for individuals interested in the same topic to share their views, however as shown in the quantitative analysis (the large number of communities and relatively high modularity) and uncovered within the interviews and timeline of activity, there are a number of distinct groups operating within this community including sole developers, lobbyist groups, non-profit organizations, commercial partners, and government, each with their own agenda and goals, linked by serendipitous events. This is a key point that needs to be discussed when addressing finding (2).

(2) The spikes in activity levels identified in Figure 5 correspond to events identified on the timeline of UK OGD activity [45], which were run or supported by the individuals identified in Table 2. The large spike in Twitter activity during January 2010 in

Figure 5 corresponded to the public launch date of data.gov.uk; identified as an important milestone in the community's activities, the online communications enabled the community to inform those interested in OGD with the news of its release, which acted as an important driver to accumulating more individuals thereafter.

However, as an interviewee 2 (Open Data lobbyist and entrepreneur) explained, the online activities do not fully reflect or provide a detailed account of the development and growth of the UK OGD community, there are social processes such as "government partnership and some other government moves that are taking place, …governments are trying to systemize this [OGD]" that are not revealed by these online activities, and fundamentally effect the online activities observe. Sustainability requires the continuous efforts of the currently involved actors, humans and technologies, and also gaining support of new, additional actors. The technologies that supported this community to grow, not only Twitter but the software technologies that underpin data.gov.uk and other government data portals, were crucial in the community's development. However, technology alone cannot be responsible for the growth of the community; this is driven by the ongoing interactions of individuals and technologies. This is a shared view amongst interviewees, especially in regards to the publication of data. Although the steady publication of government datasets – as shown in Figure 8 – is important, it is pointless–a waste of time and effort–if the data is not being used; As interviewee 3 (An Open Data Government advisor) explains "open data [is] not as an end to itself, but as a feature of something else"; It is the use of the data that will enable the community to remain stabilized and grow.

A critical finding is the lack of a feedback mechanism, not only between the online and offline world, but between the publishers and users of the information, leading to a potential instability within the UK OGD community. The quantitative analysis shows a stable community (with growing resources and activity) with potentially an end point, whereas the qualitative data has revealed many more processes that underpin the stability of the community and the fragility that inheres in this.

An additional finding that draws upon the point raised in (1), the sub-community like structures observed – which actually reflect the "top down, middle out, bottom up" (interviewee 4, civil servant of the Cabinet Office Transparency Board) structure that the OGD community consists of. Building upon this, recognizing that there are multiple communities working on their own agenda and goals here, it may be that when these sub-communities align and work on similar goals, the number of edges (the amount of communications) increase, which is shown by the fluctuating activity in Figure 4 and 5. This may be in response to some common incentive or goal that certain individual may set. This in itself is an interesting question of causality and the reflection of online and offline activity; are the tweets a way of aligning shared interest or are they a result of interests emerging elsewhere? An explanation to the abrupt and frequent fluctuations shown in Figure 4 and 5 may also be a result of the unstructured nature of such communications, as interviewee 2 suggests: "Twitter was good because it organized and allowed things to be done in a fairly fast, free flowing way". Similarly, this could be the case for the data.gov.uk deposit activities; the alliance, enrolment and mobilization of communities promotes action and change, but requires the constant attention and commitment of the individuals and technologies.

Finally, (3) discovering why some individuals were retweeted or mentioned more often than others will help inform the understanding of the structure of the community, uncovering the balance between influence and real-world impact of users. Discussing with the interviewees the importance of certain individuals, their influence and their role within the community, certain individuals were mentioned that corresponded with the those users identified in Table 2, who were said to be crucial in promoting and acting as a catalyst for development. They were identified as being involved in critical events such as government meetings and technological decisions, which were noted as important in the growth of the community. Furthermore, it was also stated that these actors were important in the dissemination of news that helped keep the community up-to-date. This line of questioning also provided some insight into the communities (the hubs of nodes and edges) that the quantitative analysis of the #datagovuk communications network revealed. Interviewees talked about the UK OGD community as one of multiple parts, "there was always these three tiers …it was a change to lock these together" (Interviewee 5, Member of UK Government). However it was the actions of the identified individuals that helped connect these together, in effect acting as the weak ties or intermediaries.

Another critical point raised a number of times by interviewees was the importance of Twitter as a tool for communication for the UK OGD community, enabling fast and direct access to news from within the community; as interviewee 1 suggested, "most people who are doing interesting stuff in the open data space are on Twitter". Additionally, the use of the retweet function was essentially a way to share important and valuable content, often produced by an individual who was high-profile, such as key government officials or representatives of the community. This reinforces the findings of the highly retweeted users, who represent the hubs of nodes and edges in Figure 4. Another important point was Twitter's ability to uncover hidden actors that were contributing towards the OGD community, as Interviewee 2 explains, "Twitter [has] un-surfaced these people" emphasising how it enables one to "find the people who get and care about open data". These individuals although not highly retweeted were provided with a way to share their knowledge and work, potentially not possible without Twitter due to its "self-selecting" nature. Illustrated here is the socio-technical relationship that underpins the communications of the community; Twitter is relied on as a tool to not only provide a way for communication and sharing of news, but to enable undiscovered actors to rise up within the community, providing them with a medium to voice their opinions.

Questions about the future stability of the community provoked mixed views from our participants– both positive and negative – but all recognized that the lack of data usage may slow down the progress of the community's current efforts. All suggested that numbers and statistics only tell half the story of what really is going on within the UK OGD community, the legal issues, policy making, interests of individual and threats to stability are only understood by immersing oneself deep within the virtual and real-world activities; only then can the community be understood.

CONCLUDING REMARKS

This paper has demonstrated how a mixed methods approach can be used to understand the Web as socio-technical phenomena. In the case study of the UK OGD community the use of qualitative and quantitative approaches provide a far richer and coherent understanding of this new movement. The mixed methods

approach offered in this paper, prompted by the disciplinary and theoretical arguments begun by Berners-Lee *et al.*[6] and Halford *et al*. [23] can provide a much richer analysis of the Web. Moreover we have shown that such an approach can address the challenge to be genuinely multidisciplinary which both Hendler *et al.* [25] and Halford *et al.* (ibid) have called for.

At first glance, quantitative and qualitative studies may appear to be studying the same phenomena. It may even be called the same thing – social network analysis – but the aims and outcomes of the research are distinct. A quantitative approach, aims to provide a mathematical and statistical account of the network's empirical structures and properties. The positivist epistemological position underlying this means that the phenomena can be reduced to empirical findings [40]. Alternatively, a qualitative approach aims to provide context, meaning and explanation of why things happen. The results show the contingency and instability of these networks, rather than focusing on their descriptive properties [43]. However, this does not suggest that quantitative and qualitative approaches are incompatible. To the contrary, each approach can be used to complement each other, not by strengthening each other's analytical weaknesses, but by providing different explanation to the phenomena that occurs.

The epistemological and ontological differences between quantitative and qualitative approaches should be embraced in Web Science rather than avoided; both approaches examine the same phenomena from a different perspective, and by combining them we can grasp the Web at micro and macro levels, we can describe its patterns and graphs whilst also attending to the complexities of social behavior and meaning. Computer science's quantitative methods, influenced by positivism, enables large-scale of datasets to be analyzed with statistical rigor, and enables networks of activity to be structurally examined. Sociology's qualitative methods, underpinned by social theory provide ways to examine the underlying content and context of these activities, to provide meaning, elicit purpose and more importantly, develop the analytical snapshot provided by quantitative approach. Together these methods provide a way of grasping the complex and dynamic properties of the Web.

No doubt some scholars will continue to argue for a strict separation of methods, a maintenance of the entrenched positions on either side of the so called 'science wars'. Web Science demands that we move beyond this. We have suggested that the underpinning differences – in research questions, methods and knowledge – that make combining qualitative and quantitative methods – and by extension computer science and sociological approaches – so desirable. By 'irritating' each other, by providing contradictory pictures and explanations, mixed methods provide a way to understand the Web through a much stronger analytical lens; a lens that Web Science needs to look through.

ACKNOWLEDGEMENT

This research was funded by the Research Councils UK Digital Economy Programme, Web Science Doctoral Training Centre, EP/G036926/1.

REFERENCES

1. Abel, F., Gao, Q., Houben, G.-jan, and Tao, K. Analyzing User Modeling on Twitter for Personalized News Recommendations. *Proceedings of the 19th international conference on User modeling, adaption, and personalization*, (2011).
2. Altheide, D.L. and Johnson, J.M. Criteria for assessing interpretive validity in qualitative research. In N.K. Denzin and Y.S. Lincoln, eds., *Handbook of Qualitative Research*. Sage, 1994, 485-499.
3. Anger, I. and Kittl, C. Measuring Influence on Twitter. *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, (2011).
4. Beekhuyzen, J. An actor-network theory perspective of online banking in Australia of online banking in Australia. *Information Systems*, (2006).
5. Berger, P.L. and Luckmann, T. *The Social Construction of Reality*. Doubleday, 1967.
6. Berners-Lee, T., Weitzner, D.J., Hall, W., O'Hara, K., Shadbolt, N., and Hendler, J. a. A Framework for Web Science. *Foundations and Trends® in Web Science 1*, 1 (2006), 1-130.
7. Boyd, D., Golder, S., and Lotan, G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *HICSS-43*, IEEE Comput. Soc (2010).
8. Broder, A., Kumar, R., Maghoul, F., and Raghavan, P. Graph structure in the Web. *Computer Networks 33*, (2000), 309-320.
9. Callon, M. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. (1986), 196-223.
10. Carey, J.W. Linking Qualitative and Quantitative Methods: Integrating Cultural Factors into Public Health. *Qualitative Health Research 3*, 3 (1993), 298-318.
11. Carr, L. and Brody, T. Size Isn't Everything: Sustainable Repositories as Evidenced by Sustainable Deposit Profiles. *Access*, (2007), 1-13.
12. Cha, M. and Gummadi, K.P. Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social*, (2010).
13. Chu, Z., Gianvecchio, S., and Wang, H. Who is Tweeting on Twitter: Human, Bot, or Cyborg? *Proceedings of the 26th Annual Computer Security Applications Conference*, (2010), 21-30.
14. Denzin, N.K. and Lincoln, Y.S. Introduction: Entering the Field of Qualitative Research. In N.K. Denzin and Y.S. Lincoln, eds., *Handbook of Qualitative Research*. Sage Publications, 1994, 1-17.
15. Depauw, J. Web2 . 0 under the Actor-Network Theory point-of-view: conceptualization and definition analysis. *Politics: Web 2.0: an International Conference*, (2008).
16. Edwards, G. ESRC National Centre for Research Methods Review paper Mixed-Method Approaches to Social Network Analysis. *Methods*, January (2010).
17. Emmel, N. and Clark, A. The Methods Used in Connected Lives : Real Life Methods , the Manchester / Leeds Node of the National Centre for Research Methods. *Methods*, September (2009), 28.
18. Esnault, L., Zeiliger, R., and Vermeulin, F. On the Use of Actor-Network Theory for Developing Web Services Dedicated to Communities of Practice. *Practice*, (2006), 298-306.
19. Freeman, L.C. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004.
20. Frick, A., Ludwig, A., and Mehldau, H. A Fast Adaptive Layout Algorithm for Undirected Graphs (Extended Abstract

and System Demonstration). *Lecture Notes in Computer Science 894*, (1995), 388-403.
21. Gross, T. and Blasius, B. Adaptive Coevolutionary Networks: A Review. *Journal of the Royal Society Interface the Royal Society 5*, 20 (2007), 13.
22. Guba, E.G. and Lincoln, Y.S. Competing paradigms in qualitative research. In N.K. Denzin and Y.S. Lincoln, eds., *Handbook of qualitative research*. Sage, 1994, 105-117.
23. Halford, S., Pope, C., and Carr, L. A Manifesto for Web Science? *Web Science 2009*, (2009), 1-6.
24. Heath, S., Fuller, A., and Johnston, B. Chasing shadows: defining network boundaries in qualitative social network analysis. *Qualitative Research 9*, 5 (2010), 645-661.
25. Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., and Weitzner, D.J. Web science : an interdisciplinary approach to understanding. *Communications of the ACM - Web Science 51*, 7 (2008).
26. Java, A., Song, X., Finin, T., and Tseng, B. Why We Twitter : Understanding Microblogging. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, (2007), 56-65.
27. Jin, E.M., Girvan, M., and Newman, M.E. Structure of growing social networks. *Physical review. E, Statistical, nonlinear, and soft matter physics 64*, 4 Pt 2 (2001), 046132.
28. Johnston, R.B. SITUATED ACTION , STRUCTURATION AND ACTOR-NETWORK THEORY : AN INTEGRATIVE THEORETICAL PERSPECTIVE. *Information Systems*, Giddens 1984 (2001), 232-242.
29. Kossinets, G. and Watts, D.J. Empirical analysis of an evolving social network. *Science (New York, N.Y.) 311*, 5757 (2006), 88-90.
30. Latour, B. *Reassembling the Social: An Introduction to Actor-Network-Theory by Bruno Latour*. Oxford University Press, 2005.
31. Law, J. Technology and Heterogeneous Engineering: The Case of Portuguese Expansion. In W.E. Bijker, T.P. Hughes and T.J. Pinch, eds., *The Social Construction of Technological Systems New Directions in the Sociology and History of Technology*. MIT Press, 1987, 111-134.
32. Leskovec, J., Backstrom, L., and Kleinberg, J. Meme-tracking and the Dynamics of the News Cycle. (2009).
33. Letierce, J., Passant, A., Decker, S., and Breslin, J.G. Understanding how Twitter is used to spread scientific messages. *October*, (2009).
34. Macskassy, S.A. and Michelson, M. Why Do People Retweet? Anti-Homophily Wins the Day! *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, (2011), 209-216.
35. Mccabe, C., Watson, R.A., Prichard, J., and Hall, W. The Web as an Adaptive Network: Coevolution of Web Behavior and Web Structure. *Web Science 2011*, (2011).
36. Mislove, A., Marcon, M., Gummadi, K.P., and Bhattacharjee, B. Measurement and Analysis of Online Social Networks. *Social Networks*, (2007).
37. Onnela, J.-P. and Reed-Tsochas, F. Spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences of the United States of America 107*, 43 (2010), 18375-80.
38. Rao, D., Yarowsky, D., Shreevats, A., and Gupta, M. Classifying Latent User Attributes in Twitter. *Proceedings of the 2nd international workshop on Search and mining user-generated*, ACM Press (2010).
39. Reid, A.J. What we want: qualitative research. Promising frontier for family medicine. *Canadian Family Physician 42*, (1996), 387-389, 397-400.
40. Sale, J.E.M. and Brazil, K. Revisiting the Quantitative-Qualitative Debate : Implications for Mixed-Methods Research. *Community Health*, (2002), 43-53.
41. Scott, J. Social network analysis: developments, advances, and prospects. *Social Network Analysis and Mining 1*, 1 (2010), 21-26.
42. da Silva, R., Buriol, L.S., Ribeiro, L., and Dotti, F.L. Modeling the webgraph evolution. *Journal of Computational Science 2*, 1 (2011), 67-79.
43. Smith, J.K. Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher 12*, 3 (1983), 6-13.
44. Tatnall, A. Actor-Network Theory and Information Systems Research. *Innovation*, (1999), 955-966.
45. Tinati, R. UK Open Government Data Community Timeline. 2012. http://users.ecs.soton.ac.uk/rt506/UKOGDTimeline.
46. Tuomi, I. Internet , Innovation , and Open Source : Actors in the Network. *Network*, (2000), 1-34.
47. Welch, M.J., Schonfeld, U., and He, D. Topical Semantics of Twitter Links. *Proceedings of the fourth ACM international conference on Web search and data mining*, (2011), 327-336.
48. Weng, J., Lim, E.-peng, and Jiang, J. TwitterRank: Finding Topic-sensitive Influential Twitterers. *Proceedings of the third ACM international conference on Web search and data mining*, ACM Press (2010).
49. Zimmermann, B. Uncovering Cultural Issues in the Internet of Things: A Design Method. *Internet of Things*, (2008).