The social sciences and the web: From 'Lurking' to interdisciplinary 'Big Data' research

John Bone, Chukwuemeka David Emele, Adeniyi Abdul, George Coghill and Wei Pang

Abstract

'Big data' is an area of growing research interest within sociology and numerous other disciplines. The rapid development of social media platforms and other web resources offer a vast and readily accessible repository of data associated with participants' activities, attitudes and personal information on a scale and depth that would have previously been difficult to access without substantial resources. However, as well as offering opportunities to social researchers, this medium also presents a significant range of challenges. Ethical issues are one much debated area where social scientists are having to reassess their longstanding modus operandi, given questions regarding access to personal data and ambiguities regarding its status and legitimate usage. In addition, the scale of data accessible and the possible skills required to collect and analyse it is also a critical issue, and is an area that, arguably, has received lesser attention. In its infancy, online research could be fairly rudimentary, employing simple techniques to gather information from weblogs, forums and so on. However, the possibilities now presented by large-scale social media platforms has created the potential for more sophisticated research that often requires specialist technical expertise, involving collaborative work by computer and social scientists working together. This is a scenario that raises its own concerns, not least in terms of forging areas of shared understanding between these disparate disciplines sufficient to facilitate such projects. This article addresses such issues, providing a reflection on the theoretical and practical experience of engaging in online research, from fledgling involvement to embarking on a current collaborative interdisciplinary project. The aim is to provide some insights to other social scientists with respect to some of the potential advantages and pitfalls of web research, while a flavour of the current project, exploring Scottish Referendum and UK General Election related Twitter data, is also presented.

Keywords

Interdisciplinary research, big data, online ethics, Scottish independence referendum, UK general election, social science, computer science

Introduction

With recent advances in Internet technology, web based applications, smart devices and the associated exponential growth of social media, the use of data drawn from these sources is now being utilised in a broad range of research fields. For example, research into crisis situations, acute events, disasters and other societal concerns are increasingly drawing on social media data to conduct their studies (Bruns et al., 2012; Guy et al., 2010; Hughes and Palen, 2009; Liu, 2009; Mendoza et al., 2010; Palen et al., 2010). Moreover, it seems clear that within sociology, as with other social scientific disciplines, there has been a significant and rapidly growing interest in online research and, more recently, in its latest incarnation in the form of so called 'Big Data', as tentative moves to explore the potential of online resources have become both more numerous and more sophisticated. Such a

University of Aberdeen, Aberdeen, UK

Corresponding author:
John Bone, School of Social Sciences, F44 Edward Wright Building, Dunbar Street, University of Aberdeen, AB24 3QY, UK.
Email: j.bone@abdn.ac.uk

development seems almost inevitable, given the growing repository of opinion, reflection and relationships that have become central features of the Internet, and a key focus of its users. That said, it has also been recognised that this new and growing resource presents a number of prodigious challenges as well as potential benefits for social researchers (Halford et al., 2013; Tinati et al., 2014).

In the first instance, it has been suggested that the standing of traditional social research, of itself, has been shaken by the increasing gathering and analysis of 'social' data by commercial organisations (Savage and Burrows, 2007; Tinati et al., 2014). There is also a recognition that the increasing prevalence of social media, as well as providing a window on a vast sea of opinion, comment and self-presentation, is also becoming a central feature of contemporary social interaction, to the of extent reshaping the latter to some significant degree (Lipinski-Harten and Tafarodi, 2012; Tinati et al., 2014). It is clear that this raises the possibility of sociologists, and associated disciplines, not only having to re-assess their methodological approaches to understanding interactive processes, but also potentially having to reassess the theoretical frameworks that have informed the latter (Halford et al., 2013).

From a more positive standpoint, the rapid expansion of this virtual realm offers a window that, in many instances, reveals both a level of intimacy and scale with respect to opinion and social interaction that was previously inaccessible and/or beyond the scope of most prior social scientific research. Given the potential of this resource, as suggested, a growing number of social scientific pioneers have dipped their toes into this stream of data. However, it may be reasonable to suggest that the methods employed in this pursuit, and the ethical questions that have arisen from this type of work, remain far from being fully formed. It may be more accurate to suggest that such social scientific research is proceeding via a process of tentative trial and error, not least as the digital landscape itself continues to evolve at a staggering pace, such that the learning curve required to enter this field may be as slippery as it is steep. These developments have also raised other issues relating to the expertise required to gather and analyse data from the digital realm, given the emergence of ever sophisticated means of gathering, sorting and visualising the latter that are being employed by computer scientists and associated information technology (IT) specialists (Halford et al., 2013).

This article considers the rapid evolution of both the digital landscape and the equipment required by the social scientific explorer in light of first hand practical experience, from initial 'rough' application of standard social scientific techniques involved in the collating and analysis of small pools of qualitative data to, in the present, social scientific collaboration with computer scientists with work aimed towards more fully exploiting this potentially rich resource. A reflection on the theoretical and practical issues involved in this type of work is presented, while offering a flavour of its potential via presentation and discussion of some preliminary findings of a current interdisciplinary work in progress. Aside from the perennial issues associated with social scientific methods and ethics, one of the major hurdles to utilising this data to its full potential arises simply in terms of social scientific engagement with colleagues in the computing sciences, given the evidently radically different skills, approaches and, not least, lexicons that prevail in each disciplinary setting.

A toe in the water

My own experience in this area has followed the sort of pattern of development suggested above. My first engagement employing web data began almost a decade ago, at a point where some of the key platforms attracting contemporary research interest, such as Facebook, Instagram and Twitter, were either yet to be founded or were at such a fledgling stage in their development that they had not yet entered wider public awareness. Initial research, conducted with a colleague, focused on applying relatively standard qualitative techniques to data gathered from a variety of well-established blogs and web forums, where debate and opinion regarding a particular topic we were exploring at that time was extensively represented. Applying relatively crude methods by contemporary standards, streams of relevant discussion were copied and pasted into word documents and analysis conducted to identify themes. One of the key sites we had accessed was also approached with a request to post a web form, with the aim of gathering subjects for more traditional data gathering, albeit that a good deal was collected via email. While bearing in mind that the subjects on this study were largely self-selecting, this approach provided an opportunity to gather data from a much larger group of subjects, drawn from a much larger geographical spread and at minimal cost in terms of both resources and time, such that this would have been impossible by other methods (Bone and O'Reilly, 2010). This is one of the more novel and potentially advantageous features of web based research, in that the web provides access to, in a sense, pre-collated 'niche' data, often already organised under particular themes and topics given the nature of web blogs, forums and now Twitter hashtags and so on. Thus, as above, the researcher, possibly with little or no funding, who may previously have struggled to gather a small sample of interview or questionnaire respondents, is now confronted with potentially hundreds, thousands or even, as is the case with Twitter, millions of 'subjects' with comparatively lesser outlay in terms of time and cost. Difficult to reach groups also become more accessible online (Brownlow and O'Dell, 2002).

It was also recognised, as other researchers in this area have noted, that 'lurking' on web forums also avoids the longstanding problems associated with the 'halo' or 'Hawthorn' effect, given that subjects responses cannot be influenced by the presence of the researcher (Mann and Stewart, 2000 as cited by Brownlow and O'Dell, 2002).

Conversely, however, the advantages of being an unnoticed observer of online opinion and interaction also means that the subjects' control the dialogue, raising the possibility that precise questions that the researcher may have in mind may not be pursued. There is also a question of the 'honesty' of subjects online interactions and self-presentations, in that there has been considerable media and academic debate re the capacity for individuals to construct false impressions, and indeed false identities online, potentially creating a gulf in the consistency between these interactional realms. The self-selecting nature of online groups may also skew perceptions of particular issues, while the generalisability of findings may be more difficult to assess. Conversely, on this point, some commentators on the evolution of 'big data' suggest that this approach may eventually reduce the biases that emerge from self-selection and even sampling (Mayer-Schönberger and Cukier, 2013). In effect, they note that the increasing capacity for big data research to draw on ever larger samples will reduce the potential error inherent in current techniques. While presently the fact that online research evidently skews any sample towards the Internet 'savvy', much as current political polling may disproportionately represent the view of those who continue to use telephone land lines, these sorts of effects are liable to reduce over time as web activity becomes more ubiquitous (Mayer-Schönberger and Cukier, 2013):

> We tend to think of statistical sampling as some sort of immutable bedrock … (b)ut the concept is less than a century old, and it was developed to solve a particular problem at a particular moment in time under specific technological constraints. Those constraints no longer exist to the same extent. Reaching for a random sample in the age of big data is like clutching at a horse whip in the era of the motor car. We can still use sampling in certain contexts, but it need not – and will not – be the predominant way we analyse large datasets. Increasingly, we will aim to go for it all. (Mayer-Schönberger and Cukier, 2013: 31)

Aside from these types of considerations regarding the approach, our own early research conducted in this area also introduced us to some of the particular ethical dilemmas presented by online studies (Bone and O'Reilly, 2010; Brownlow and O'Dell, 2002; Snee, 2013).

Online ethics

At the time of conducting the original study, it might be suggested that in terms of ethics there was a kind of rule of thumb approach taken to dealing with the data, largely via applying standard ethical guidelines as best we could. Data gathered from online forums, email interviews and web forms (as with telephone interviews) was anonymised and subject to the same principles of informed consent that would apply to traditional qualitative research. It must be conceded, however, that at that time we tended to regard web forum and blog data as being of a different nature, as being in a sense 'fair game' as it was already 'published' and publicly available. As we were dealing largely with very public fora, and large volumes of responses, we considered that obtaining informed consent seemed both impractical as well as being unnecessary. In the first instance, we were of the view that the extracts being reproduced from these sources had already been subject to far greater public exposure than would be the case with inclusion in an academic publication. We also anonymised subjects while giving details of the sources and time of capture which, in retrospect, may have been less than wholly satisfactory given that a word search could reveal the identity of posters who had not used an online pseudonym (Bone and O'Reilly, 2010).

On further reflection, the overall approach to issues of privacy and informed consent that informed this early research might appear a little naïve as, in light of subsequent debate, such questions seem much less clear-cut than they were once regarded. As Snee (2013) observes, there remains a question mark over whether blog data can be regarded as private or public, legitimate or intrusive. The key question here appears to be the extent to which the latter is treated as private opinion, that is being intruded upon and/or exploited by the researcher, or published text that is merely being quoted in line with regular academic practice. This has been documented by other researchers, reflecting the fact that online research has created a situation where '… there appears to be a 'technology lag' where ethics has played catch-up to the various methodological options available to the researcher' (Convery and Cox, 2012: 51).

As is clear from a good deal of the growing debate around online ethics, some of the perennial questions governing conventional social research re-emerge while the domain raises a range of novel dilemmas. This is something that I have become very aware of when engaging on a current online research project, discussed below, which was embarked upon after a considerable hiatus in terms of my initial foray into web-based research. Thus, as will be discussed, when re-engaging with this type of work after a gap of a few years, it was clear that not only had the digital landscape undergone a considerable metamorphosis – and consequently the research techniques and the skills required to apply them – but the ethical issues had also been similarly transformed. For example, the potential of social media data to be exploited for commercial advantage was something that became clearly apparent to me while attending a 'big data' conference in early 2014.

The conference was attended by a mixed audience, including former academics now employing their research skills on behalf of marketing organisations, the latter underlining some of the prescient concerns re the commercial incursion into previously academic social scientific space that has accompanied e-research (Savage and Burrows, 2007). Some indication of the level of the increasingly sophisticated techniques being commercially applied became clear to me when

one presenter noted that a major retailer had been employing collection and analysis of informal exchanges on social media regarding desired products, including location data, to make decisions regarding the stock allocation of particular lines in specific stores across the country. This struck me as being not only clever but somewhat unnerving, given the way in which public opinion could be so precisely explored, while possibly raising the potential for this to be exploited by carefully placed target marketing and 'viral' online advertising.

This raises the important question as to whether the emergence of this new resource provide us with a route to greater understanding and more rational, evidenced-based decision and policy making, as its many supports claim, or will it lead merely to exploitation and control by powerful interests amid a 'fetishization' and 'dictatorship' of data; the view espoused by more guarded observers (Mayer-Schönberger and Cukier, 2013). Regarding the latter, in one sense, the online data revolution might be regarded as potentially representing the apotheosis of what Kornhauser (1959) once raised concerns about in the mid-20th century, of a 'mass society' where a highly atomised population is subjugated by the carefully crafted, and largely mediated, manipulation of an exploitative elite. On this point, there are also growing concerns with respect to the use of 'dataveillance' techniques, including their deployment as a means of extending control, micro-management and exploitation in the workplace (Connolly, 2014).

While such alarming considerations re the possible ethical, or dramatically unethical, uses of big data must be borne in mind, it is also important to consider that, as suggested above, these same techniques can be employed to swiftly identify, intervene and control epidemics and contagious diseases (a broad area where many of these techniques were developed), and can be employed to both treat and confine their effects, understand their aetiology, while evidently similar strategies can be applied to identify and develop remedial solutions with respect to a range of social 'ills'. From a political point of view, social media research techniques may provide a window on public opinion, enabling policymakers to become more responsive to the needs and wishes of electorates or, as noted above, can be employed as a vehicle to subvert democratic processes. In short, this returns us to the notion that the uses of any investigative technique, large or small scale, ultimately depends on the integrity, good faith and ethical stance of practitioners, who must constantly weigh the advantages and disadvantages of collection and analysis and the implications of its potential usage in light of the effects on both subjects and those who may benefit or, alternatively, be disadvantaged by its employment.

Re-entering the digital field

My re-engagement with online research can to some extent be regarded as a fortuitous accident. As noted, it is clear that the technological progress that has taken place in the interim

between simple data gathering and the collection and analysis of 'big data' is substantial. In fact, it may be suggested that this has occurred to such an extent that much of the technical work involved has moved significantly beyond the skill set of the majority of social scientists and into the domain of the computer sciences. The collaborative project discussed below was instigated by the latter with the principal aim of developing new techniques and programmes that might be used by social researchers. This evidently necessitated input from social scientists to aid the development of the proposed research tools. As one of the few social scientists at our institution who had any prior experience in this area, I was enlisted.

A stranger in a strange land

The current project involves a substantial group of researchers of which our team, composed of four computer scientists and one social scientist, is a subset. The first few meetings of the team was in many ways unnerving for a social scientist, given the gulf of perspective, focus and, disciplinary lexicon that exists between the disciplines, an issue returned to at various points below. A good deal of the discussion, initially at least, was conducted in what, from the perspective of the social scientist, may as well have been impenetrable computer code. Moreover, as initial meetings progressed, it became clear that my interests, focusing specifically on the analysis, meaning and interpretation of data as opposed to its manner of collection and presentation, appeared at odds with that of my colleagues who, understandably, seemed more engaged with technique and process. Thus, in my experience, this has been one of the main obstacles that had to be overcome.

For some time, after the commencement of the work, there were various points where I would be presented with output that appeared to be almost incomprehensible and whose utility seemed similarly unfathomable. One notable example related to visualisations of keyword 'clustering' drawn from social media exchanges around our chosen topic. This process seemed technically as well as aesthetically elegant while being, nonetheless, ultimately meaningless from a social scientific perspective. In qualification, there was some sense in which similar techniques might be applied to produce useful social scientific output in very specific cases. However, in this instance, it seemed clear that this type of approach was inappropriate for our particular purposes. This apparent disciplinary chasm led me at various points in the initial stages to question whether we might not be too far apart to engage in meaningful collaboration and, thus, whether I had been wise to become involved in the project from the outset, while these reflections in one sense could be perceived as both a cautionary and advisory note for other social scientists considering this type of cross-disciplinary collaborative venture.

I suspect that the above scenario, as other researchers have also observed, is something that is likely to be experienced by many social scientists engaging with 'big data'

research, albeit that this may be situation that is alleviated or even avoided, altogether should these early projects, as is intended at present, lead to the development of greatly simplified applications, programmes and techniques that might be employed solely by social scientists (Halford et al., 2013). A further possibility, however, and one which I suspect may be more realistic, is that effective large-scale online data research may not be wholly achievable without interdisciplinary collaborative effort or, alternatively, some social researchers may confine themselves to utilising pre-existing simplified applications to conduct fairly basic investigations while others work more closely to the developmental horizon with more specialist assistance and on more technically sophisticated research. Even in the event that powerful but sufficiently simplified tools do emerge, given the rapid evolution of the web it may well be the case that there will also be an ongoing requirement for social scientists to work in developmental projects, as a means of ensuring that our methods keep pace with rapidly changing technologies and that we do not fall behind or are excluded from this potentially fertile field (Halford et al., 2013).

By contrast, I would also note that these reflections have also been advanced from a more positive perspective, as a means of highlighting the fact that difficulties of this nature are to be expected, and should not be considered as insurmountable – as grounds for avoiding interdisciplinary web based research or for abandoning it at the first hurdle – as they can be overcome with time and mutual goodwill. In this I was possibly fortunate, in that my computer scientist colleagues have been highly amenable and generous in terms of explaining technical processes and patient when trying to reconcile my need for data to be collected and presented in specific ways with their seeming, and quite understandable, aforementioned preoccupation with technique. Gradually over time, however, and while significant areas of opacity remain, a growing zone of mutual understanding has emerged together with a sense of collective engagement. As noted, the point of this brief discussion is to highlight the fact that embarking on this type of work is not an easy process but, as is discussed below, with time, effort and patience on both sides, good working relationships can be established, while the potential bounty for the social sciences makes this very worthwhile. In fact, as discussed above, and echoing the concerns expressed by Savage and Burrows (2007), to do otherwise may potentially lead to a narrowing of the academic social scientist's domain over time with respect to a range of empirical studies.

The project

One of the other issues that had to be resolved early on by the current team was how we might pursue the primary goal of the research – the development of a new techniques for web-based social scientific research – while being able to advance this within the context of a substantive/empirical project, where the efficacy of specific techniques and approaches could be assessed together with the practicality and utility of collaborative projects of this nature in a more general sense. Evidently, we felt that the more topical the project the better the outcome given that this would be much more likely to offer common ground, providing a more likely a basis for building mutual understanding and collective interest, than a more 'niche' sociological inquiry. Fortuitously perhaps, the timing of this phase of the project coincided with the lead in to the Scottish independence referendum, where it seemed clear that something of significant sociological as well as political and wider public interest was taking place that had ramifications for the United Kingdom as a whole and, crucially, might lend itself to the sort of case study we had in mind. Below an overview of our current project, its background and some preliminary findings have been presented as a means of exemplifying aspects of this type of work in terms of process, problems and potential.

Some substantive background

The decision to explore data relating to the independence referendum was taken in the months leading up to the vote, when the public and media debate was increasing, and it was beginning to become apparent that, what was once assumed to have been pretty much a foregone conclusion in October 2012 when UK Prime Minister David Cameron acceded to the vote, had taken an unexpected turn.

Since the late 1970s, it has been argued that there has been a growing divergence of the mainstream national and political cultures between Scotland and the rest of the United Kingdom and, in particular, between the former and that of southern England (Johnston and Pattie, 1989). Prior to this point, it was suggested that, while Scots had always retained some sense of national identity, this was considered to be reconciled with a notion of Britishness. Thus, according to Smout (1994), the Scots have been bearers of a dual national identity. Moreover, in line with the broadly accepted view, within the social sciences that modern nation states can be characterised as 'imagined communities', it has been argued that Scotland's self-identification has been informed by the often contradictory symbolism and imagery of tartanry, the highlands as well as the gritty legacy of Scotland's industrial heartlands. Simultaneously, however, a majority of Scotland's populace have also considered themselves to be part of a broader British community, represented by its associated icons and symbols of the Royal Family, the BBC, The NHS, the Welfare State and other UK public services (Anderson, 1983; McEwen, 2010):

> In Scotland, the welfare state contributed to reinforcing identification with and belonging to the UK state by promoting a social conception of British nationhood which could rest alongside and stretch beyond the territory and boundaries of the Scottish nation. It is not suggested that the welfare state represented an explicit nationbuilding strategy to reinforce Scots

commitment to the British state structure. Rather, the expansion of state welfare in the post-war period had fortuitous nation-building consequences, by strengthening a sense of solidarity across the national state territory and reinforcing an attachment to the state as the guarantor of social and economic security. (McEwen, 2010: 86)

From the 1980s, however, it seems clear that this reconciliation of national identifications has begun to slowly unravel, punctuated by the establishment of the Scottish Parliament in 1997 and critically underlined by the unanticipated scale of the Yes vote in the Scottish referendum and, recently, the Scottish National Party's (SNP's) historic landslide result in the UK General Election of 2015. It may be suggested that what was a slow-burning ember of national self-determination in Scotland, and an associated rejection of the trajectory of the political culture that has developed at Westminster, was suddenly and full ignited in the months preceding the referendum. It has been suggested that views on Thatcherism, and indeed 'Blairism', had long been divided to some extent, with the more prosperous middle classes being more supportive (or at least less averse) with respect to the United Kingdom's neoliberal political consensus and, thus, less likely to support independence (Curtice, 2014; Johnston and Pattie, 1989). However, as is discussed in more detail below, an increasing unease across a large swathe of, in particular, Scotland's less affluent populace with respect to UK politics since the rise of Thatcherism – a trend punctuated by the imposition of the poll tax in the late 1980s – appeared to have gained further traction in response to the post-crisis austerity politics embraced by the Conservative led coalition government. In effect, the referendum appeared to emerge as a focus for current and longstanding discontents that had hitherto lacked a political vehicle or outlet for expression (Davidson, 2014; Keating, 2010). Moreover, growing support for the SNP's 'offer', and independence itself, appeared further influenced by the perception that the Labour Party, the once unassailably dominant party in Scotland, no longer offered a significant opposition or alternative that chimed with the political culture of increasing sections of the Scottish public, and particularly those from working-class backgrounds (Curtice, 2014; Davidson, 2014; Davidson et al., 2010). Some commentators have suggested that these trends may be fairly complex in that, at an individual level, Scots may only be marginally more 'leftist' than their English counterparts. Nonetheless, there appears to be a significant divergence with southern England in terms of the broader political culture (Keating, 2011):

> It is not that Scots are further to the left than the English. Surveys show them only very slightly to the left and close to voters in the North of England (Rosie and Bond, 2007). In Scotland, however, resistance to Thatcherism had a vehicle in national identity, with the nation being reinvented on social democratic lines. Since devolution, policy in Scotland under both Labour/Liberal Democrat and SNP administrations has been more traditionally social democratic than in England, with no quasi-privatization in the health service, support for comprehensive education and a generally more Universalist attitude to social services. (Keating, 2010 as cited in Keating, 2011: 8)

Aside from such nuances in terms of the underpinnings of Scotland's emerging self-determination, the growing extent of the cleavage with the South and Westminster politics appeared came as a surprise to the UK government and opposition parties, who had been reasonably sanguine regarding the outcome just months before:

> David Cameron is growing increasingly confident the UK government will win its battle for a single-question referendum asking the Scottish people to simply vote Yes or No to independence. The Scotsman has learned the Prime Minister is prepared to let the SNP delay the poll until its preferred date of October 2014 in the knowledge he will win the crucial argument in favour one straightforward question. Senior No 10 sources suggest Mr Cameron's belief that he will outflank Alex Salmond on the framing of the ballot paper has been bolstered by the SNP's cave-in on the Scotland Bill. (The Scotsman, 24 March 2012)

In fact, at the time of assenting to the referendum, as the above implies, David Cameron, was considered to have laid the Scottish question largely to rest by avoiding a 'devo max'[1] option on the ballot, where a yes vote looked like a clear possibility, in favour of a straightforward full independence question, where the latter had looked virtually impossible at that time.

The increasing fragility of this assumption became ever clearer in the weeks prior to the referendum vote, precipitating a 'panicked' trek northwards from leaders of all of the main Westminster parties to present what became known as 'the vow' on increased Scottish parliamentary powers. This was accompanied by a concomitant series of emotional pleas from Westminster politicians, highlighted in the national media, aimed at averting the break-up of the United Kingdom:

> I speak for millions of people across England, Wales and Northern Ireland – and many in Scotland, too, who would be utterly heartbroken by the break-up of the United Kingdom. Utterly heartbroken to wake up on Friday morning to the end of the country we love. (Prime Minister, David Cameron, The Independent, 16 September 2014)

In terms of our project, as noted, both the apparent speed of this political transition, and its potential impact or UK politics as a whole, captured our interest. Principally, we were concerned with the factors that had shaped the pre-referendum debate in Scotland and how might we come to an understanding of its wider effects, if any. This latter considerations led to us widening the scope of the study to explore Twitter data related to the subsequent UK general election (taking place some 8 months later), as well as referendum related data, with the aim of identifying synergies and disparities with respect to the issues being raised in relation to

each poll. As can be identified from the sample of our preliminary findings presented below clearly underline the view that, in line with much of the public debate at the time, and not least reflected in the SNP's historic landslide general election result at the general election, that the Scottish issue, and the debate surrounding it, appear to have had a substantial impact on the UK political scene as a whole.

Methodology

Data gathering and technical issues

For this project, we opted to focus on Twitter data, a platform whose utility has been recognized by a range of sociological researchers (Tinati et al., 2014). This social network platform employs flat and flexible communicative structures. For example, users can include hashtags (i.e. keywords, prefixed with the hash symbol '#') in their tweets to make these messages visible to others following the hashtag, and the establishment of shared hashtags is rapid and ad hoc. Hashtags could be related to anything (e.g. an event, etc.). Crucially, in the Twitter platform, hashtags provide a mechanism for conversation and update threads between users even if these users are not already 'following' one another in the social network.

The major challenge in doing research on the use of Twitter for communication is to capture a comprehensive (or at least representative) sample of tweets, which relate to the event in question and/or the population under study. We adopted a relatively simple and straightforward approach to this challenge by focusing on tweets, which contain the relevant topical hashtags (and keywords) related to the study. For example, discussions around the Independence Referendum had the #Indyref hashtag associated with a substantial proportion (as well as additional, adjunctive and sometimes overlapping discussions using #YesScotland, #BetterTogether, etc.), while the General Election discussions used the #ge2015 and other ancillary hashtags.

By tracking topical hashtags (and associated keywords) and capturing identified tweets, we have attempted to generate a dataset of the most visible tweets relating to the event in question. The hashtags and keywords were collaboratively gathered from news bulletins, party manifestos, political debates and so on. This does not mean that we were able to capture all messages relating to the event in question or its implications, however, as some users tweeting about a particular issue will be unaware of the existence of the central hashtag, will use a different hashtag variant, or even be unfamiliar with the concept of hashtags altogether. We were aware of these limitations and attempted to mitigate them by tracking a range of relevant hashtags and other keywords.

Twitter provides access to public tweets through two key elements of its Application Programming Interface (API): the search API and the streaming API. Of these, the former can be used to retrieve past tweets according to a range of criteria (including keywords/hashtags, senders, location, etc.), within set limits: in the first place, the search API will only return a limited number of tweets, and therefore cannot be used to retrieve a comprehensive archive of past tweets containing specific hashtags, for example. Additionally, there are in-built limits on how many keywords or users can be queried at any one time or within set timeframes. Where the search API is focused on past content, the streaming API, by contrast, can be used to subscribe to a continuing stream of new tweets containing specific keywords or originating from specific users or locations; here, too, however, significant limits on the number of users or keywords which can be followed do apply. (It should be noted that some such limits can be overcome, at a cost, by accessing the Twitter API through one of a number of third-party resellers of Twitter content.)

Given these limitations of the Twitter API, any research method which seeks to establish a reasonably comprehensive dataset of tweets related to a specific event will need to begin tracking the event as it happens, or otherwise it will risk missing these early tweets as they will eventually no longer be retrievable using the search API. Furthermore, follow-up tweets must be captured either by using the streaming API to subscribe to an ongoing update feed of relevant tweets, or by regularly retrieving the latest past tweets through the search API. Even such retrieval methods cannot guarantee a comprehensive capture of Twitter data, however, outages on the side of server or client, or transmission problems between them, cannot be ruled out altogether, and may result in message loss. Furthermore, there are very few reliable means of comprehensively cross-checking the dataset for its veracity, since the Twitter API constitutes the only point of access to the Twitter stream, which is available to researchers. No dataset captured by using the Twitter API is guaranteed to be entirely comprehensive, therefore, especially where research focuses on identifying broad patterns in Twitter activity from a large dataset. Nonetheless, such research remains valid and important as a large sample of relevant tweets can be captured despite these limitations, which usually go well beyond the level of qualitative responses one could gather from standard social scientific approaches.

We developed an in-house tool (using R language) for capturing tweets. The tool utilises twitter's Search API. Data captured through the tool contains the following data points for each tweet retrieved:

- **created_at:** tweet timestamp in human-readable format
- **isRetweet:** a Boolean that says whether this tweet is a retweet or not
- **retweetCount:** the number of retweets of the original tweet
- **text:** contents of the tweet itself, in 140 characters or less

Data analysis

Many existing studies use custom-made research tools that are barely discussed and so cannot be validated by other scholars and the excuse often hinges around the argument that Twitter and other social media platforms are relatively new, and interdisciplinary approaches for the qualitative and quantitative study of datasets drawn from social media platforms are not standardized. This undermines the potential to replicate and translate such studies to similar contexts. In this section, we, therefore, outline the research approaches adopted in this work, and provide more detailed methodological background to our research. This will allow other researchers to apply these approaches to their own areas of research, to generate comparable datasets, and to replicate or challenge our findings.

Text and opinion mining sits at the heart of social media analytics because most platforms utilized the micro-blogging approach, which allows users to express their views and opinions with a mixture of text, images, and gestures (in the form of emoticons). We utilized a variety of tools and packages in R to process the dataset gathered from Twitter. For example, we employed basic statistics and text mining techniques to count and compare specific communicative patterns within the dataset. Such analysis could provide metrics and statistics that describe the Twitter activities captured in the dataset. Specifically, we counted the frequency of relevant terms/keywords within the dataset and used standard information retrieval techniques such as stemming and stem completion to normalise the keywords. The volume of the dataset was large and time consuming to process and so a representative sample was chosen (that is tweets communicated one day before, during and after the Scottish Independence Referendum and UK General Election). To do this, the dataset was filtered for relevant tweets using specific keywords and specified timeframes only (since we did not collect the user details including screen name, User ID, location, and so on due to ethical considerations discussed in the Ethics Section). In this light, we considered tweets, which had at least one of the hashtags in the list of hashtags and was tweeted or retweeted for these events during the selected time frame.

As outlined below, we conducted some content analysis and could identify most prominent keywords within the corpus. For the next phase, we are utilising *latent dirichlet allocation* (LDA) to identify topics in the dataset and create topic models, which allow the probabilistic modelling of term frequencies in a corpus. The fitted model can be used to estimate the similarity between tweets as well as between a set of specified keywords using an additional layer of latent variables, which are referred to as topics. We also intend to employ sentiment analysis techniques to gauge attitudes to particular themes associated with keywords that have emerged as being prominently represented in the data.

Ethics revisited

As noted above, the ethical scenario with respect to web data has shifted to some extent in tandem with the rapid development of both techniques, the digital landscape itself and, to some extent, in the way in which many universities have extended oversight in this area. One particular hurdle we had to negotiate was that, in line with the latter, ethical decisions no longer rest with individual academics, research teams or even disciplines but often have to be ratified by institutional committees prior to bids being submitted and/or work commencing. This was the case for our project team. While we did not experience serious difficulties in this regard, it can be suggested that the requirement to achieve ethical clearance can be seen to be particularly problematic for online researchers, given the fact that, as noted above, this type of work can raise quite novel considerations while the application of standard ethical guidelines may not be appropriate or workable.

In our case, when applying for ethical approval, it was anticipated that there may have been some issues regarding the fact that we would be exploring personal political opinions and, further down the line, potentially presenting responses from specific individual posters as we 'drilled down' through various levels of the data. Certainly, we were aware that any personal identifiers would need to be excluded. What we were less prepared for was a suggestion that there might be an issue with us collecting data from minors (under 13), as this was brought to our attention by the committee. While this was always a possibility, we considered it fairly unlikely or extremely minimal in relation to a study of this nature, while our ethics committee came to the view that this was not a serious issue. However, one condition of approval that, in retrospect, does seem problematic is that, with the aim of supporting posters' anonymity, we were not allowed to collect global positioning system (GPS) location data with the tweets. We could not then subsequently conduct one aspect of our proposed analysis that might provide insight with respect to regional variations in keyword/topic frequencies that may have further illuminated our findings. Evidently, we would not have presented location data associated with individuals, but the ability to group responses together on a regional basis is something that would not have compromised the interests or privacy of subjects. This is something that we intend to address in relation to future work.

Preliminary findings

For the specific purposes of this piece, as noted, it is only possible to offer a flavour of ongoing work at this stage, given that both the data analysis and the tools being developed are very much works in progress, albeit that to attempt to go further would also have been unwieldy and beyond the scope of this largely reflective piece. However, we have

presented some preliminary findings drawn from a selection of the data at the 'top level' of analysis, covering the relative frequency of keywords related to two of the key hashtags being explored on dates around the Independence Referendum 2014 and General Election 2015. As can hopefully be gleaned from the discussion below, this preliminary stage in the analysis has nonetheless revealed some interesting results.

We have approached this stage, as suggested above, on the assumption that the frequency of aggregated keyword usage within Twitter postings is broadly representative of an associated issue's significance in relation to the political processes being commented upon and discussed, that is, in relation to associated hashtags. As noted, the keywords/topics were those that the team considered were reflective of some the key contemporary political and socio-economic issues permeating the referendum and subsequent general election debates among politicians, online pundits, web forums and as represented in the mainstream mass media. By applying the same set of topic related keywords in relation to referendum *and* general election associated hashtags, we also hoped to gain an impression of consistencies and disparities in terms of the relative frequency and, thus, potential significance of political and socio-economic concepts and issues in each case. Evidently, this stage of the research process is necessary in terms of identifying topics and issues for further exploration and excavation within the data, although this does not reveal attitudes in relation to particular concepts and issues, as deeper (contextual and sentiment) analysis is evidently required here, which will be conducted in a further phase of the analysis. Nonetheless, we consider that it is legitimate to offer some provisional observations based on the prominence, or lack of, with respect to particular keywords, when considered against the backdrop of the wider debate associated with the referendum and general election (see Table 1).

The Scottish independence referendum (18 September 2015)

Figures 1 to 3 illustrate, as obviously anticipated, many of the terms being circulated on Twitter around the period of the referendum poll refer directly to the political parties and politicians involved. It is also the case that, as might be expected, 'nationalistic' terms, that is, Scots, English and so on are highly prominent. The specific context in which these phrases were employed will be the subject of further analysis as we 'drill down' through the data. What will be interesting is the extent to which the employment of these terms is simply descriptive or, in fact, reflects something more meaningful in terms of a pre-existing or emerging ethno-national dimension to the independence process. This is interesting given that numerous commentators, including prominent members of the SNP, retain the view that the independence

Table 1. Search terms for Independence Referendum and GE2015.

Hashtags, Keywords and Search terms		
Debate	Rich	deficit
Westminster	housing	austerity
Cameron	second homes	independence
Miliband	banks	nationalism
Sturgeon	banking	Scotland
Clegg	bankers	England
Farage	economy	Wales
UKIP	crisis	Ireland
Plaid Cymru	recession	Scottish
Libdem	recovery	Scots
Liberal	bailout	English
Communist	interest rates	Irish
Labour	socialist	Welsh
Labor	socialism	cuts
Conservative	red tories	health
Tories	neoliberalism	NHS
Liberal Democrat	globalisation	privatisation
UKIP	globalization	oil
SNP	consumption	energy
Green	protest	housing
left wing	budget	house price
right wing	business	renting
left-wing	consumer	rents
right-wing	benefits	mortgages
employment	scroungers	buy to let
unemployment	skivers	BTL
work	hard working	tax
jobs	immigration	taxation
wages	emigration	inequality
pay	pension	tax evasion
salaries	pensioner	tax avoidance
low Pay	public service	insecure work
part time		insecurity
		zero hours
		welfare
		public spending
		debt

movement has been largely mobilised via a rejection of UK politics and economic policy, as suggested above, as opposed to nationalistic or, indeed, anti-English sentiment per se.

As to the prevalence of more generic political and socio-economic keywords, a few notable terms were consistently employed across this selection. Oil features prominently in each data selection, reflecting its longstanding centrality to the Scottish independence debate. Work, tax, pay, NHS, cuts (and austerity), banks, pension related terms also feature across the 3 days selected, pointing towards some of the key socio-economic factors that concern those engaged on the online independence debate. These are areas where we would expect to find that pro-independence commentators would
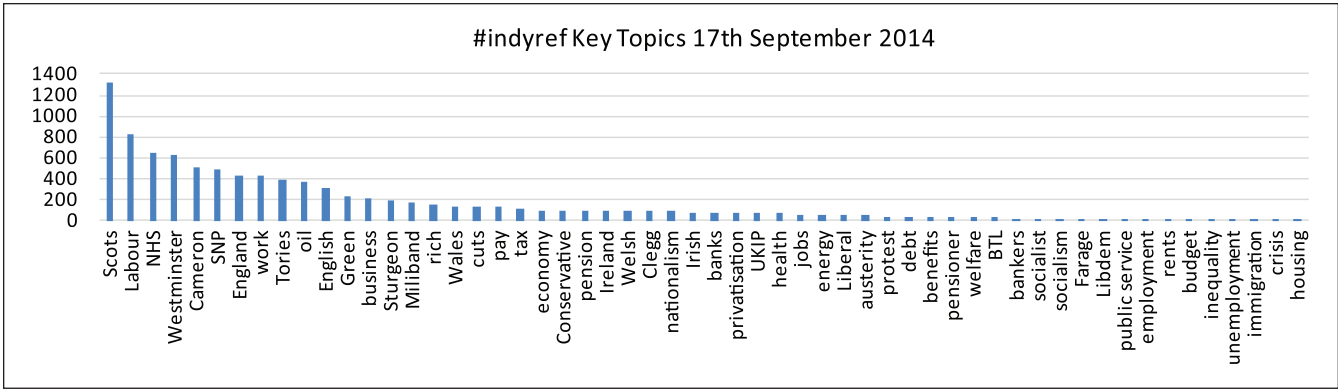
Figure 1. Drawn from filtering of 28,630 tweets – Generic terms Scotland; Independence; Referendum; Scottish excluded from data as high responses skewed presentation – frequencies <10 also excluded.
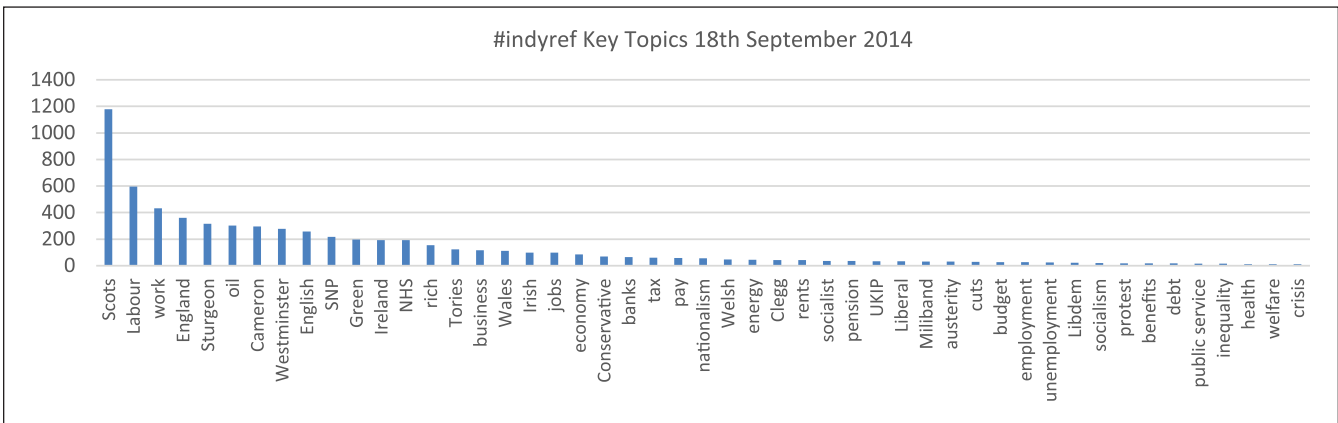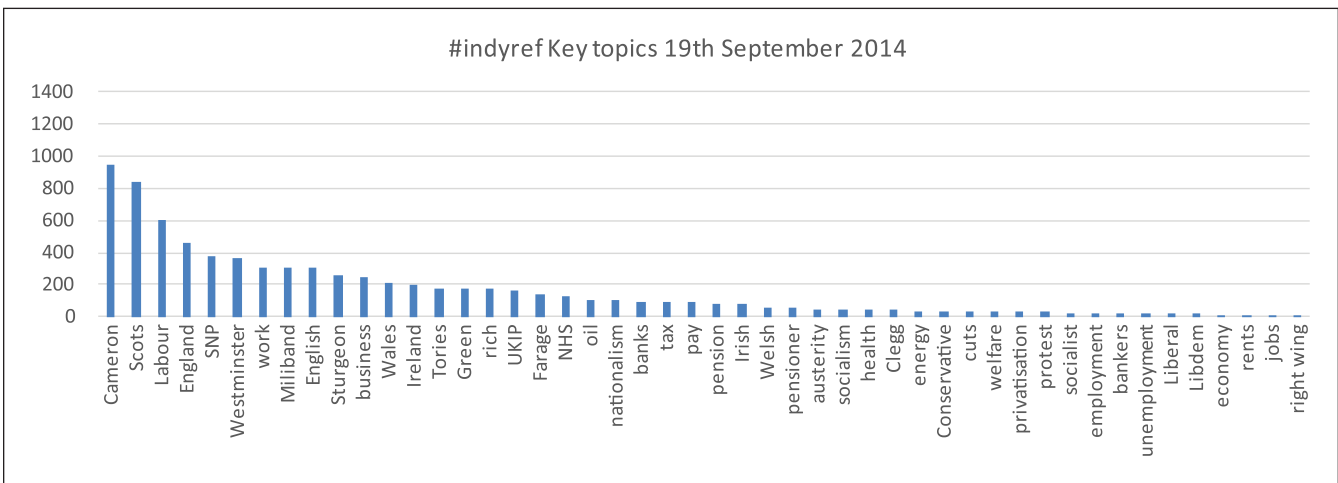

Figure 2. Drawn from filtering of 28,611 tweets – Generic terms Scotland; Independence; Referendum; Scottish excluded from this data as high responses skewed presentation – frequencies <10 also excluded.
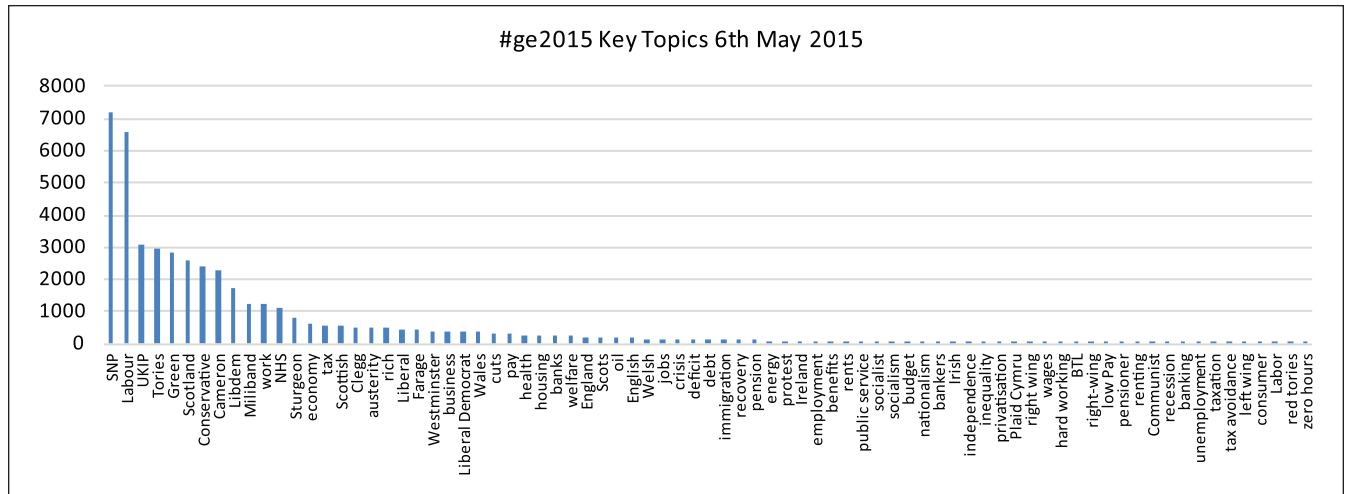

Figure 3. Drawn from filtering of 28,466 tweets – Generic terms Scotland; Independence; Referendum; Scottish excluded from this data as high responses skewed presentation – frequencies <10 also excluded.

Figure 4. Drawn from filtering of 53,421 tweets – frequencies <10 are excluded.

take issue with Westminster policies, which is something we will explore further.

One term where a relatively high frequency came as a bit of a surprise was 'rich'. As can be identified from all three data selections, this term featured prominently. Unlike the keywords discussed above, anti-wealth rhetoric was not an especially prominent feature of the referendum debate in the wider public and mass media spheres. Our initial assumption, however, was that this might possibly be suggestive of some sort of nascent unease or discontent with 'the rich' or, indeed, an association between the current UK political establishment and its leading politicians with a privileged elite. An important consideration here is that it may also be recognised that this is potentially the most ambiguous of our keyword search terms, given its possible range of associations. On examining the raw tweets, however, it appears that our initial assumptions seemed plausible, albeit with a few caveats. Across the various dates for both #indyref and #ge2015, we had captured a few 'Richards'. In addition, and particularly with reference to the #indyref samples taken in the days prior to the referendum, a significant number of the incidences of this term were related to the notion of an independent Scotland being a potentially 'rich' or 'oil rich' nation. However, a number of these tweets also referred to the referendum as representing a challenge to Westminster, and particularly the Conservatives, as representatives of the wealthy, with numerous post poll tweets (see Figure 3) in particular indicating that the 'No' vote had represented a victory of the rich over the poor. As discussed below, an even greater predominance of such sentiments evident among the #ge2015 tweets consolidated our view that this may be an emerging political issue in Scotland and across the United Kingdom as a whole.

With respect to the above, the incidence of unusual or unexpected factors emerging from this level of analysis provide indicators that point towards areas that can be pursued to explore, potentially, unrecognised or at least under

acknowledged social currents. Moreover, this highlights one of the principal advantages of this type of research suggested at the outset, in terms of affording social researchers the capacity to tap into a very large sample of public discourse easily, cheaply, frequently and at short notice.

Evidently, returning to the data, what is also interesting are those keywords that were expected to feature prominently but which did not. We may have expected housing, immigration, welfare and benefits to feature highly in discussions but, as above, this was not the case despite these issues being highly prominent in United Kingdom wide political discourse. This may indicate that these factors have less resonance in the Scottish context.

The UK general election (7 May 2015)

Turning to the presentation of the same set of keywords on the broader national forum, debating the UK general election, in Figures 4 to 6, once more, key terms relating to the main parties and figures feature highly. However, as above, we were somewhat surprised that the SNP/Scotland factor appeared to be prominent, even in light of the media's highlighting of Scotland during the campaign. In fact, when viewing this output, we checked to ensure that this had not been produced by a technical or procedural flaw. Having ruled this out, this outcome perhaps seemed to underscore just how successful the Conservatives' strategy of focusing on constitutional matters – and the 'spectre' of rule by a 'tartan terror' dominated SNP/Labour coalition – had played with the public in the period surrounding the poll, given that this seemed to overshadow discussion of key socio-economic factors, at least among our sample of Twitter posters. Conversely, the highlighting of 'the deficit' as an area of central political debate in the media was not significantly reflected in our data, albeit this may have been overshadowed by the constitutional focus in the closing stages of the campaign when our sample was drawn.
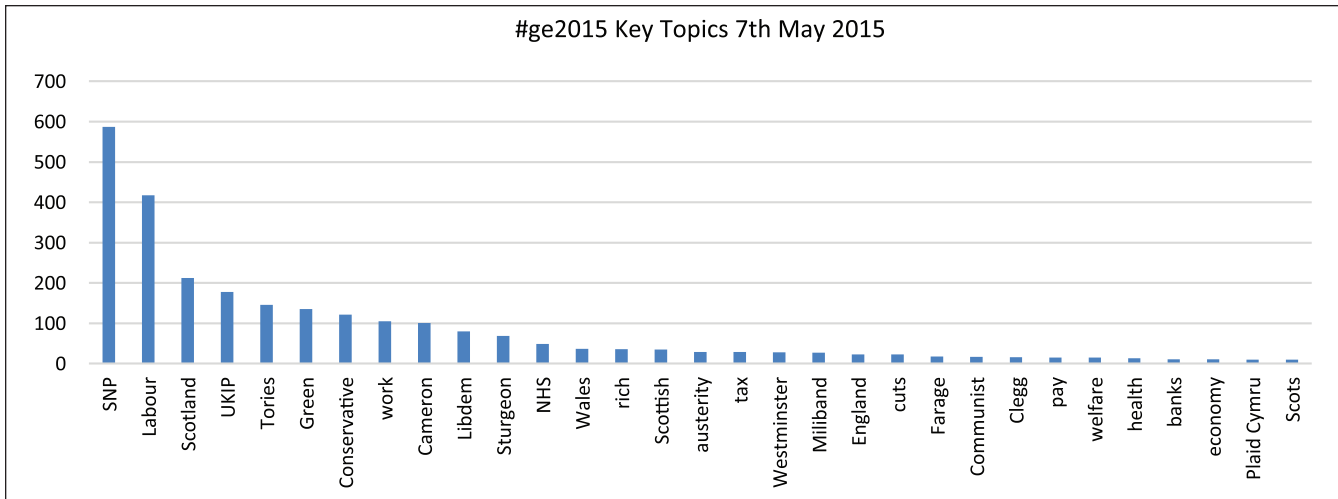
Figure 5. Drawn from filtering of 2959[2] tweets – frequencies <10 are excluded.
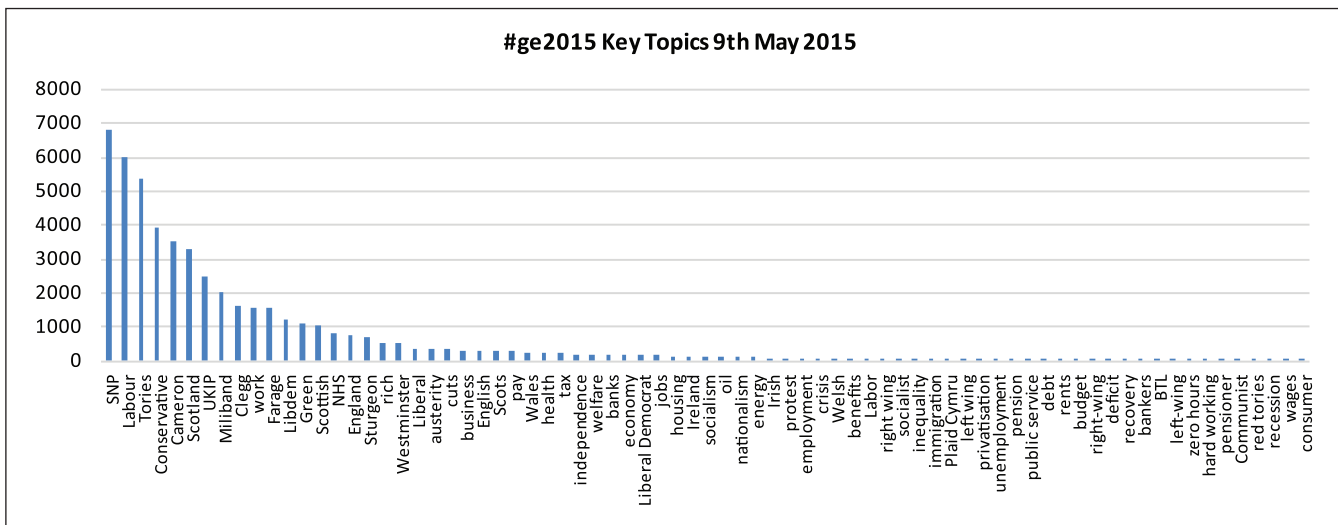

Figure 6. Drawn from filtering of 66,260 tweets – frequencies <10 are excluded.

Interestingly, turning to other key socio-economic terms, as the figures illustrate, the overall spread of keyword frequencies was not vastly out of line with those collected from #indyref at the time of the referendum, perhaps reinforcing the notion advanced by Keating (2011), above, that the concerns of the Scottish electorate are not so radically different from those of their English counterparts. Notably, however, austerity, welfare, benefits and housing issues were somewhat more prominent and, perhaps unsurprisingly, oil was much less so. Finally, it might also be noted that the 'rich' term also appears prominently here, further reinforcing the impression that there may possibly be something interesting happening that is not as clearly reflected in mainstream political and media debate as it is in the minds of the (at least social media savvy) public. With respect to the #ge2015 data, as intimated, antipathy towards the rich and growing social inequality appeared more pronounced in the general election

tweets than with the referendum samples, with a majority of posters depicting the election as a contest between social classes, and with a number depicting the winners as partisan representatives of the comfortable middle and upper classes. On reflection, while possibly contentious, it might even be suggested that the apparent frustrations and discontents displayed by posters around issues of socio-economic inequality and the concomitant rise of a privileged elite at the time of both the referendum and general election may reflect the seeming resurgence of left leaning sentiment that has been evidenced by the election of Jeremy Corbyn as leader of the Labour Party.

Overall, with respect to this ongoing project, at present we are developing techniques with the aim of uncovering some of the more detailed and subtle implications of the data we have collected and the social indicators that appear to have significance thus far.

Conclusion

This article is intended to convey both a sense of the growing opportunities as well as difficulties inherent in engaging in this particular type of cross-disciplinary work, principally in terms of bridging disciplinary boundaries, accommodating to radically different concerns, methodological approaches, skills and techniques, disciplinary lexicons and so on. The article also attempts to offer a flavour of the latter through these reflections on the research process and presentation of preliminary findings from a current collaborative work in progress.

At present our empirical study is progressing in tandem with the development of technical procedures aimed at facilitating and refining our methods, as a key aim of our overall project, as well as deeper excavation of the captured data in a dialectical process of trial and error. On this point, as alluded to above, each stage in this process has regularly involved a good deal of discussion and, at times, false starts before a procedure emerges that is both technically viable and, crucially, is also of social scientific utility. Generally, for both social and computer scientists, there can be some considerable potential for frustration here, as the gulf between disciplinary perspectives generates scenario's where a good deal of tactful debate and patience on both sides is required to move things forward in a positive direction. As also suggested, in my view, the difficulties that this kind of work presents entails that the building of mutually respectful relationships and understanding necessary for achieving gains cannot be accomplished overnight, but can be a painstaking process that in my view may ultimately depend upon sound interpersonal affiliations being established.

There are also other prodigious challenges and anxieties inherent for researchers in reaching beyond one's respective disciplinary 'comfort zone' in projects of this nature, where there is a constant awareness that failure to find mutual understanding and a way through may result in a project stalling and/or failing to meet its objectives while, in this new territory, the option of falling back on tried and tested methods or previously well-honed solutions may not be available or viable. Conversely, however, there is also a strong element of satisfaction in those 'eureka moments' when an impasse is resolved and an exciting solution emerges. Our project has, and continues to be, a roller coaster ride between these polarities. However, at present, our goals in terms of both the specific empirical study and the development of new methods appears to be travelling in the right direction. I would conclude by suggesting that this type of work is worth the journey while it may be both bountiful and, inevitably, essential work in terms of advancing social scientific enquiry in future.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

1. Maximum devolution within the United Kingdom, understood as effective 'home rule', with a few 'umbrella' responsibilities such as defence and foreign policy being shared with the rest of the United Kingdom.
2. We experienced technical issues on 7 May which lead to a much smaller sample of tweets being collected on that day. However, we decided to include them with the evident caveat that they are less reliable than data produced from the larger samples on 6 and 8 May.

References

Anderson B (1983) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Bone J and O'Reilly K (2010) No place called home: The causes and social consequences of the UK housing 'Bubble'. *British Journal of Sociology* 61(2): 231–255.

Brownlow C and O'Dell L (2002) Ethical issues for qualitative research in online communities. *Disability & Society* 17: 685–694.

Bruns A, Burgess J, Crawford K, et al. (2012) *#qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods*. Brisbane, QLD, Australia: ARC Centre of Excellence for Creative Industries and Innovation.

Connolly R (2014) Dataveillance in the workplace: Privacy threat or market imperative? Gupta M (ed.) *Handbook of Research on Emerging Developments in Data Privacy*. Hershey, PA: IGI Global, pp. 70–85.

Convery I and Cox D (2012) A review of research ethics in internet based research. *Practitioner Research in Higher Education* 6(1): 50–57.

Curtice J (2014) It all depends on your perspective: Economic perceptions and the demography of voting in the Scottish Independence Referendum. *Fraser of Allander Economic Commentary* 38(2): 147–152.

Davidson N (2014) A Scottish watershed. *New Left Review* 89(2): 5–26.

Davidson N, McCafferty P and Miller D (2010) *Neoliberal Scotland: Class and Society in a Stateless Nation*. Newcastle upon Tyne: Cambridge Scholars.

Guy M, Earle P, Ostrum C, et al. (2010) Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies. In: Cohen PR, Adams NM and Berthold MR (eds) *Advances in Intelligent Data Analysis IX: Lecture Notes in Computer Science*, vol. 6065. Berlin, Heidelberg: Springer, pp. 42–53.

Halford S, Pope C and Weal MJ (2013) Digital futures? Sociological challenges and opportunities in the emergent semantic web. *Sociology* 47(1): 173–189.

Hughes AL and Palen L (2009) Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management* 6(3–4): 248–260.

Johnston RJ and Pattie CJ (1989) A nation dividing? Economic well being, voter response and the changing electoral geography of Great Britain. *Parliamentary Affairs* 42(1): 37–57.

Keating MJ (2010) 'The Strange Death of Unionist Scotland'. *Government and Opposition* 45(3): 365–385.

Keating M (2011) *Scotland and Independence*. Montréal, QC, Canada: The Federal Idea.

Kornhauser W (1959) *The Politics of Mass Society*. Glencoe: Free Press.

Lipinski-Harten M and Tafarodi RW (2012) A comparison of conversational quality in online and face-to-face first encounters. *Journal of Language and Social Psychology* 31: 331–341.

Liu SB (2009) Informing design of next generation social media to support crisis–related grassroots heritage. In: *PhD colloquium of the 6th international ISCRAM conference*, Gothenburg, 10–13 May.

McEwen N (2010) State welfare nationalism: The territorial impact of welfare state development in Scotland. *Regional & Federal Studies* 12(1): 66–90.

Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

Mendoza M, Poblete B and Castillo C (2010) Twitter under crisis: Can we trust what we RT? In: *SOMA '10: 1st workshop on social media analytics*, Washington, DC, 25 July.

Palen L, Starbird K, Vieweg S, et al. (2010) Twitter-based information distribution during the 2009 Red River Valley flood threat. *Bulletin of the American Society for Information Science and Technology* 36(5): 13–17. Available at: http://dx.doi.org/10.1002/bult.2010.1720360505

Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.

Smout TC (1994) Perspectives on the Scottish identity. *Scottish Affairs* 6: 101–113.

Snee H (2013) Making ethical decisions in an online context: Reflections on using blogs to explore narratives of experience. *Methodological Innovations Online* 8(2): 52–67.

Tinati R, Halford S, Carr L, et al. (2014) Big data: Methodological challenges and approaches for sociological analysis. *Sociology* 48(4): 663–681.

Web sources

http://www.scotsman.com/news/politics/top-stories/pm-confident-of-winning-battle-over-referendum-1-2194055

http://www.thenews.com.pk/Todays-News-9-332323-Rich-privileged-and-in-power

http://www.independent.co.uk/news/uk/scottish-independence/scottish-independence-full-text-of-david-camerons-no-going-back-speech-9735902.html

Author biographies

John Bone's research interests are in political economy, social and biosocial theory, well-being and social media research. In terms of theoretical work, he has a specific interest in the way in which human neurobiological capacities and propensities shape and delimit the dialectical relationship between individual and society, with implications for understanding social organisation, social interaction and well-being. From a more empirical perspective, he is interested in the way in which the contemporary trajectory of global development and neo-liberal economic policy is affecting organisations, individuals and communities in terms of identities, working lives, well-being, social polarisation and social cohesion.

Chukwuemeka David Emele is a computing scientist whose principal research field is multi-agent systems, with a focus on computational models of policies (or norms), argumentation and complex systems. His interests include machine learning, visualisation and data mining – how can we learn from patterns that we have noticed in systems (or in a dataset) and use such knowledge to aid decision making and Digital Economy and Open Data. Dr. Emele is particularly keen on exploring how social media and digital channels such as Twitter, Facebook, LinkedIn, Instagram and so on are changing the way we do things including elections, e-commerce, transportation, planning future cities, crowdsourcing, healthcare and cloud computing.

Adeniyi Abdul is a software engineer with experience in designing, developing and testing software across a variety of platforms. His research interest are in cloud computing, grip computing, machine learning using predictive modelling, agent-based scheduling and augmented reality.

George M. Coghill is SICSA Chair and Director of Research in School of Natural and Computing Sciences at the University of Aberdeen. He has nearly 30 years of research experience in the area of Model-based Reasoning, and rich experience in Computational Intelligence. He has contributed immensely by applying his expertise to various fields as diverse as process engineering and system biology. He is currently leading the System Modelling research group at the University of Aberdeen, and the group is engaged in the theoretical foundation, pragmatics, and use of tools for analysis and modelling abstract, conceptual and real-world system.

Wei Pang's research interests are in bio-inspired (especially immune-inspired) data mining and machine learning, especially as applied in social, biological, health, and environmental sciences. He's also interested in qualitative and semi-quantitative modelling and simulation of complex systems. His substantive research is in social network analysis, systems biology, anomaly detection in environmental sensor data and clustering analysis in prenatal data with missing values.