THE STATE OF THE ART: A LITERATURE REVIEW OF SOCIAL MEDIA INTELLIGENCE CAPABILITIES FOR COUNTER-TERRORISM

Jamie Bartlett

Carl Miller

November 2013

DEMOS

PARTNERS CREDITS

Supported by Public Safety Canada

Published by Demos 2013

Third Floor
Magdalen House
136 Tooley Street
London SE1 2TU

T 0845 458 5949
F 020 7367 4201

hello@demos.co.uk
www.demos.co.uk

BACKGROUND, RESEARCH DESIGN AND STRUCTURE

This paper is a review of how information and insight can be drawn from open social media sources. It focuses on the specific research techniques that have emerged, the capabilities they provide, the possible insights they offer, and the ethical and legal questions they raise. The relevance and value of these techniques are considered for the purpose of maintaining public safety by preventing, pursuing, protecting and preparing against terrorism.

Social media research has emerged as a practice, but it is not yet a coherent academic discipline or distinctive intelligence tradecraft. It is neither a distinct area of study, nor driven by a united research community. It is conducted across the public, private and academic sectors, spanning disciplines from the computer sciences and ethnography to advertising and brand management. Its aims range from understanding the topography of social networks comprised of millions of individuals to the deep, textured knowledge of the social worlds of individuals and small groups.

As such, techniques and approaches often reflect specific disciplinary traditions and rarely refer to those found elsewhere. Social media research is also fragmented by platform. Twitter already has a distinct nascent discipline, driven by free access to millions of tweets, an easily available Application Programming Interface (API), and fewer concerns with privacy and intrusion. Since 2008, 'Twitterology' has grown from a handful to hundreds of research papers, covering everything from topic identification to event detection and political forecasting.[1]

Research on Facebook – either about it or using it – has struggled in the face of technological difficulties in acquiring the data and Facebook's corporate orientation towards advertising rather than research. As of 2011, there were 42 peer reviewed journal articles about Facebook research, although this is growing quickly.[2]

The overall aim of this review is to describe the emerging contours of social media research: to codify the capabilities that have emerged and the opportunities they have created, and the risks and

hurdles that they must commonly face and surmount – methodological, legal and ethical – in order to usefully contribute towards countering terrorism in a way that is publicly supported.

A semi-systematic literature review methodology was employed. The purpose of the review was defined with an explicit statement of focus and further refined following a series of short meetings with a small group of likely consumers of this paper in March 2013.[3] On the basis of these meetings, studies were consistently included and excluded on the basis of the following criteria:

- The paper had to have used 'social media' as the primary or sole focus. Following a relatively consensual definition, social media was defined as both the technology and the use of a varied category of internet services inspired by 'the participatory web' or 'web 2.0'[4] which enable users to create and share digital content, whether textual, audio or video.[5]

- Where possible, the paper should have been published within the last three years, especially if it was related to newly emerging techniques or areas of rapid development.

- The paper had to suggest a method, capability, technique or usage considered by the researchers to be broadly relevant to the purposes of countering terrorism (with particular stress on the prevention of violent extremism and the broader task of understanding the phenomenon).

- Notably, given the diversity of the literature, no judgment of relevance was made *a priori* regarding research design, methodology or technique.

All studies found to meet the criteria above were incorporated and considered. These were found through a variety of techniques:

- Scholarly searches using keywords relevant and terms relevant to the purpose as defined above.

- Experts in the field were approached for the purpose of bibliographical recommendation.

- Publication outlets and research centers known to conduct relevant work were searched according to the relevancy criteria defined above.

- The proceedings and published proceedings of conferences that were subject matter-relevant to the paper were gathered.

In total, 112 papers were analysed, and their key contribution to the question to counter-terrorism capability was identified and recorded. Further notes were incidentally made on location, date, method, and overall thesis. The results were synthesised into categories of capability, as set out below.

Caveats

It is notable that very little social media research found was directly related to counter-terrorism work, but much had, when extrapolated, implications for counter-terrorism. Therefore, we have provided reflections where necessary based on our research work and judgment. We have made this clear throughout.

Finally, there is a large difference between current capabilities, and what are published capabilities. We do not have access to a great deal of use-cases – including novel techniques, novel applications of techniques or substantive findings – that are either in development or extant but unpublished. Academic peer-reviewed publishing can take anywhere from six months to two years, while many commercial capabilities are proprietary. Furthermore, much social media research is conducted either by or on behalf of the social media platforms themselves and never made public. The growing distance between development and publishing, the increasing role of proprietary methodologies and private sector ownership and exploitation of focal datasets are important characteristics of the social media research environment.

This paper does not consider techniques to acquire or use closed or private information, or methods by which detailed profiles of individuals can be built. These techniques are more readily situated within the gamut of secret intelligence work rather than research, and beyond the scope of the authors' expertise.

Although the contents of this paper are relevant to a wide variety of agencies, this work was commissioned by Public Safety Canada, so subsequently there is a focus on Canadian specific examples of case studies throughout.

Structure

The paper is structured as follows:

Part 1 is an overview of social media use, focused on how it is used by groups of interest to those involved in counter-terrorism.

Part 2 provides an introduction to the key approaches of social media intelligence (henceforth 'SOCMINT') for counter-terrorism.

Part 3 sets out a series of SOCMINT techniques. For each technique a series of capabilities and insights are considered, the validity and reliability of the method is considered, and how they might be applied to counter-terrorism work explored. The techniques examined in this manner are:

- Machine learning & Natural Language Processing
- Event detection
- Predictive analytics (notably non machine learning based)
- Network Analysis
- Manual analysis / 'netnography'
- Solicited / 'crowd sourced' insight

Part 4 outlines a number of important legal and ethical considerations when undertaking SOCMINT work.

PART 1: OVERVIEW OF SOCIAL MEDIA USE AND BEHAVIOUR

Trends in use

Every month, 1.2 billion people now use internet sites, apps, blogs and forums to post, share and view content. Loosely grouped as a new, 'social' media, these platforms provide the means for the way in which the internet is increasingly being used: to participate, to create and to share information about ourselves and our friends, our likes and dislikes, movements, thoughts and transactions. Although social media can be 'closed' (meaning not publically viewable) the underlying infrastructure, philosophy and logic of social media is that it is to varying extents 'open': viewable by certain publics as defined by the user, the user's network of relationships, or anyone.

The most well-known are Facebook (the largest, with over a billion users), YouTube and Twitter. However, a much more diverse (linguistically, culturally and functionally) family of platforms span social bookmarking, micromedia, niche networks, video aggregation and social curation. The specialist business network LinkedIn has 200 million users, the Russian-language VK network 190 million, and the Chinese QQ network 700 million. Platforms such as Reddit (which reported 400 million unique visitors in 2012) and Tumblr, which has just reached 100 million blogs, can support extremely niche communities based on mutual interest. For example, it is estimated that there are hundreds of English language pro-eating disorder blogs and platforms.

Social media accounts for an increasing proportion of time spent on-line. On an average day, Facebook users spend 9.7 billion minutes on the site, share 4 billion pieces of content a day and upload 250 million photos. Facebook is further integrated with 7 million websites and apps.

Trends in Canadian use

80 per cent of Canadians are connected to the internet and spend on average 17.2 hours online every week, which includes watching

an average of one hour of online videos every day (80 per cent of it on YouTube). Furthermore, 80 per cent of all Canadians that use mobile devices have a Smartphone. General internet usage is higher among Anglophone than Francophone Canadians. This gap, slowly diminishing, is more pronounced in older age groups, and non-existent in the 18-34 age group.

Canadians are among the earliest and most enthusiastic adopters of social media. One study found that around half of all Canadians have a social media profile, projected to grow by another two million to 18.5 million by 2014, although other research puts this figure higher. The average Canadian spends almost 8 hours a day on social media each week, a figure that is growing. Future trends of social media use, global and Canadian, are unclear. Facebook membership uptake, which has driven social media uptake and accounts for a large proportion of total social media use is slowing in Western Europe and Northern American with signs of market saturation. A recent survey about social media use in Canada found that 44 per cent of users said they were less enthusiastic about social media than they were a year earlier, which could be an early indicator of the onset of social media fatigue.

Age, unsurprisingly, strongly characterises social media use in Canada: 18 to 25 year olds spend almost twice as much time on social media network sites as those over 55. (Nonetheless, every age group in Canada is above the worldwide average). In the younger age groups, male and female users are roughly similarly represented, but in older age cohorts women tend to use social media in significantly higher numbers than men. Anglophone Canadians seem more active social media users than Francophone ones, although the difference is relatively minor.

In terms of use, 61 per cent of Canadians use social media to stay connected with friends and family, 39 per cent to stay connected with professional contacts, and 55 per cent to stay up to date on news and general items of interest. In any typical month, 44 per cent update their status on one platform or another, 38 per cent post photos, 17 per cent post videos, and 14 per cent share their GPS location on a social media network.

Like in many other countries, Facebook is the most popular social media platform, although the precise numbers, especially when concerned with actual use rather than formal membership, are controversial. A recent AskCanadians survey found that 73 per cent of Canadian social media users were on Facebook, 35 per cent use YouTube, 21 per cent use LinkedIn and Twitter, 19 per cent use Google+, 5.3 per cent use Pinterest and Flickr, 3.3 per cent use Tumblr, 3 per cent use Instagram, 2.4 per cent use MySpace, and 1.7 per cent use Foursquare. An analysis of visits to social media sites in Canada undertaken by Hitwise in January 2012 found that Facebook received 63 per cent of all social media website visits. YouTube came second with 22 per cent of visits, with all other sites receiving less than 2 per cent of visits.

This paper's analysis of Canadian Facebook users (using Facebook's Advertising function) revealed that 17,863,080 people are on Facebook in Canada, of which 45 per cent are men and 55 per cent women. Thirty-two per cent are under 25; 31 per cent are between 26 and 39, 28 per cent are between 40 and 59, and 9 per cent are 60 or over.

As of February 2013, Twitter is the ninth most popular website in Canada. The most followed Canadian political account is Nidal Joad's (@pm2025), a political figure from Quebec, who unsuccessfully ran as an independent in the 2003 Quebec provincial elections, and a commentator on the Arab Spring (who is currently particularly focused on Syria). After the US, India and the UK, Canada has the fourth highest number of LinkedIn accounts (6,514,327, which is 4 per cent of all 163.5 million LinkedIn users). LinkedIn use correlates with business centres like Montreal (where one in four people use the networking site) and Toronto (where one in five use the site).

The use of social media by extremist and terrorist groups

Extremist and terrorist groups use the internet for a myriad of purposes, including the dissemination of propaganda, the recruitment of new members and the development of operational planning. Online activity is a critical part of almost every national security investigation. By 1999 nearly all known terrorist groups

had established a presence on the internet. Nevertheless, the extent to which the internet affects radicalisation into violence is contested.

The picture is less clear in respect to social media specifically. Detailed empirical research into how extremist and terrorist groups have reacted to the rise of social media is limited, but markedly growing. The shift from text-heavy traditional websites to social networks built around interactive forums allowing the sharing of mixed media (often where leaders posted stories and steered discussions) came in the mid 2000s. Recent analysis suggests that since the late 2000s activity has increasingly shifted to social media platforms.

According to Aaron Zelin, 'it is only a matter of time before terrorists use Twitter and Instagram as part of ongoing operations'. Zelin charts an increase in activists using Twitter as a tool of communication, motivated perhaps by the need to appeal to a younger demographic that prefers this medium. A MEMRI report has documented the use of Instagram by al-Qaeda leaders to share images and quotes, glorify imprisoned fighters, and disseminate images of dead 'martyrs'. The international prominence (and highly cited case study, although recently discontinued) of al-Shabaab's Twitter accounts has been used by the group to present a professional and united image, obtain support from the Somalia diaspora, offer dialogue with supporters and rebut critics in real time.

Similarly, far right analysts have agreed that while right-wing extremist communities have had an online presence for years through dedicated websites, there has been increased activity on social media in recent years. According to O'Callaghan, social media is used especially by neo-Nazi groups to redirect users to content hosted on external websites. Indeed, it is this ability to share news items, original articles and essays and tribute videos that is perhaps key.

From right-wing to al-Qaeda inspired extremism, social media may 'lower the bar' for participation, making the involvement of low-

level, semi-radicalised or previously disengaged individuals a new feature of transnational extremist conversations and movements. Although extremist forums are still dominated by Arabic language content, the opposite is true of Twitter feeds. According to Michel Juneau-Katsuya, social media is playing a growing role in reaching out to vulnerable young people: 'a means of privileged communication…which excludes their family and isolates them with others who sympathise with their cause and [who] think in a similar fashion.'

Social media platforms are believed to have helped extend the reach of hate groups more broadly. According to Christopher Wolf, the online world 'has become a technology embraced by racists, anti-Semites, homophobes and bigots of all kinds to spread their message of hate'. Holocaust deniers, the Identity Church, KKK Members, neo-Nazis and racist skinhead groups are all believed to be particularly active. Anders Breivik, for example, drew much inspiration and impetus from his interactions online, including from the new 'counter-Jihad movement' – an international collection of Islamophobic bloggers, which, according to Hope not Hate, comprise over 200 organisations worldwide.

No research was found that comprehensively measures the amount of hate speech that occurs online. The Simon Wiesenthal Centre's annual Digital Terror and Hate Report from 2012 is based on 15,000 'problematic' websites, social networks, forums, online games and apps. They believe this has seen an increase of around 3,500 problematic outlets since 2010. Similarly, the International Network Against Cyberhate has argued that over recent years 'the amount of cyber hate has grown to enormous proportions', with 'Islam, Jews, lesbians and gays, blacks, Roma, liberals' and 'left-wingers' representing the main targets of online abuse. It is of note that of all the referrals made by the UK's counter-terrorism internet Referral Unit (which seeks material that glorifies terrorism and asks for its removal from internet service providers), Facebook, Twitter, Blogger and/or Blogspot were most frequently identified as the hosts of the problematic, referred material.

Social media and law enforcement

More generally, social media use is affecting other types of law enforcement activity: criminal organisations and gangs exploit the internet and social media. Well-organised and longstanding groups have stable social media presences, usually used for advertising their organisation, or in some cases 'cyber banging' – levying threats against rival groups, or individuals. Indeed, in November 2012, the British Justice Secretary announced a crackdown on the use of social media by criminals to intimidate witnesses. Additionally, the amount of personal information posted on social media has been shown to influence the risks of individuals to burglary.

Social media is also of growing relevance for public disorder policing. In both the August 2011 riots in the UK, and in Vancouver following the Stanley Cup the same year, a common tendency was identified. During the early stages of disorder, participants and uninvolved observers recorded and shared information about the event. As the disorder increased, information describing the apparent impunity of the rioters, visibly shared on social media, may have escalated the disorder further. In the aftermath similar themes of a united community coming together to condemn the riots and organise a clear up were seen both in London and Vancouver. Moreover, the confusion of modified digital content, rumour and hearsay were noted as having slowed down the policing procedures following both riots.

It is important to note that both the technological infrastructure of social media and the way that this infrastructure is used changes quickly. Research suggests that users have increasingly become aware of the privacy risks and reacted by placing more of their social media content onto higher privacy settings with more restricted possible readerships. A study of 1.4 million Facebook users in New York showed that in 15 months between 2010 and 2011 users who kept key attributes on their profiles private rose from 12 per cent to 33 per cent. Users are taking more care to actively manage their online accounts; figures for deleting comments, friends, and tags from photos are all increasing

according to a recent Pew survey. Equally, the nature of the terror threat is likely to evolve in future, and could include groups that are expert in various counter-surveillance and 'sous-veillance' techniques (monitoring agents of the state). For example, a new platform which sprang to prominence during the Occupy Protest movement in 2011 was Vibe; an app for smartphones which allows users to send short anonymous messages to users within a pre-defined geographical proximity which are automatically deleted after a pre-determined period of time.

PART 2: AN INTRODUCTION TO SOCIAL MEDIA INTELLIGENCE

SOCMINT covers a wide range of applications, techniques and capabilities available through the collection and use of social media data. The term was first coined by the authors in a 2012 report, *#Intelligence*.[6]

Some analysts have suggested SOCMINT to be a branch of open source intelligence (OSINT), which has been defined as 'information that is publically available and can be lawfully obtained by request, purchase or observation'.[7] SOCMINT does not easily fit into the category of open or secret intelligence. SOCMINT is defined, not by the openness of the information on which it is based, but by its existence on a social media platform. As either open or closed intelligence, SOCMINT requires very specific considerations of validity and interpretation.

This paper does not discuss closed or secret SOCMINT, which by definition would require access to communications which are not publicly available. Instead, this paper focuses only on open SOCMINT as define above. We believe this type of SOCMINT is potentially a useful and important part of counter-terrorism and public safety efforts. In the United States, OSINT is considered to be of considerable and increasing value, covering commercial, procurement and trade data, expert opinion data and a variety of types of 'gray' literature produced by the private sector, government agencies and academics.[8] The US Committee on Homeland Security considers OSINT to be a tool that federal state and local law enforcement agencies should use to develop timely, relevant and actionable intelligence; especially as a supplement to classified data.[9]

As with any intelligence, SOCMINT should improve decision-making by reducing ignorance.[10] There are many different types of open SOCMINT. We believe the most significant, capable of reducing ignorance and improving decision-making for the purposes of counter-terrorism are:

- Natural language processing – a branch of artificial intelligence involving the computational analysis (often using machine learning methods) of 'natural' language as it is found on social media.

- Event detection – The statistical detection analysis of social media streams to identify offline 'events', whether natural, political, cultural, commercial or emergency to provide situational awareness, especially in dynamic and rapidly developing contexts.

- Data mining and predictive analytics - The statistical analysis or 'mining' of unprecedentedly large ('big data') datasets, including social media and other 'big' or open data sets (such as Census data, crime, health, environmental and transport data), to find the dynamics, interactions, feedback loops and causal connections between them.

- Social network analysis: the application of a suite of mathematical techniques to find the structure and topography of the social networks found on social media. These networks are then subjected to analysis, which can identify a range of implications and conclusions (including predictive ones) on the basis of the characteristics of the network structure and type.

- Manual analysis /'netnography': drawn from qualitative sociology and ethnography, this is a broad collection of manual approaches to collecting and analysing data concerning social media data. It often aims for depth over breadth in order to reveal and untangle the hidden, obscured, overlooked or contingent social significances, meanings and subjectivities experienced by individuals on social media.

- Solicited / 'crowd sourced' insight: refers to the emerging practice of a number of public and private agencies to use social media to ask citizens or social media users for information directly.

PART 3: TECHNIQUES AND METHODS

We critically discuss the state of the art in each category of open SOCMINT. Each section considers capabilities generally and, where possible, specifically for the purpose of countering terrorism. First, we present the main ways of accessing and analysing data sets.

Social media data collection and retrieval

It is possible to manually collect social media data in a number of ways - copying, screen grabbing, note-taking, and saving web-pages. However, where large volumes of data are involved, the most appropriate method is to collect the data automatically. This is done through connection to a platform's 'Application Programming Interface' ('API').[11]

The API is a portal that acts as a technical gatekeeper of the data held by the social media platform. They allow an external computer system to communicate with and acquire information from the social media platform. Each API differs in the rules they set for this access: the type of data they allow researchers to access, the format they produce this data in, and the quantities that they produce it in.

Some APIs can deliver historical data stretching back months or years, whilst others only deliver very recent content. Some deliver a random selection of social media data taken from the platform, whilst others deliver data that matches the queries – usually keywords selected by the analyst - stipulated by the researcher. In general, all APIs produce data in a consistent, 'structured' format, and in large quantities. Facebook and Twitter's APIs also produce 'meta-data' – information about the data itself, including information about the user, their followers, and profile. Along with Facebook and Twitter, most major social media platforms allow API access for researchers in some form.

There are seven types of API access to Facebook data, most of which have been designed for app makers.[12] The Facebook API relevant to social media research is the 'Graph API', which can be directly accessed online with Facebook's Graph API Explorer, or via Facebook-approved third party commercial re-sellers of data, like DiscoverText or DataSift. The difference between Graph API

Explorer and a third party front end is that the third party software is designed to gather large amounts of data via the Explorer and present them in a way that is conducive to detailed analysis. There is no additional functionality, and Facebook retains all control over what kind and how much data can be collected.

Graph API allows posted text, events, or URLs, plus any comments on posts to be accessed, along with metadata on user information, including gender and location.[13] It operates like database interrogation software: a user asks it for information using the relevant coding language, and Explorer finds where on Facebook that information is stored (i.e. the web address) and returns the information. Facebook API is sometimes considered opaque by researchers that use it. There is no detailed written record of how it works, which potentially introduces bias to any data gathered through the API.

Access to all Facebook data is predicated on the user's settings and who has agreed to share information with them. Facebook's privacy structures are complex - potentially, any single user can have a distinct privacy setting for every piece of data they share. They can, for example, decide that only their 'close' friends (a user-defined group of 20 people) can see a single post, all posts, or posts on a particular page. API searches only return data that is public, and fails to quantify the information that has remained uncollected due to privacy restrictions. This is a significant weakness in methodological terms.

The most prolific and heavily researched provider of social media data for research is Twitter. Twitter has been operating since 2006 and its 200 million active users have posted over 170 billion tweets since the platform was first created. As a platform experiencing extremely rapid growth, the demography – geography, language, age and wealth – of its users is constantly changing. Major studies, whilst struggling to keep pace with this rapid change, have found that over 100 languages are regularly used on Twitter. English accounts for around half of all tweets, with other popular languages being Mandarin Chinese, Japanese, Portuguese, Indonesian, and Spanish (accounting together for around 40 per cent of tweets.)

These languages are geographically spread, with concentrations in Europe, the United States, Latin America and South East Asia. China, with 35 million users, has more users than any other country.[14]

Twitter has three different APIs that are available to researchers. Twitter's 'search' API returns a collection of relevant tweets matching a specified query (word match) from an index that extends up to roughly a week in the past. Its 'filter' API streams tweets that contain one of a number of keywords in real time. Its 'sample' API returns a small number (approximately 1 per cent) of all public tweets in real time.

Each of these APIs (consistent with the vast majority of all social media platform APIs) is constrained by the amount of data they will return. Twitter provides three volume limits. A public, free 'spritzer' account is able to collect one per cent of the total daily number of tweets. White-listed research accounts may collect 10 per cent of the total daily number of tweets (known informally as 'the garden hose') while the commercially available 'firehose' collects 100 per cent of daily tweets. With daily tweet volumes averaging roughly 400 million, many papers do not find any of these restrictions to be limiting to the number of tweets they collect (or need) on any particular topic.

Each of Twitter's APIs produces up to 33 pieces of meta-data with each tweet (far exceeding in length the content of the tweet), including (if it exists) the geo-location of the author (expressed as longitude-latitude coordinates), their profile's free-form text location, their time-zone, the number of followers they have, the number of tweets they've sent, the tweet's creation date, the author's URL, the creation date of the account, even the author's wallpaper on their Twitter homepage.[15] A full list of available data in every tweet is included in the annex.

To set up a stream or search to collect the data, it is typical to create a user interface which is built around the underlying API provided by Twitter. The API is a series of http 'end points' that return data according to the parameters that are provided with the request.[16]

One of the key advantages of acquiring data via a social media platform's API is the consistent, 'structured' nature of the data that is provided. This advantage becomes important when gathering high volumes of data from dynamic platforms such as Twitter. Alongside direct API access, a number of licensed providers make available raw data to multiple APIs. These include DataSift, Gnip and DiscoverText.

Web scrapers and crawlers

For the purpose of this overview, 'scrapers', 'crawlers', 'spiders' and 'bots' are all automated programs which are used to find and catalogue information stored on websites. This is typically achieved through transforming website data (usually expressed in a language called 'HyperText Markup Language', or html) into structured datasets.

A basic crawler of this type is usually a relatively simple piece of code that employs a technique to collect and process data on different websites. Programmers can use regular expressions (or 'regex') to define a pattern (for example, any acronym involving at least 3 letters) to enable the crawler to execute a particular predefined action when a match is identified on a webpage. This allows the crawler to copy or index specific information from any page on the World Wide Web to which it is directed. These and many other associated techniques are subject to constant development.

Someone with little experience can in a short space of time build their own bespoke crawler using only freely available programs and tutorials. A very basic crawler can be created with only a few lines of code on platforms such as Scraperwiki or using programming languages such as Python, Java, or PHP. These crawlers can be built very quickly and cheaply, and increasingly the code is open source.[17] Despite their relative simplicity, basic crawlers and the vastly more complex crawlers employed by commercial and public organizations have the potential to unlock data on the way communities interact, the information they seek, and the sources of information they use.

Information retrieval

In general, information retrieval refers to a body of techniques employed to identify those documents in a large collection that are relevant to a specific information need. Patterns and 'objects' within documents are often found by rules-based algorithms, which allow documents to be ranked according to their relevance.[18] This is a rapidly developing area of work.[19] Retrieval techniques are designed to allow for more powerful meaning based searches. For example, running a search for conversations related to Jihad and filtering the subsequent results based on clustered groups of identical or near identical material highlights those retrieved items that include new material.[20]

Search engines still do not always effectively search social media content, even though it might be highly relevant. For example, photos with a relevant title or geo-location often contain little textual narrative making them difficult to search for. Improving the accuracy of social media searching is also an emergent field of considerable interest. Current developments focus on 'similarity clustering', which facilitates the identification of relevant clusters of social media data considered to be importantly similar, either in their content, or when or where they the content was posted.

According to Tim Berners-Lee, automated search techniques require further development. Information embedded in a document is still not easy to find. Berners-Lee believes the Web will evolve from a 'web of documents' to a 'web of data' – underpinned by Universal Resource Identifiers (URIs) to allow for a consistent reference. Simple Protocol and Resource Description Framework Query Language will allow this semantic web to be searched.[21]

Machine learning/Natural Language Processing

Introduction

Natural Language Processing (henceforth, NLP) is a long-established sub-field of artificial intelligence research. It combines approaches developed in the fields of computer science, applied mathematics and linguistics. It 'teaches' algorithms to automatically detect the meaning of 'natural' language, such as that found on social media. These algorithmic models look for statistical correlations between the language used and the meaning expressed on the basis of previous examples provided by human analysts, and then, building on this, automatically (and therefore at great speed) make decisions about the meaning of additional, unseen messages. NLP is increasingly and necessarily used as an analytical 'window' into datasets of social media communication that are too large to be manually analysed.

This training of NLP algorithms – a technique called Machine Learning – is conducted through a process called 'mark up'. Messages are presented to the analyst via an interface. The analyst reads each message, and decides which of a number of pre-assigned categories of meaning it best fits. After the analyst has made a decision, they click on the most relevant tweet and it is 'annotated', becoming associated with that category. The NLP algorithm then appraises the linguistic attributes, that, depending on the specific algorithm, often includes words (or unigrams), collection of words (such as bigrams and trigrams), grammar, word order or emoticons – that correlate strongly with each category. These measured correlations provide the criteria for which the algorithm then proceeds to make additional automatic judgments about which category additional (and un-annotated) pieces of social media data best fit into.

The statistical nature of this approach renders it notionally applicable to any language where there is a statistical correlation between language use and meaning. NLP programmes vary in the way they make their decisions: some place more weight on specific words, others on structural or grammatical features.

The operational opportunity of NLP for countering terrorism is to use these algorithmic models as 'classifiers'. Classifiers are applied NLP algorithms that are trained to categorise each piece of social media data – each post or tweet – into one of a small number of pre-defined categories. The earliest and most widely applied example of this technology is 'sentiment analysis', wherein classifiers make decisions on whether a piece of social media data is broadly positive or negative in tone. However, the kinds of distinctions that a classifier can make are arbitrary, and can be determined by the analyst and the context.

The performance of NLP classifiers is often quantified by comparing a body of automatically classified data against a 'gold standard' set of human classifications. On this measure, their accuracy – the ability of the NLP algorithm to classify any given message the same way a human would – varies considerably. There are many scenarios where 90 per cent accuracy would be expected. However, an accuracy of around 70-80 per cent in a three-way classification task would often be considered excellent.

Classifiers are sensitive to the specific vocabulary seen in the data used to train them. The best classifiers are therefore also highly bespoke and trained on a specific conversation at a specific time to understand context-specific significance and meaning. As language use and meaning constantly change, the classifier must be re-trained to maintain these levels of accuracy. The more generic and expansive the use of any NLP classifier, the more likely that it will misunderstand language use, misclassify text and return inaccurate results.

In many situations, the performance of these classifiers is sufficient to produce robust aggregate findings, even when the accuracy of any given singular classification is quite low. This arises because the data sets are sufficiently large that even relatively inaccurate individual estimates lead to an accurate assessment of the overall trend. Assessing when this technology tends to work well and when it does not is an area of active research.

A key area of active research is in the reduction of the time, effort and cost required to train and maintain an NLP classifier. It is typically very expensive to produce the labeled training data that these supervised machine learning algorithms require. In complex tasks, it would not be unusual for this to take multiple person-months of effort. The novel introduction of an information-theoretic technique called 'active learning' is beginning to allow classifiers to be built much more rapidly and cheaply – often in a matter of hours, and sufficiently quickly to meet changing operational requirements prompted by rapidly shifting situations and contexts.[22]

There are three emerging uses of NLP that we considered particularly relevant. The first is to classify tweets into categories other than positive, negative and neutral: such as urgent, calm, violent or pacific.[23] The second is to use NLP to dramatically reduce the amount of data that an analyst must sift through in order to find messages of relevance or interest. In this respect, classifiers can also be 'tuned' to perform at high precision (only highlighting messages very likely to be of interest) or high recall (highlighting all messages conceivably of interest).[24] This form of relevancy filtering is sometimes known as 'disambiguation'.[25] The third is to create layers of multiple NLP classifiers to make architectures capable of making more sophisticated decisions.

Attitudinal data/sentiment analysis

Perhaps the largest body of attitudinal research on social media has focused on the use of NLP to understand citizen attitudes on Twitter. This research has been driven by the view – implicit or explicit in most of the research papers - that attitudinal datasets on Twitter are different to those gathered and understood by conventional attitudinal research – interviewing, traditional polling or focus grouping. This is because the size of available data sets are huge,[26] naturalistic (meaning that they are not exposed to observation bias) and constantly refreshing in real time. Furthermore, because of the increasing ease of data access and dramatic reductions in computing costs, these data sets are notably more analysable.

Harnessing social media datasets of this kind stands to have a transformative impact on our ability to understand sentiments and attitudes. However, no published output has yet been able to understand attitudes on social media using methods that satisfy the conventional methodological standards of attitudinal research in the social sciences, or the evidentiary standards of public policy decision-makers. There remain a number of methodological problems.

Perhaps the most important methodological challenge is sampling. Twitter's API delivers tweets that match a series of search terms. If searches are subjected to Boolean operators similar to search engines, searching for 'Canada' returns tweets that contain 'Canada' in either the username of the tweeter, or the text of any tweet. A good sample on Twitter must have both high recall and high precision. 'Recall' is the proportion of possibly relevant tweets on the whole of Twitter that any sampling strategy can find and collect. 'Precision' is the proportion of relevant tweets that any sampling strategy selects.

A high recall, high precision sampling strategy (measured together as a single mean score called F1) is therefore comprehensive, but does not contain many tweets that are irrelevant. Arriving at a series of search terms that return a good sample is far from easy. Language-use on Twitter is constantly changing, and subject to viral, short-term transformations in the way language is mobilised to describe any particular topic. Trending topics, '#' tags and memes change the landscape of language in ways that cannot be anticipated, but can crucially undermine the ability of any body of search terms to return a reasonably comprehensive and precise sample.

Current conventional sampling strategies on Twitter construct 'incidental' samples using search terms that are arbitrarily derived. They do not necessarily return precise, comprehensive samples, and certainly do not do so in a way that is transparent, systematic or reliable. Furthermore, it is becoming clear that the way Twitter is

used poses first-order challenges to discerning genuine attitudes of people. Indeed, a lot of Twitter data does not actually include any attitude at all – is often just general broadcasting or link shares.[27]

It should be noted that work on sentiment analysis has begun to drawn upon other methodologies beyond NLP. Some studies have drawn upon network analytics (see below) and specifically theories of emotional contagion to inform sentiment analysis algorithms.[28]

Latent insight

NLP works on the premise that certain features of a text can be statistically analysed to provide probabilistic measures of meaning. One rapidly emerging area of study in NLP is to run classifiers on large training data sets in order to generically reveal 'latent' meaning, especially features about the author – age, gender and other demographics – which are not revealed explicitly in the text or captured by the social media platform and provided as meta-data, but which can be probabilistically determined by the underlying structures of language use. The development of latent NLP classifiers is an area of intensive investigation by university research institutes and social media platforms themselves.[29]

One university, for example, has developed a fairly accurate gender estimator, based on around 50 characteristics that tend to be associated with male or female language use (there is a free test interface available; http://stealthserver01.ece.stevens-tech.edu/index), trained against a large data set of emails. On Twitter, the main way to spot gender is by user name: which is possible using an automated system and is correct around 85 per cent of time. One research team, using NLP on just the tweet content, achieved 65 per cent accuracy, achieving 92 per cent accuracy when further meta-data was included).[30]

Information about a users' location is another important area of work. Around 2-3 per cent of tweets include latitudinal and longitudinal meta-data, allowing tweets to be located very precisely. A larger body of tweets is possibly resoluble to a location through the use of additional meta-data. An academic study found that approximately 15 per cent of tweets can be geo-located to a specific

city, based on the cross-referencing of other fields of meta-data: location (of the account, recorded as free-form text) and time zone (structured).[31] Another study demonstrated that resolving place names to longitude/latitude coordinates have been shown to increase the ability to geo-locate social media documents by a factor of 2.5.[32]

Other techniques have been applied to determine latent details from online data. A 2013 report by Berkeley and Cambridge universities found that it was possible to deduce personal information about people through an analysis of their 'likes' on Facebook, including sexual orientations, ethnicity, religious and political views, and some personality traits. [33] The model correctly discriminated between homosexual and heterosexual men in 88 per cent of cases, African Americans and Caucasian Americans in 95 per cent of cases, and between Democrat and Republican in 85 per cent of cases. Drugs use was successfully predicted from likes in 65 per cent of the time.

However, personality traits – such as conscientiousness or happiness – were less easily deduced. It appears that simple demographic data – especially dichotomous variables – are more amenable to this type of analysis, but behaviour less so.[34] Similar studies have found that personality can be predicted with a reasonable degree of accuracy on the basis of web browsing, music collection, or friend numbers and networks.[35]

The use of automated language recognition to spot certain types of 'risky' behaviour or criminal intent is also a developing application of the NLP. Some linguists argue that certain structural, underlying features of a sentence and related syntax can be broadly correlated to general behaviour types, such as anger or frustration, subconscious states of mind.[36] We are not able to locate any academic peer reviewed papers that test this hypothesis in detail. A series of recent reports about 'predictive policing' are not based on social media data sets but the use of existing crime data and other data sets to predict crime hot spots.[37]

However, based on our experience training classifiers, the extent to which this might be amenable to practical application will depend on the existence of training data – the information fed into the classifier to allow it to spot patterns. There is no reason that a classifier with enough training data would not be able to spot language use known to be correlated with certain behaviours (for example criminal activity); and assess the confidence with which it had made these decisions on the basis of quantifiable values. This would allow an analyst to effectively target further investigation.

Indeed, in a well-documented case in 2012, Facebook worked with the police to apprehend a middle age man talking about sex with a 13 year old girl and trying to meet her. The details of the case are not clear, but it appears likely a machine learning algorithm would have been used.[38] (It is of note that Facebook has access to a far larger data set than independent researchers, as their data set will include all private accounts and messages, where behaviour of this type is *prima facie* more likely to occur.)

Event detection and situational awareness

Introduction

Social media can be viewed as an information platform that contain 'events', defined as discrete incidents of, for example, a political, cultural, commercial or emergency nature. These events may be intrinsic to social media, such as a particular type of conversation or trend; conversely, they might be indicators or proxies of events that have occurred offline.[39] During the 2011 Egyptian revolution, for instance, 32,000 new groups were formed and 14,000 new pages created on Facebook in Egypt.[40]

Event detection technology attempts to identify and characterize events by observing the profiles of word or phrase usage over time - usually anomalous spikes of certain words and phrases together – that indicate that an event may be occurring. Broadly there are two styles of positively identifying an event; query drive and data driven. Query driven event detection is akin to waiting for a fairly specific 'thing' to happen and report that it has when enough evidence that matches the event 'query' has been recorded over a

short enough time period. A purely data driven event detection system has no preconceived notion what type of event it is meant to report. Rather it has a preconceived notion of what an event 'looks like' in terms of the statistical characteristics that are elicited in the text stream.

Situational awareness via Twitter

Twitter is by far the platform of greatest interest in terms of event detection.[41] Of all the uses of event detection technology, building situational awareness of rapidly developing and chaotic events – especially emergencies – is perhaps of most clear application to counter-terrorism. Emerging events are often reported on Twitter (and often spike shortly thereafter as 'Twitcidents') as they occur.[42]

Social media users (especially Twitter users) can play a number of different roles in exchanging information that can detect events. They can generate information about events first-hand. They can request information about events. They can 'broker' information by responding to information requests, checking information and adding additional information from other sources and they can propagate information that already exists within the social media stream.

Multimedia content embedded on social media platforms can add useful information – audio, pictures and video – which can help to characterise events. One crucial area of development has been to combine different types of social media information across different platforms. One study used YouTube, Flickr and Facebook, including pictures, user-provided annotations and automatically generated information to detect events and identify their type and scale.[43]

Due to the user generated nature of on social media there is a pervasive concern with the quality and credibility of information being exchanged. Given the immediacy and easy propagation of information on Twitter, plausible misinformation has the potential to spread very quickly, causing a statistically significant change in the text stream. Confirming the validity of the positive system response is a crucial step before any action is to be taken on the basis of that output. A vital requirement of event detection

technology is the ability to verify the credibility of information announcing or describing an event. Some promising work has been done to statistically identify first-hand tweets that report a previously unseen story; however it is unclear how that system would perform with the relatively small amounts of data available in an emergency scenario.[44]

Generally speaking, untrue stories tend to be short lived due to some Twitter users acting as information brokers, who actively check and debunk information that they have found to be false or unreliable. One study, for instance, found that false rumours are questioned more on Twitter by other users than true reportage.[45] Using topically agnostic features from the tweet stream itself has shown an accuracy of about 85 per cent on the detection of newsworthy events.[46]

One 2010 paper, 'Twitter under crisis', asked whether it was possible to determine 'confirmed truth' tweets from 'false rumour' tweets in the immediate aftermath of the Chilean earthquake. The research found that Twitter did tend toward weeding out falsehoods: 95 per cent of 'confirmed truth' tweets, were 'affirmed' by users, while only 0.3 per cent were 'denied'. By contrast, around 50 per cent of false rumour tweets were 'denied' by users. Nevertheless, the research may have suffered a number of flaws. It is known, for example, that the mainstream media still drives traffic – and that tweets including URL links tend to be most re-tweeted, suggesting that many users may have simply been following mainstream media sources. Moreover, in emergency response, there tends to be more URL shares (approximately 40 per cent compared to an average of 25 per cent) and fewer 'conversation threads'.[47]

One important factor, especially important for situational awareness, is the ability to identify the geo-spatial characteristics of an event. Many of the techniques described above to infer the location data of social media content are also used in the field of event detection.

However, the reliability of event detection and situational awareness techniques may be context or even event specific. It appears especially useful in emergency response where a large number of people have a motive to produce accurate information. By contrast, one recent (unpublished) thesis analysed the extent to which useful real time information about English Defence League protests could be gleaned from Twitter.[48]

In the build-up to three demonstrations for which data were collected in 2011, most tweets were negative; and very few were geo-located to the event venue. A very large number of tweets were re-tweets (49 per cent compared to 24 per cent during a control period), and on further analysis, a significant proportion of the re-tweets were negative, inaccurate rumours. Moreover, a very large proportion of tweets (50 per cent) came from a very small number (5 per cent) of – usually negative – commentators.[49] The recent crowd-sourced effort to positively ID the suspects in the recent Boston terror attacks on Reddit were also less successful – although it is not clear whether and how information gained through the exercise was of use or value to the police.

This is only one of a number of difficulties relating to the validity and reliability of data sets. There are now, for example, systematic, highly organised operations to create fake reviews,[50] although other researchers are using natural language processing to determine fake reviews from real ones – including verifying an IP address to determine frequency.[51] Of course, at the very large scale, data can be widely skewed by automated information bots.[52] Facebook recently revealed that seven per cent of its overall users are fakes and dupes.[53]

The validity of large scale data sets partly relies not on the fact that every single data point will be taken as accurate, but that when aggregated and combined, large scale data sets can produce valid and robust results – or at least results more robust than any single, even expert, observation. This is the principle that 'the wisdom of the crowds' produces more accurate descriptions than any single observer when certain conditions are met: diversity of opinion, independence, decentralisation, and aggregation.[54]Social media, as

a social network, does not always meet these conditions. One recent study of 140,000 Facebook profiles looked at the first three months of use and found that new members were closely monitoring and adapting to how their friends behaved, suggesting that social learning and social comparison are important influences on behaviour.[55] The 2011 London riots were widely discussed – and perhaps partly organised – via social media networks. It does not appear that Twitter was able to 'dispel' misinformation quickly. Indeed, rumours spread rapidly, and although some disagreement was found, they were within different, sealed networks.[56]

Predictive analytics

Introduction

Broadly, there is a growing sense that the 'big data' revolution – the ability of humans to make measurements about the world, record, store and analyse them in unprecedented quantities – is making new kinds of predictions possible.[57] This, 'predictive analytics', brings together a wide range of intellectual and technical infrastructure, from modeling and machine learning to statistics and psychology.[58]

The explosion of social media is part of the big data revolution. More and more of our intellectual, cultural and social activity is being captured in digital form on social media platforms. It represents the 'datafication' of social life. It renders social life measurable and recordable.

Interest in harnessing these social-digital traces by predictive analytics was sparked by a paper published in 2009 by Hal Varian, Google's chief economist, who argued that Google search terms can sometimes predict real world behaviour (such as searches relating to jobs preceding and predicting unemployment figures). Since then, there has been an interest in applying predictive analytics to social media datasets to predict a range of social behaviours and phenomena, from election results, to box office numbers, to stock market trends.[59]

One recent article reviewed the areas and ways social media information can be used to make predictions. It identified

commercial sales, movie box offices, information dissemination, election results, and macroeconomic developments as being particularly amenable to predictions on the basis of social media data. The paper concludes that while none of these metrics seem to have sufficient predictive power by themselves, they can work quite well when combined.[60]

Politics

Correlations of social media sentiment are also subject to predictive analytics. Eric Siegel, in *Predictive Analytics – The Power to Predict Who Will Click, Buy, Lie, or Die* explains how Obama's predictive analytics team predicted those 'swing voters' who had the greatest likelihood of being influenced to vote for Obama. They used data from Twitter and Facebook to predict which people were strong influencers of the swing voters, and targeted them, not the swing voters themselves (an example of the 'Persuasion Effect'). That approach is at the very cutting edge of predictive analytics today, largely because of its development and successful deployment within American electoral campaigns.[61]

Research from Tweetminster during the 2010 UK general election found that volume of mentions on Twitter – at the national but not candidate level – is associated with overall election results.[62] A similar study was undertaken in the German Federal election of 2009, although these results were critically analysed by other researchers, who found that the relative frequency of mentions of political parties had no predictive power, and argued the results were contingent on the arbitrary choices of the researchers.[63] On replication, the researchers included the online group the Pirate Party, which the original research team failed to do, and found that it secured the greatest share of Twitter mentions and yet failed to secure a single seat.

Zeynep Tufekci has made the argument that in the recent Arab Spring uprisings, Facebook and Twitter have played a crucial role in a 'collective action / information cascade' that created a momentum that helped transform groups of dissidents acting alone into a widespread revolution, applying Malcolm Gladwell's idea of a 'tipping point' to social media.[64] However, Malcolm Gladwell

himself is sceptical of the idea that social media influences 'real' revolutions, positing that the tools with which people within revolutionary events communicate are not in themselves important or interesting.[65]

Health

One area that has received a lot of attention is the use of Twitter data to understand the spread of infectious disease, known as 'public health monitoring'. Some analysts believe this will become a vital part of spotting and tracking health trends. Google search terms for flu symptoms – although not technically social media – are already found to identify outbreaks faster than doctor's records.[66]

One 2012 paper found that, based on an analysis of 2.5 million geo-tagged tweets, online ties to an infected person increased the likelihood of infection, particularly where geographically proximate (due, of course, to the increased incidence of physical transmission). The analysis was based on 6,237 'geo-active users', who were tweeting with geo-location enabled Twitter accounts more than 100 times per month. While the results are fairly obvious; the researchers suggest that these findings demonstrate that Twitter analysis can help model global epidemics.[67]

This study was undertaken through the analysis of only open, geo-located Twitter accounts, and using machine learning as outlined above to identify tweets which appear indicative of flu. Some papers have suggested ways to geo-spatially characterising social media, combining text features (e.g. tags as a prominent example of short, unstructured text labels) with spatial knowledge (e.g. geo-tags, coordinates of images and videos).[68]

Crime detection

Most of the work that has been done on criminal incident prediction relies primarily on historical crime records, geospatial information and demographic information, and does not take in to account the rich and rapidly expanding social media context that surrounds many incidents of interest. One paper presents a preliminary investigation of Twitter-based criminal incident prediction. The

model analysed the tweets of a single feed (Charlottesville, Virginia news agency), but believed an adapted version could potentially be used for a larger-scale analysis of tweets.

Rather than keyword volume analysis and sentiment analysis, which are unhelpful to predict discrete criminal incidents that are not mentioned ahead of time, the authors used NLP techniques to extract the semantic event content of the tweets. They then identified event-based topics and used these to predict future occurrences of criminal incidents. The performance of the predictive model that was built was evaluated using ground-truth criminal incident data, and compared favourably to the baseline model that used traditional time series methods to study hit-and-run incidents per day.[69]

Raytheon's Rapid Information Overlay Technology (RIOT) was widely reported in UK media in early 2013 as signaling a new type of social media mining that would be of interest to security services for predictive purposes.[70] Based on a video posted on the *Guardian*'s website, Raytheon's principle investigator suggested that RIOT could be used to closely track a person's life, down to their daily gym schedule. It is not clear precisely what techniques or functionalities are used in RIOT.

The problem of prediction

Nate Silver has described how big data driven predictions can succeed but also fail in his recent book *The Signal and the Noise*. He argues that 'prediction in the era of big data is not going very well'. Silver attributes this to our propensity for finding random patterns in noise, and suggests the amount of noise is increasing relative to the amount of signal, resulting in enormous data sets producing lots of correlative patterns which are ultimately neither causal, accurate, nor valuable. [71]

Correlations, without either sound theoretic underpinning or explanation, are common in many branches of social media research. Incidental correlations of this kind – such as an apparently strong relationship identified in one Facebook study between high levels of intelligence and the liking of 'Curly Fries' -

add little insight or value.[72] Silver's suggestion is that we use more Bayesian mathematics: probabilistic predictions of real world events based on clear expressions of prior beliefs, rather than statistical significance tests or dichotomous predictions. Interestingly, as Silver points out, big companies spend less time modeling than running hundreds of data experiments to test their hypotheses.[73]

Indeed, predictive analytics have rarely been used experimentally, and then tested in reality. All studies cited in this paper have been based on a 'retrospective fit' – where researchers, acting with the benefit of hindsight, construct post-event analyses of pre-event data. This is obviously ill-suited to many of the operational needs of counter-terrorism agencies, who have to make time-dependent forecasts in chaotic, unpredictable and fundamentally uncertain circumstances.

Network Analysis

Introduction

Social network analysis (henceforth, SNA) is at its root a sociological and mathematical discipline that pre-dates the internet and social media. It aims to discern the nature, intensity, and frequency of social ties, often as complex networks. Its premise is that social ties influence individuals, their beliefs, and behaviours and experiences. By measuring, mapping, describing and modeling these ties, social network analyst attempt to explain and indeed predict the behavior the individuals that comprise the network.

In order to derive SOCMINT, SNA can be conducted on different types of datasets of online activities, including blogs, news stories, discussion boards, and social media sites. It attempts to measure and understand those 'network links' both explicitly and implicitly created by the features of the platform, and how the platform is used. These include: formal members of particular movements; followers of Twitter feeds; members of forums; communities of interests; and interactions between users. Sometimes these are referred to as 'explicit' or 'implicit' communities depending on the degree of involvement in or commitment to the group in question.

Explicit communities tend to refer to groups where members have made an explicit decision to join a blog-ring, group, or network, while implicit communities refer to the existence of broader interactions such as linking, or commenting.

The network characteristics of digital information is often measured using a technique, pre-dating social media, to map the relationship between internet websites. Crawlers follow hypertext links from one site to the next, recording whether and how each links to others. In general terms, a crawler tends to start from a small number of carefully selected seed sites and then continuously find the links from there to other sites. There is a range of methodologies for effective crawl 'depth' in research (meaning how many steps should be crawled from the seed sites). The design of the data capture and selection of seed sites for a web crawl stems from the perspective created by the research question.[74]

Borgatti, in his famous analysis of 200 Conservative bloggers, used a crawl depth of two in order to balance the risk of a sample being too shallow - a significant risk when the crawl depth is one - with the risk of a sample being too deep, introducing a high degree of noise, or mapping neighboring issue networks. Indeed, a crawl depth of two was also used in a number of recent studies concerning a variety of political networks, including pro-gun control networks and the mapping of the Norwegian Blogosphere.[75]

Linkages can be split into three classes: content, structure and usage. The identification of these kinds of linkages allows the user to build a dataset of online activities, whether they take place on blogs, news sites, discussion boards, or social media sites.[76] Once the data is gathered it can be used for a number of purposes, ranging from the analysis of how many individuals are engaged in a specific activity online, to the assessment of information flows and influence in complex systems.[77] Indeed, it is possible to map even covert networks using data available from news sources on the World Wide Web, as shown by researchers including Valdis Krebs.[78]

Once the data is gathered it can be used for a number of purposes, ranging from estimating how many individuals are engaged in a specific activity online to understanding the flow of information and influence in complex systems.[79]

Typical activities include:

- Tracking increases in content produced about a specific issue or location.
- Tracking the spread of a specific piece of information.
- Tracking the sharing of information between individuals.
- Understanding the complex structures created by the behaviour of individuals which influences the information other users receive, and subsequently the behaviours those communities adopt.

There are a number of mathematical techniques that can be used to understand and describe social networks expressed in social media data. Centrality analysis is a well-established technique that describes position of any given node in a network to other nodes through three measures.

First, the 'degree' - or how many links a node has to other nodes. High degree nodes are sometimes described as 'Achilles' heels' within a network, and often represent 'leaders' or 'influencers' of various types.

Second, 'betweeness' measures how far a node lies between other nodes in a network. Nodes with high betweeness are sometimes considered the gate-keepers between different, tighter clusters within a looser network, and act as important channels of influence between them.

Third, 'closeness' is measured as the sum of the length between a node and the other nodes (low scores means it may be hard to communicate).

Another commonly used type of analysis is known as 'community analysis', which is designed to identify social groups in a network. A 'community' is identified where members of a group have a higher density of links than with those outside a group; the specific limits of a group can be accurately divined by the establishment of a 'threshold' which determines at what point a node is part of a group.

Followers and affiliates – understanding the loose network

Several groups likely to be of interest host open social media accounts. The network of open account followers of Al-Shabbab – easily downloadable – is highly diverse, with many likely to be curious spectators, journalists, researchers or analysts as well as supporters and ideologically aligned fellow-travellers. There is not, as far as we know, any technique for making these distinctions, beyond careful and manual reconstruction of each individual.

It is for this reason that the free, automated analytics tools of Twitter followers, such as *mapmyfollowers* or *Twittalyser* can be highly misleading. When making policy decisions, it is often good practice to use systems that are transparent about the way influence or 'influencers' has been calculated. Some more detailed academic studies have been able to rank users' influence on a specific subject area, rather than more simplistic measures such as engagement and follower numbers. By analysing their followers, and whom they follow, on a thematic basis, it is possible to observe clustered relationships based on particular themes.[80]

A recent paper published an analysis of the 3,542 followers of 12 White Nationalist Twitter accounts, and a random sample of each of their 200 most recent tweets. It was found that around 50 per cent did not overtly subscribe to White Nationalist ideology (although these were not removed in the final analysis). The researchers created their own compound measure of network influence. Rather than using the existing centrality measures detailed above, they measured 'influence' through the combination of two metrics, 'engagement' - the amount of times a user's tweets resulted in a response of any kind (for example in the form of a reply, retweet or

favourite); and 'exposure' by the number of times a user responded to other people's tweets in the same way.

As noted above, it is possible to create new measures of understanding networks in this way through Twitter. This research found the most 'engaged' also tended to be the most overt supporters of White Nationalism: 93 of the 100 most engaged accounts were also those who appeared the most overt supporters of White Nationalism.[81] When the same method was applied against anarchist accounts, results were less clear-cut. The data set was less coherent, and there was less covert self-identification as anarchist; as a result, top engagement was not as closely correlated with active involvement.[82]

This research also found a large number of link shares. The authors argued that by identifying the key content among radical and extreme groups, through the links that they share, it would be possible to understand in greater detail their ideology. Furthermore, the paper recommended that targeting shared links for disruption through terms of service violation reporting would be an important potential counter-extremism tactic.

A similar study of White Nationalist Twitter accounts started with a core or seed set of accounts. In this case, social ties were measured through the phenomenon of one user mentioning another through the use of a Twitter handle (@<username>) in a tweet; in this context, reciprocal mentions can be considered a dialogue. A network was then created based on these collected reciprocities. A 'highly stable' network based on significant dialogue was thereby mapped out, and an analysis undertaken on common keywords employed, in order to determine the common themes of communication within the community. The research team then conducted analysis on the location of members, with some success.

The research found that the dialogue network tended to be among people from the same country, in contrast to a simple network of followers (although this allowed the researchers to identify a user acting as an English language translator for a Swedish nationalist group). However, the work has a caveat, recognising the likely

incompleteness of datasets it used, presumably based on the imperfect choices made when selecting the initial core seed accounts.[83]

Relational networks

An important application of SNA for SOCMINT has been to estimate the strength of relationships on the basis of different forms of social media activity, as a precursor to more detailed understanding, (often through predictive modeling), of how strong and loose networks influence individual behavior.[84]

In very general terms, Facebook evidences high rates of connection: 92 per cent of users are connected by four degrees of separation. It also appears to support the 'weak ties' argument – many users relate intensively and constantly with a small group of friends (10-20) but follow more loosely a larger group (150-200). Other research has found that there is a tendency to join a community based on both the number of friends they have in the community, but also, importantly, by how those friends are connected to one another.[85] Jure Leskovec's work has been especially prominent in this regard, applying social psychological principles to machine learning to interpret and predict the positive and negative feedback on the Epinions, Slashdot and Wikipedia platforms.[86]

In one study of White Nationalist blog sites, researchers manually identified a series of seed blogs and blog-rings that used White Nationalist terminology in their title or description.[87] The subsequent crawl of linked sites identified 28 groups, comprising 820 individual bloggers, and found more blog groups than were listed on hate speech directories, suggesting that the use of blog spiders was as useful in identifying groups as understanding them.

The researchers noted how many users were active on each blog, and further extracted all profile information about each blogger, including their user ID, date of birth, city, and real name. This profile information is self-reported, and thus of dubious accuracy, although the 'blog creation date' is automatically recorded by the host site and is therefore a reliable source. In this case the 'link'

analysed to understand the network was whether the blogger had subscribed to another blog-ring.

The researchers found the community was well connected internally – the average number of links that would link any member of the network to any other was only 2.89. The clustering coefficient – a measure of how tightly grouped together nodes in a network are – was 0.37, characteristic of a small, nascent community. This dataset, similar to many others found on social media, was subject to a 'power-law distribution': the top bloggers had many more direct links than other members of the community.[88] Those with high in-degree scores might be usefully subject to further detailed analysis of their blogs to understand better how ideologies, motivations and messages are formed and spread throughout the group.

A strong sub-discipline within social media research has sought to identify why and how people are influenced (either ideologically or behaviorally) on social media. This has often been driven by a desire to market and reach 'key influencers' within a particular field.[89] A 2010 study built a series of 'models of influence' that strongly predicted on a probabilistic basis whether a user would perform an action on social media on the basis of their position on a social graph.[90]

Understanding a social network based on flow of content

A significant component of an individual's information environment is the relationships that affect how they acquire information and knowledge.[91] Research in this field has focused on how information (and therefore influence) flows in multiple directions, and how it coordinates around hubs or focal points. Crawlers, API calls and network analysis are jointly being used to develop insight, locate influential individuals or communities of influence and understand the hubs around which these social media users coordinate.

For example, information flow, if represented by a directional network, allows influential users to be identified. Influential users can be discerned through a number of measures of 'centrality', as

discussed above. All such measures are useful in different ways; 'degree centrality', for example, shows those with the greatest number of connections.[92] An analyst using a combination of these measures and community detection algorithms can gain extensive insight into the interconnected communities which exist within the network, along with the level of influence of different members.

The combination of network analysis and sentiment analysis has been used in order to identify the focal points for specific communities within a wider trending topic or complex issues. For example, research into the use of Twitter during the 2009 Iranian election revealed quite a distinct set of different communities that were using the tag #Iranelection, and very different topics that 'trended' within them.[93]

Content sharing and diffusion

One of the most remarkable emerging areas of interest has centered around the analysis of what information is 'shared' - content that has been posted and then re-posted, re-tweeted or otherwise further disseminated by individual users. According to one analyst, more detailed understanding of what is being shared can provide insight into a groups' changing beliefs and views, and is the most interesting element of social media analysis, with greater promise than social media sentiment analysis.[94]

There are a number of free or cheap tools that provide simple data on the trends of link sharing. Seesmic, for instance, provides a number of useful tools for understanding of media consumption behaviours within communities. Cascade, a piece of software developed by the *New York Times,* shows who shares each story and when, in order to understand the structure of sharing. This is done by analysing the trajectory of bit.ly URL shortened links. This helps identify influencers (who shares the most and who drives subsequent traffic); and which variables appear to affect this.

Unedited, user generated content adds to the challenge as people often copy and paste entire articles or parts of articles into blog posts without providing a hyperlink to the source.[95] Inter-media have explored the nature of media consumption during the Arab

Spring by examining patterns of sharing. They found that some journalists were important in driving traffic to particular news stories, blogs and tweets, and became information brokers, aggregating, filtering and disseminating relevant content.[96]

Discussion

There has not yet, however, been a full and detailed study into the sociology of the phenomenon of sharing. The social significance of sharing content – news stories or otherwise – is little understood. Given this, any extrapolation from content sharing into the purposes or motivations of the sharers must be treated with care.[97]

This has important implications for understanding the research design of SOCMINT SNA. As with machine learning, determining the nature of a network depends partly on the initial decisions of an analyst in deciding what sort of link is important and selection of seed accounts/sites. One key, recurring theme in network analysis study is the extent to which a crawler captures everyone in a network. One of the key challenges in analysing a network is how and where to define the boundary of the network. There is rarely a simple boundary to a network, and the larger the networks, the less likely there is to be a clear boundary, as recognised by Malcolm Sparrow.[98] The case for embracing the concept of more fluid network of relations rather than conceiving of groups in a formal, rigid structure has been further reiterated by the 'Fijnaut Group' and others studying organised crime in the Netherlands.[99] Indeed, as Borgatti has argued: 'the choice of nodes should be dictated by the research question and one's explanatory theory', rather than arbitrary, inflexible conditions.[100]

Automated network analysis can produce both strategic and tactical insight, but only in the appropriate context. Real-time monitoring, which tracks shifts in the volume of terms and content produced around a specific issue, can have significant tactical value, for example through the provision of real-time intelligence regarding changes in locations being mentioned by groups seeking to create or exploit public disorder.

One of the most useful aspects of automated network analysis is the identification of information, groups and individuals. The rapid identification of the most engaged individuals in certain ideas is an extremely simple and cheap type of analysis, which can be done via access to APIs without any machine learning or NLP. However, this would still require a great deal of analytical review because of the lack of clarity about who these people are. On the strategic level, on the other hand, analysis can also be usefully done through big data sets which comprise content aggregated from a range of tactical sources and over a longer time period than (near) real-time. At this level, it is possible to analyse the wider information system through the fluctuations in volume flow, and in doing so identify users who have different influential roles within that system.

Netnography

Introduction

'Netnography' broadly refers to the application of ethnographic and qualitative sociological methodologies to the study of social media data. Consistent with the theoretical commitments of these disciplines, netnography usually avoids the quantification or numerical measurement of social media data and instead sees it as part of an individual's social and cultural life that is textured, complex and often only understandable when studied in depth. Netnography often practically takes the form of 'participant observation' – sustained contacts between the researcher and members of a digital community.

Forum ethnography

The careful study of behaviours within forums of 'communities of interest' is a potentially powerful way of gaining insight into attitude formation and behaviour. One recent detailed study of 126 pro-eating disorder sites, for example, found that there were very different types of platform being used: furthermore many had not been updated, while others were in daily use. The sites varied greatly in content and tone, in their use of images, and the nature of the disclaimers and terms of service governing their use. The study found that the type and format – whether a blog, website, social network – was directly related to the nature and type of content.

The researchers applied mature offline sociological social research methods: using contextual (type of site, regularity of updates, functions used) and conceptual (nature and categories of content) codes. This generated new insight into the cultural norms and social makeup of these sites.[101]

The kinds of social space that forums represent are changeable. Many chat room forums include several sub-forums, some of which are public and others private. For counter-terrorism purposes, and more generally the study of discussions based on socially problematic or stigmatised views, closed forums are often more valuable than open ones. One report concerning al-Qaeda forums found that 'it is not possible to have a rounded sense of what is taking place through only the public sites' – although such a level of access does offer useful insight into the 'zeitgeist' of the movements, including broad ideological shifts.[102]

Reconstructions of offline groups from social media

In some instances, very careful reconstruction of the social interactions contained within a forum can help researchers understand the specific membership and hierarchy of a group. In 2009, Strathclyde Police launched *Operation Access,* which used social networking sites such as Facebook to uncover criminal activity by identifying weapons carriers, especially in the context of urban gang memberships and inter-gang feuding. As part of the programme, police officers searched through images to find users who had posted pictures of themselves with weapons. The Superintendent in charge of the operation stated that as a result, 400 people were questioned.[103]

Similar work – collecting valuable evidence on criminal activity such as illegal gun ownership – has been undertaken successfully elsewhere.[104] It has been well established that detailed analysis of forums is a useful way to collect intelligence at both a strategic and operational level. For example, the Islamic Awakening forum may have been a useful source to calculate the affect on the movement of the controversy surrounding Hammami.[105]

Data verification problems

There are lots of examples of inaccurate information or misinformation being widely believed, even by subject specialists. Tweets by imposter accounts have been picked up, believed and reported on by major news outlets.[106] A number of Facebook studies have asked whether users tend to portray an accurate picture of themselves on the site, and a literature review of these publications suggests that Facebook profiles convey fairly accurate personality impressions of users. This may be because, unlike other online groups, people tend to become Facebook friends only after being offline friends.[107]

In many respects, determining the veracity of any single source is much the same as usual – and would require the same standards and methods applied in any human intelligence source: track record, known capabilities, motivations, and so on. As such, most important techniques appear to be fairly obvious: images can be cross-referenced against known landmarks, and through the checking of unique URLs. Social media adds some technological components that might be more at home in intelligence fields such as imagery intelligence (or, IMINT). For example, a basic knowledge of current capabilities of widespread imagery manipulation software such as WARP – a perspective modification tool – is necessary. Other techniques might be specifically related to certain social media platforms. Producing visually convincing photography of inauthentic tweets and Facebook content is straightforward and has been used in the past.[108]

'Crowd sourced' information

Background

It has been recognised in recent years that public safety requires the involvement of a large number of different actors. For example, the British counter-terrorism strategy relies on the active engagement of citizens.[109] In the US, the gang-prevention initiatives that work most effectively are those that have 'all-community' involvement from the police, social support services, charities, youth groups, local churches, parents' organisations, rehabilitation centres and schools.[110] There is significant potential for the police to create and

curate networks of citizens cooperating to keep their community safe. Indeed, this 'co-production' of safety and security has already developed in many areas, often at the instigation and insistence of civilian participants, not the police.

Social media can be, and often is, used to inform the public rapidly and directly, and as a way of directly asking the public for help and assistance in keeping the public safe. Possible applications are as diverse as policing itself, from reporting successes and providing reassurance to promoting community activities and engagement or delivering statements, particularly following a major incident (for example, the 2013 terrorist attack in Boston).

Dispelling rumours

A more controversial method of official engagement with the public is to dispel rumours and conspiracy theories, including by proactively intervening in discussions and conversations. Staffordshire Police have been using Twitter to dispel rumours since January 2010, particularly in relation to English Defence League protest and counter-protest events.[111] Most effective debunking efforts involve crafting counter-messages that are as appropriable and shareable – as 'sticky' - as the misinformation they seek to confront. For example, West Midlands Police Force used social media, particularly Twitter, to counter rumours of an attack on the police station by posting 'Twitpics' of officers standing outside the station.[112]

Any form of appropriate engagement with the public on social media requires a number of risks to be addressed and managed. The established culture of policing is necessarily based on command and control, hierarchy and operational security. It is conditioned by the role of the police as agents of the criminal justice system and hence the need to preserve the integrity of evidence and the rights of suspects and victims. These cultural values often sit uncomfortably with the openness, informality and public nature of communications on social media.

Direct solicitation for information

Perhaps the greatest way crowd sourcing is currently being used to collect information on individuals is the simplest: asking the public. Recently, some US police forces have also used content sharing sites, such as Pinterest, to ask for the public's assistance in identifying criminals. Following the 2011 UK riots, police uploaded photos to a Flickr stream and a dedicated website that compiled images of people thought to be involved in looting. As a result 770 people were arrested and 167 charged. Furthermore, up to 2,800 images were uploaded to the smartphone app Facewatch ID, created in partnership with Crimestoppers, which allowed users to sort the images via postcode and then inform on those they recognise by sending a name and address to the police. The app also included 2,000 or more images of people wanted for offenses not connected to the riots.[113] Similarly, in the aftermath, citizens organised themselves using #riotcleanup, and staged public demonstrations in condemnation of the criminality and the violence.

The development of mobile applications for the purpose of supporting incident reporting has been an area of significant activity. In the US, further development has focused on making the use of these apps more simple and ensuring more accurate reporting of incidents, events, and tip-offs (known as e-tips). Anonymous reporting functions have also been developed in an effort to extend popular participation.[114] The continued development of useful reporting apps could help create a citizens' network of reliable human intelligence sources for event detection. Anonymisation might encourage more accurate and more problematic reporting – but does not help in verifying the information itself. [115]

Survey solicitation

Research work by the authors has revealed that one useful way to collect information about groups through the use of social media is to target individuals for direct and open survey recruitment. In one study, researchers collected 1,200 completed survey responses from Facebook supporters of the English Defence League which revealed

new insights about the group, its members, its likely size, and its motivations.[116]

Challenges of two-way communications

The challenge is to set the right balance of central control. Counter-terrorism police operations increasingly work to secure positive citizen engagement, but police forces have understandably sought to limit the risks of this new environment by issuing guidance and establishing internal control procedures. Police forces usually issue strict guidance which requires police officers to protect the reputation of the force and to pay proper attention to operational considerations such as the protection of the identities of victims and witnesses, the protection of the integrity of current operations, as well the avoidance of comment which might be prejudicial to legal proceedings.

However well controlled, the opening of direct channels of communication between the public and the police poses inherent risks. Responsibility rests with the police to respond to emergency call, and with less degrees of urgency, other non-emergency forms of contact by the public. Most police forces reviewed by the authors have taken the view that tweets directed at an official account should not be treated with the same degree of urgency as other forms of communication – indeed, most police sites on Twitter contain a warning not to use the channel to report crime. Twitter feeds are not routinely staffed 24 hours a day or integrated into force control centres. Nevertheless, a number of forces are reporting a significant increase in the amount of information requests coming to them through social media – and there are not, as far as we know, systems in place to manage and filter these requests.[117] While this is not a significant problem yet, we anticipate it might become one in the near future.

Social media can also disrupt other forms of communication and engagement. Social media is challenging both for the press and for force press officers – journalists and reporters increasingly find breaking stories online, and seek police verification before the force is ready to confirm or deny a particular instance.[118]

PART 4: LAW AND ETHICS

The use of social media to collect information by law enforcement agencies presents challenges to the existing legal and ethical frameworks that manage the various types of harm that can result from intelligence gathering. Below, each framework is discussed in turn.

Overall, open SOCMINT does pose new challenges to existing legal frameworks that govern intelligence work, and ethical consideration for research work. If SOCMINT is to be used as a valuable and legitimate form of insight, we believe it must be based on a clear legal, publicly argued footing.

Lawful access and social media

Like all intelligence work, SOCMINT must be carried out within a legal framework. Most OECD countries have legislation that covers the collection and use of private information, which is intended to ensure that state agencies can only access citizens' private information in a legal, proportionate way, with various mechanisms of oversight and scrutiny designed to minimise potential abuses of power. Different countries have different legal frameworks underpinned by slightly different principles. In the UK for example, the collection of information that might reasonably be considered 'private' requires the utilization of a strict authorisation process and oversight by legitimate bodies, as well as that the intelligence is used for appropriate purposes and gathered using appropriate methods.

The main difficulty facing most law enforcement agencies when collecting information of any kind is the extent to which certain types of data collection might require a legal authorisation. In making this decision, the very broad principle to which most legal frameworks adhere is that of 'reasonable expectation' of privacy. Of course, there is sometimes a difference in how this principle is applied. In the UK, RIPA authorisation is required where there is a likelihood that 'private information' will be obtained, even if it comes from a public source. In Canada under the 'Intercept' parts of the Criminal Code, it appears the key consideration is whether the communication itself - rather than the content - might be

reasonably considered private.[119] In Belgium, whether 'technical means' are used has a bearing on whether a judicial or department warrant is necessary.

Despite these different considerations 'reasonable expectation' is a useful starting point in respect of social media. Based on our previous work on the subject, we believe that the specifics of any judgment about reasonable expectation will rely upon a number of distinctions and assessments – from what is proportionate to what is a private space – that are contextual, mutable and a matter of degree. In respect of social media, these assessments are extremely difficult to make because:

- SOCMINT covers both open-source data and closed networks. Sometimes, however, the distinction is not clear. For example, Facebook accounts and groups often have varying degrees of openness, and different platforms often have quite different terms and conditions and norms of use that might determine the degree of intrusion.

- Social media analysis software and tools allow for far greater surveillance than ever before, with concomitant risks and opportunities. The increased use of automated software to collect and analyse information (inevitable in the age of terabytes of unstructured data) poses additional risks of misuse.

- Public attitudes toward data sovereignty and privacy (even on open platforms) change quickly, and there are reputational risks for law enforcement agencies seen as 'snooping' online.

Determining 'reasonable expectation'

The academic Susan Brenner has highlighted two questions which are specific to US law enforcement, but which can provide a useful format through which to frame the consideration of reasonable expectation of privacy. Writing about 'search' under the 4th Amendment, Brenner draws on *Katz v. US*, 389 US 347 (1967),[120] and suggests that someone has a reasonable expectation of privacy in a place/thing if two conditions are met: (i) he thinks it

is private; and (ii) society accepts as objectively reasonable his belief that the place/thing is private.

It was re-affirmed as an important principle in a 2012 US Supreme Court decision relating to the FBI's use of warrantless GPS tracking devices placed on the underside of cars parked in public places: although the Court did not rule on the reasonable expectation consideration (being limited to whether the use of GPS constituted a 'search'), the principle was discussed in detail in the ruling and affirmed as the key principle at stake by Justices Alito and Sotomayer.[121]

In light of Brenner's interesting distinction, below we discuss some considerations that may help determine both an individual's and society's reasonable expectation of privacy.

Individual expectation

One way to determine an individual's expectation of privacy on social media is by reference to whether that individual has made any explicit effort or decision in order to ensure that third parties cannot access this information. This could be manifested in a series of ways:

- Any data coming from closed accounts, or any account or group where a restriction has been placed limiting the access (for example 'friends only' settings). This means that an explicit decision has been made to limit the access of outside parties and thus can be considered a 'private communication' even if the group involved is extremely large.
- Where a password is required in order to enter a site.
- Any robot.txt restriction that has been placed by the site administrator in order to prevent permission for a search bot or scraper to access data.

An individual's reasonable expectation can also be determined by reference to the terms and conditions of use of a forum or site and the typical behaviour of users, as these will often help to shape the expectations that an individual has about the nature of the

interaction. Not all 'open' platforms are the same in terms of the reasonable expectations of the user. This too can be manifested in several ways:

- Some chat room forums and threads have fairly explicit instructions that request that users sign in to take part, and that data and conversations are not shared outside of the group, while others (such as Twitter) make it clear that they will encourage people's personal information to be widely shared – and that users should be prepared for that.

- Often, targeting certain individuals results in the obtainment of information on other, non-targeted individuals who interact with the targeted individual, information which these non-targeted individuals may reasonably consider to be private.

- Most terms and conditions of sites – including Facebook – clearly state that users are expected to be honest about their profile information. Therefore, the creation of fake / pseudo social media accounts (on Facebook these are sometimes called 'ferret' accounts) in order to join a closed group or chat room, including when an individual joins using a blank or anonymous account, might be considered unreasonable.

- Similarly, any direct interaction in any forum – open or closed – in which an officer seeks to elicit information and are not explicit about their real identity can be problematic.

Social expectation

Brenner also argued that a reasonable expectation is driven by society's view of what is acceptable. Recent debates in relation to internet surveillance - in particular the Edward Snowden allegations – have demonstrated that public acceptability is fundamental to any measures being undertaken. Polling data reveals that fears over online privacy have become an issue of growing public concern.[122] Polling undertaken in 2012 suggests that the erosion of privacy is the second most important worry Canadians have, just behind the global financial crisis (but ahead of climate change and terrorism). Seventy-two per cent of Canadians

express concern about this compared to 73 per cent who worry about the financial crisis.[123] They worry specifically about online privacy too: a 2011 poll commissioned by Canada's federal privacy watchdog found that 82 per cent of Canadians were against giving police and spy agencies the power to access emails and online data without court authorisation.[124]

At the same time however, citizens expect security services and law enforcement agencies to have the necessary powers to fulfill their obligations in regards to public safety and security. The challenge posed to authorities, therefore, is to balance these potentially conflicting societal expectations of promoting safety and security while at the same time protecting citizens' online privacy. There are indications that the idea that law enforcement agencies should be able to access social media data for public safety and security purposes is generally supported.

For example, a 2008 Eurobarometer poll found that 80 per cent of European citizens trust the use of citizens' personal information in a proper way by police, while the same survey found that a majority supported the monitoring of internet activity to protect society against terrorism. Seventy-six per cent of UK respondents said they believe monitoring internet usage is acceptable in order to fight terrorism.[125] Similarly, a poll undertaken by consultancy firm Accenture in five countries, including Canada, found that 72 per cent of respondents believe social media can aid in criminal investigations and prosecutions.[126]

Taken together, therefore, we consider that the following types of SOCMINT collection might be reasonably considered as genuinely open-source and non-intrusive, where there is little or no expectation of privacy:

- Volunteered crowd-source intelligence through direct and explicit solicitation. This should be employed wherever possible instead of 'listening in' technologies and techniques. In these instances, the expectation of privacy can reasonably be considered to be very low.

- Where social media users have no reasonable expectation of a right to privacy, because they understand this content is likely to be shared and used. That condition is met if any terms of agreement establish that content uploaded is public and will be made available through an Application Programme Interface access. A good example is Twitter, which makes it clear that it will actively encourage sharing, which means that data collected from open Twitter accounts would not require authorisation. In addition, no privacy blocks or walls (often effected through the use of robot.txt restriction) or password requirements should exist.

- Network analysis through the use of 'crawlers' or 'spiders' (automated programmes to map a network of individual accounts), providing no individuals are named, no private information about an individual is collected, and providing API access is granted through robot.txt, and it is made clear in the terms and conditions that data are shared.

The use of automated systems can make these decisions more difficult. As a robots.txt file is not enforceable, they can be ignored when crawling. However, many would consider that 'bad manners'. On the other hand, this is not a privacy control, as the page is viewable publicly and the purpose for which open source intelligence is being used may outweigh the desire to 'be polite'.

Furthermore, the legal and ethical dimensions shift when the content or behaviours on a specific site are considered criminal. Additionally, every port requires a txt file, or else the bot assumes no restrictions are in place. Since 2011, for example, Facebook has banned the collection of data using automated means (without their explicit approval). Crawlers will tend to issue multiple requests to the site for information. Sometime a site administrator will notice multiple requests and then decide to prevent access. However, it is possible – and inexpensive – to use Virtual Private Networks which reroute requests, to make it appear to the site that the request is coming from multiple different sources. Again, this may result in serious privacy concerns.

In addition to robot.txt the server and API etiquette is important. Many systems pause, or sleep, briefly between calls. This is to prevent the crawl, or API calls, from putting unnecessary strain on the servers where the data is stored. The reason for this is partly etiquette, and partly practical. Some sites have the capacity to ban users that are making too great a demand on a server from accessing further information on that server. This is done through an IP ban, denying anyone using a specified IP address from accessing data on a server, or rate limiting when data is accessed through a platform API. There are a number of ways around these limitations, but the decision to get round an IP ban should be taken after consideration of the ethics of overriding the attempt by a system administrator to block further data collection – along with any legal considerations which relate within the relevant jurisdiction.

Given the reputational considerations set out above, public acceptability and proportionality should inform any decisions taken in respect of even open SOCMINT – even where a decision has been made that no warrant is required. Agencies undertaking this type of research work should try to conduct open SOCMINT work according to good ethical and professional research standards:

- Being explicit and public about the research aims and methods used where possible.

- Considering whether the measures taken might reasonably be seen as proportionate by those potentially monitored, and could be defended as such. Even where data is anonymised, there is increasing public concern about 'pseudo' anonymous data, where individuals can be identified by cross-referencing data sets.

- Assessing if any measures might undermine the existence of a free and open internet, which would cause damage to the economic and social well-being of the nation. It is our view the benefits of collecting and storing large amounts of open data in a general, non-targeted way, should be carefully weighed against this possible risk.

- Assessing whether such measures are an effective use of public money.

Research ethics

Research ethics are not legally binding. Rather, they are a set of commonly agreed principles by which academic research institutions undertake research. Similar to legislation to cover intelligence work, they aim to measure and minimise harm, and in this instance balance the need to undertake socially useful research against possible risks to those involved. In contrast to the law, they cover more varied harm considerations than the law, and are usually affected by university ethics committees or institutional review boards.

Principles of research ethics and ethical treatment of persons are codified in a number of policies and accepted documents such as the UN Declaration of Human Rights and the Declaration of Helsinki, which aim to uphold the principles of human dignity, safety, respect for individuals, and maximise benefits while minimising harms. In the UK, the standard best practice is the Economic and Social Research Council (ESRC) ethical framework, composed of six principles that express this broader moral-ethical doctrine.[127] The ESRC will not offer funding to research projects that do not demonstrably adhere to these principles.

Social media research is a new field, and the extent to which (and how) these ethical guidelines apply practically to research taking place on social media is as of yet unclear. Because the nature of social media research is highly varied – ranging from large quantitative data analysis down to very detailed anthropology – there is no single approach that can be applied.

One significant (and unresolved) debate within the academic community in the UK is the extent to which social media analytics requires approval from ethics committees at all. This is because most universities consider that ethics committees are required where research is undertaken on 'human subjects'. Where social media analytics research is led by computer science departments,

this is generally viewed as research on a non-human subject. Indeed, the term 'human subject' as applied to the consideration of research ethics itself emerged from harmful medical experiments, and may therefore be ill-suited for inquiry where there is no direct interaction with people. This dynamism is reflected in the fact that as of the time of this writing, no official frameworks regarding internet research ethics have been adopted at any national or international level.[128]

There is, however, a growing group of internet researchers who have issued various types of guidance themselves. The Association of internet Researchers (AoIR) released its latest ethical framework in March 2013.[129] This guidance – first issued in 2002 and frequently updated – is commonly used by institutional review boards when making decisions. The AoIR note that ethical principles cannot be applied universally, but must rather be understood inductively and through the use of applied practical judgments.

Crucially, they note that because all digital information at some point involves individuals, consideration of principles related to research on human subjects may be necessary even if it is not immediately apparent how and where persons are involved in the research data. [130]

The most commonly applied principles for human subject research are: a) any possible harms to participants must be measured, managed, and minimised; b) informed consent should be sought when and where possible. These guidelines are considered below.

Harm to participants

There is a broadly agreed obligation for researchers to avoid research that is harmful to its subjects, irrespective of how research is collected. Harm is difficult to measure in respect of social media research. There may be new harms related to mass data extraction, such as a loss of confidence in the platform. Extraction tools need to be designed to avoid accidental extraction from non-public

accounts, and new forms of collection – such as extracting profile information – might in some instances require explicit consent.

This type of ethical consideration is important when using crawlers and other automated bots, particularly when one considers how the digital medium can potentially dehumanise research subjects. Crawlers can collect information about user profiles, types, videos uploaded, and place individuals within a network. Individuals within that network - especially on the fringes - might very reasonably consider this to be a significant harm. If the crawler is looking at linked sites, it will quickly fall upon rather more moderate forums and individuals who may find themselves on the fringes of a network of extremist sites.

The presentation of the data is a critical consideration. According to the British Psychological Society, researchers:

Should avoid using quotes that are traceable to an individual posting via a search engine unless the participant has fully understood and consented to this.

As such, quotes should be paraphrased, and not linked to the forum they were gathered from.[131]

Informed consent

In traditional research methods, the principle of informed consent refers to the need for researchers to be open about who they are, about the purpose of their work and about how it will be disseminated. Informed consent is considered of vital importance as a way to minimise harm, and is meant to ensure that there is no explicit or implicit coercion so that research subjects can make an informed and free decision on their involvement in the research. They should therefore be informed about the fact that information they share is being used for research purposes. Informed consent is not always necessary, however, and in certain cases such consent is widely acknowledged to be impracticable or meaningless, for example in research on crowd behaviour.[132]

While such research should not be undertaken without particular caution and consideration, research without informed consent can

be justified when no details about an individual are likely to be divulged, and where the risk of harm to research subjects is fully minimised.

The application of informed consent is likely to vary depending on the type of research being undertaken. The developing field of internet research poses various new challenges to this basic research principle because of the ambiguity of the concepts of privacy and informed consent in online settings, and the difficulties of establishing the real identity of research subjects and of obtaining their consent.

In the UK, the British Sociological Association and the British Psychological Society (BPS) argue that full ethical frameworks should be adhered to when undertaking social media research. BPS guidelines specifically state that unless consent has been sought, observation of public behaviour needs to take place only where people would 'reasonably expect to be observed by strangers'. As such, similar considerations relating to reasonable expectation as set out above might apply here, such as the express decisions of the user and the norms and expectations of the site. A recent Canadian case *Century 21 Canada Limited Partnership v. Rogers Communications Inc*, 2011 BCSC 1196 (CanLII) also appears to show that the collection and reproduction of proprietary information through automated bots can result in liability.[133]

The extent to which all of these principles are applicable will partly depend on the extent to which the platform is open or closed. Similarly to the legal framework, one useful heuristic is the social networks' own privacy policies, for example, Facebook's restriction of the use of web-crawlers in March 2011. This is useful because it is an important indication of the expectation of privacy. As it stands, it is generally agreed that Twitter data is in the public domain and can therefore be treated as carrying implicit informed consent. Twitter's Terms of Service[134] and Privacy Policy[135] clearly state:

What you say on Twitter may be viewed all around the world instantly. We encourage and permit broad re-use of Content. The Twitter API exists to enable this.

Even with open source data, however, certain conditions still ought to be met. Because this are of research is so changeable and often difficult to interpret, principles rather than hard and fast rules are most suitable (for example, the use of ‘situational ethics’). These principles need to take into account frequent technological changes, the medium involved and the expectations of the research subjects.

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.

b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.

c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.

d 'Original Author' means the individual or entity who created the Work.

e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.

f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work,or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive,perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b to distribute copies or phonorecords of, display publicly,perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised.The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a You may distribute,publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phonorecord of the Work You distribute, publicly display,publicly perform, or publicly digitally perform.You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder.You may not sublicence the Work.You must keep intact all notices that refer to this Licence and to the disclaimer of warranties.You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement.The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licencor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.

b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation.The exchange of the Work for other copyrighted works by means of digital filesharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

C If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works,You must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

A By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

B except as expressly stated in this licence or otherwise agreed in writing or required by applicable law,the work is licenced on an 'as is'basis,without warranties of any kind, either express or implied including,without limitation,any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental,consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

A This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence,however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

B Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

A Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

B If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

C No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

D This Licence constitutes the entire agreement between the parties with respect to the Work licensed here.There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You.This Licence may not be modified without the mutual written agreement of Demos and You.

NOTES

[1] Early and emerging examples of Twitterology were presented at the International Conference on Web Search and Data Mining 2008.
[2] Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science, 7*(3), 203-220. doi: 10.1177/1745691612442904
[3] The explicit purpose was agreed as "To prepare a paper that reviews the state of knowledge about what the increasing range of social media analysis software tools can and cannot do; what social science questions can and cannot be answered with such technologies, including what does and does not make for useful evidence; how such research questions can inform contemporary government priorities for addressing knowledge gaps on terrorism and counter-terrorism; and implications for conducting this kind of research, including a focus on research ethics."
[4] O'Reilly, T. 'What is Web 2.0 Design Patterns and business models for the next generation of software'. http://oreilly.com/1pt/a/6228
[5] Kaplan, A.M and Haenlein, M. (2010) Users of the world, unite! The challenges and opportunities of social media. *Business Horizons* 53(1) 59-68
[6] Omand, D., Bartlett, J & Miller, C (2012) *#Intelligence*
[7] See Schaurer, Florian (2012) 'Social Media Intelligence (SOCMINT) – Same song, new melody?', *Open Source Intelligence,* http://osintblog.org/?p=1462.
[8] Stephen C. Mercado, CIA Center for the Study of Intelligence, *Studies in Intelligence,* Vol. 48 No.3, https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol48no3/article05.html
[9] Lynda Peters 2012 'utilising osical media to further the national suspicious activity reporting initiative'.
[10] As defined by David Omand in Securing the State.
[11] It is also possible to acquire large amount of social media data via a licensed data provider. These are often third party re-sellers.
[12] Apart from the Graph API discussed here, there is a Facebook Query Language API, an Open Graph API, a Dialogs API, a Chat API, an Internationalization API, and an Ads API. For more on these, see: https://developers.facebook.com/docs/reference/apis/
[13] Example screen grabs of the format in which information is given through an API request for both Facebook's Graph API and Twitter's search API are included in Annex III.
[14] For a map of current Twitter languages and demographic data, see:
http://www.flickr.com/photos/walkingsf/6277163176/in/photostream/lightbox/
http://expandedramblings.com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats/
[15] Simon Fodden, Slaw, Anatomy of a Tweet: Metadata on Twitter
http://www.slaw.ca/2011/11/17/the-anatomy-of-a-tweet-metadata-on-twitter/, and see
http://www.slaw.ca/wp-content/uploads/2011/11/map-of-a-tweet-copy.pdf
[16] For more information, see: https://dev.twitter.com/docs/api/1.1
[17] https://scraperwiki.com
[18] Chau, M & Xu, J 'Mining communities and their relationships in blogs: A study of online hate groups' *International Journal of Human-Computer Studies*, p62
[19] Mining Social Media: Tracking Content and Predicting Behavior, Manos Tsagkias. Ph.D. thesis, University of Amsterdam
[20] Stuart Shulman 'Keeping Humans in the Machine Learning Loop' (March, 28, 2013). Paper presented at the *Paper presented to the social text workshop (closed)*, University of Birmingham, 28 March 2013
[21] O'Hara, K., Berner-Lee, T., Hall, W & Shadbolt, N (2011) "The use of the semantic web in e-Research", in Dutton, W & Jeffreys, P (ed.) *World Wide Research: Reshaping the Sciences and Humanities*
[22] Stuart Shulman 'Keeping Humans in the Machine Learning Loop' (March, 28, 2013) Founder of 'Discover Text'.
[23] http://www.relevantdata.com/pdfs/IUStudy.pdf
[24] http://aclweb.org/anthology-new/E/E12/E12-1062.pdf

[25] Stuart Shulman 'Keeping Humans in the Machine Learning Loop' (March, 28, 2013). Paper presented at the *Paper presented to the social text workshop (closed)*, University of Birmingham, 28 March 2013
[26] Viktor Mayer-Schonberger and Kenneth Cukier, 'Big Data'.
[27] Stephen Jeffares, 'Coding policy Tweets', Paper presented to the social text analysis workshop, univeristy of Birmingham, 28 March 2012
[28] Exploiting Social Relations for Sentiment Analysis in Microblogging, WSDM Conference, Xia Hu, Lei Tang, Jiliang Tang, Huan Liu, 2013
[29] http://research.google.com/pubs/NaturalLanguageProcessing.html
[30] *www.mitre.org/work/tech_papers/2011/11_0170/11_0170.pdf*
[31] Kalev H. Leetaru, Shaowen Wang, Guofend Cao, Anand Padmanabhan, Eric Shook (2013) 'Mapping the global Twitter heartbeat: the geography of Twitter', *First Monday Vol. 18, No.5,* pp.1-33.
[32] Watanabe, K., Ochi, M., Okabe, M., & Onai, R. (2011). Jasmine: a real-time local-event detection system based on geolocation information propagated to microblogs. Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 2541–2544). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2064014
[33] Kosinskia, M., Stillwella, D & Graepelb, T (2013) 'Private traits and attributes are predictable from
digital records of human behavior', *Proceedings of the National Academy of Science*, (submission)
[34] The data, which other researchers have subsequently used for further work is all available here: http://mypersonality.org/wiki/doku.php
[35] Kosinskia, M., Stillwella, D & Graepelb, T (2013) 'Private traits and attributes are predictable from
digital records of human behavior', *Proceedings of the National Academy of Science*, (submission)
[36] Oliver Mason, *Paper presented to the social text workshop (closed)*, University of Birmingham, 28 March 2013
[37] See for example: http://www.cnn.com/2012/07/09/tech/innovation/police-tech
[38] Morozov, E (2013) 'How Facebook could get you arrested', Observer, March 2013. http://www.theguardian.com/technology/2013/mar/09/facebook-arrested-evgeny-morozov-extract Accessed 18/9/2013
[39] Xuning Tang, Christopher C. Yang , TUT: a statistical model for detecting trends, topics and user interests in social media, International Conference on Information and Knowledge Management 2012,
[40] A review of Facebook research in the social sciences, Perspectives on Psychological Sceince 7(3) 203-220
[41] Sakaki, T., Okazaki, M & Matsuo, Y (2010) 'Earthquake Shakes Twitter Users: Real-Time Event Detection By Social Sensors', International Conference on the World Wide Web,
[42] Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman and Ke Tao, Fighting Fire with Information from Social Web Streams, International Conference on the World Wide Web, 2012
[43] Learning similarity metrics for event identification in social media, WSDM Conference, Hila Becker, Mor Naaman and Luis Gravano, 2010.
[44] Osborne, M., & Lavrenko, V. (2010) 'Streaming First Story Detection with application to Twitter. Computational Linguistics,' (June), 181–189.
[45] Mendoza, M., Poblete, B., Castillo, C , Twitter Under Crisis: Can we Trust What we RT?, KDD Workshop on Social Media Analytics, 2010
[46] Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. Proceedings of the 20th international conference on World wide web - WWW '11, 675. doi:10.1145/1963405.1963500
[47] A Hughes "Twitter adoption and use in mass convergence and emergency events" IGCRAM, 2009
[48] Heeley, R (unpublished, 2011) *Mobile Collaborative Filtering Protest analysis using Twitter and Android.* Submitted in partial fulfilment of the requirements for the MSc Degree in Computing Science of Imperial College London
[49] *Ibid*
[50] Streitfeld, D. (2012). The Best Book Reviews Money Can Buy. *The New York Times.* http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html

[51] Liu, B., Mukherjee, A. & Glance, N. (2012). Spotting Fake Reviewer Groups in Consumer Reviews. *International World Wide Web Conference Committee.* http://www.cs.uic.edu/~liub/publications/WWW-2012-group-spam-camera-final.pdf; Ott, M., Cardie, C., Choi, Y., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 309-319. http://projects.iq.harvard.edu/files/ptr/files/cardie_deception.pdf; Ott, M., Cardie, C., & Hancock, J. T. (2012). Estimating the prevalence of deception in online review communities. *International World Wide Web Conference Committee*. http://www.cs.cornell.edu/home/cardie/papers/www-2012.pdf; Streitfeld, D. (2011). In a Race to Out-Rave, 5-Star Web Reviews go for $5. *New York Times*. http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html?_r=1&hp

[52] Jakobsson, M. (2012). *The Death of the internet*, John Wiley, New Jersey, p.234.

[53] Eaton, E. (2012). There are more 'Fake' People on Facebook than Real Ones on Instagram. *Fast Company*

[54] Reips, U. (2011) Mining Twitter: a Source for Wisdom of the Crowds. *Behavior Research Methods,* 43 (3), pp.636-642.

[55] Burke, M 2009 Feed me: motivating newcomer contribution in social network sites in *Proceedings of the 27th Conference on human factors in computing systems*

[56] See http://www.guardian.co.uk/uk/interactive/2011/dec/07/london-riots-twitter for the visualisation of the spread of misinformation

[57] Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data,* John Murray, London: 2013.

[58] http://smartdatacollective.com/ericsiegel/113761/new-predictive-profession-odd-yet-newly-legitimate?utm_source=feedburner&utm_medium=feed&utm_campaign=Smart+Data+Collective+%28all+posts%29

[59] Wang, X., Gerber, M., and Brown, D.E., (2012) 'Automatic Crime Prediction Using Events Extracted from Twitter Posts', Social Computing, Behavioral - Cultural Modeling and Prediction Lecture Notes in Computer Science Volume 7227, 2012, pp 231-238

[60] Yu., S & Kak, S (unpublished, 2012) 'A Survey of Prediction using Social Media'.

[61] http://smartdatacollective.com/ericsiegel/113761/new-predictive-profession-odd-yet-newly-legitimate?utm_source=feedburner&utm_medium=feed&utm_campaign=Smart+Data+Collective+%28all+posts%29

[62] See: http://www.guardian.co.uk/media/pda/2010/may/13/twitter-tweetminster-election for an overview of Tweetminster's study attempting to predict the UK 2010 general election.

[63] Jungherr, A., Jurgens, P, Schoen, H (2011) "Why the Pirate Party Won the German Election of 2009, or The Trouble with Predictions" *Social Science Computer Review*

[64] http://www.technologyreview.com/view/425280/new-media-and-the-people-powered-uprisings/?p1=A3

[65] See: http://www.newyorker.com/reporting/2010/10/04/101004fa_fact_gladwell?currentPage=all

[66] Available here at www.google.org/flutrends

[67] Sadilek, A, Kautz, H & Silenzo, V (2012) 'Modelling Spread of Disease from Social Interactions" *Association for the Advancement of Artificial Intelligence*.

[68] Latent Geospatial Semantics of Social Media, Sergej Sizov, WSDM Conference, 2010.

[69] Wang et al (2012) *Automatic Crime Prediction Using Events Extracted from Twitter Post*

[70] See: http://www.guardian.co.uk/world/2013/feb/10/software-tracks-social-media-defence, Accessed 21 March, 2013

[71] Silver, N (2012) *The Signal and The Noise*, Penguin: London, p.452

[72] Kosinskia, M., Stillwella, D & Graepelb, T (2013) 'Private traits and attributes are predictable from digital records of human behavior', *Proceedings of the National Academy of Science*, (submission)

[73] Silver, *The Signal and The Noise*, p.452

[74] Borgatti, Stephen P., and Daniel S. Halgin. "On network theory." *Organization Science* 22.5 (2011): 1168-1181.

[75] Bruns, Axel (2007) 'Methodologies for Mapping the Political Blogosphere: An Exploration Using the IssueCrawler Research Tool'. *First Monday 12(5).* Devereaux, Zachary, Wendy Cukier, Peter Ryan, and Neil Thomlinson. "Using the Issue Crawler to Map Gun Control Issue-Networks." (2009); Moe, H "Mapping the Norwegian blogosphere: Methodological challenges in internationalizing internet research." *Social Science Computer Review 29.3* (2011): 313-326.

[76] As discussed later, on many social media platforms, calls to the API (application programming interface) are frequently more effective for gathering structured data than crawlers.
[77] Sparrow, Malcolm K. "Network vulnerabilities and strategic intelligence in law enforcement." *International Journal of Intelligence and Counter Intelligence*5.3 (1991): 255-274.
[78] Valdis Krebs. "Uncloaking Terrorist Networks" *First Monday* [Online], Volume 7 Number 4 (1 April 2002) Valdis Krebs, 'Mapping Networks of Terrorist Cells' *CONNECTIONS* 24(3): 43-52 2002
Richard Medina and George Hepner 'Advancing the Understanding of Sociospatial Dependencies in Terrorist Networks', *Transactions in GIS*, vol. 15, no 5, 2011
[79] Malcolm Sparrow, 'Network vulnerabilities and strategic intelligence in law enforcement', *International Journal of Intelligence and Counter Intelligence* (1991), Vol.5, No.3, pp.255-274.
[80] Weng, J., Lim, E & Jiang J (2010) 'TwitterRank: Finding Topic Sensitive Influential Twitterers' *WSDM 10,* February 2010
[81] Berger, J & Strathearn, B (2013) *Who Matters Online*, ICSR: London
[82] Berger, J & Strathearn, B (2013) *Who Matters Online*, ICSR: London
[83] O'Callaghan, D., Greene, D., Conway, M., Carthy, J, Cunningham, P (2013, forthcoming) 'An analysis of interactions within and between extreme right communities in social media'.
[84] Xiang, R., Neville, J. & Rogati, M (2010) 'Relationship Strength In Online Social Networks', *International Conference on the World Wide Web*.
[85] Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science, 7*(3), 203-220. doi: 10.1177/1745691612442904; Tsagkias, M (2012) *Mining Social Media: Tracking Content and Predicting Behavior*, Ph.D. thesis, University of Amsterdam, p10
[86] Positive and Negative Links in online Social Networks, International Conference on the World Wide Web, Jure Leskovec, Daniel Huttenlocher, Jon Kleinberg, 2010
[87] All blogs were hosted on xanga.com.
[88] Chau, M & Xu, J 'Mining communities and their relationships in blogs: A study of online hate groups' *International Journal of Human-Computer Studies*, p67
[89] For example, Weng, J., Lim, E., Jiang, J & and He, Q (2010) 'TwitterRank: Finding Topic-sensitive Influential Twitterers', *Paper presented at the WSDM Conference, 2010*; Bakshy, E., Hofman, J., Mason, W & Watts, D (2010)'Everyone's an influencer: quantifying influence on Twitter', *Paper presented at the WSDM Conference, 2010*.
[90] Goyal, A, Bonchi, F & Lak, V, (2010) 'Learning Influence Probabilities In Social Networks', *Paper presented at the WSDM Conference, 2010*.
[91] Rob Cross, Andrew Parker and Stephen P. Borgatti, 'A bird's-eye view: Using social network
analysis to improve knowledge creation and sharing', *IBM Institute for Business Value*, http://www.analytictech.com/borgatti/papers/cross,%20parker%20and%20borgatti%20-%20A_birds_eye_view.pdf
[92] Node centrality in weighted networks: Generalizing degree and shortest paths: http://toreopsahl.com/2010/04/21/article-node-centrality-in-weighted-networks-generalizing-degree-and-shortest-paths/
[93] Fisher, A *Everybody's getting hooked up; building innovative strategies in the era of big data*, 2012 p.49
[94] Interview, Alberto Nardelli, founder *Tweetminster*
[95] Mining Social Media: Tracking Content and Predicting Behavior, Manos Tsagkias. Ph.D. thesis, University of Amsterdam, 2012.
[96] Ibid., p.54
[97] Bartlett, J *et al* (2013, forthcoming) *Politics by Twitter? Understanding public attitudes using social media,* Demos
[98] Sparrow, M. "Network vulnerabilities and strategic intelligence in law enforcement." *International Journal of Intelligence and Counter Intelligence* 5.3 (1991): 255-274.
[99] Klerks, P (2001) "The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands." *Connections* 24.3 (2001): 53-65.
[100] Stephen P. Borgatti, Daniel S. Halgin, On Network Theory (2011) http://www.steveborgatti.com/papers/orsc.1110.0641.pdf

[101] Bond, E (2012) *Virtually Anorexic – where's the harm*, Nominet Trust
[102] Meleagrou-Hitchens, A., Maher, S % Sheehan, J (2012), *Lights, Camera, Jihad: Al-Shabaab's Western Media Strategy*, ICSR, London.
[103] See: http://www.journal-online.co.uk/article/5410-police-use-facebook-to-identify-weapon-carriers (accessed 21 March 2013)
[104] Public Safety Canada (2011) *Social Media Sites: New Fora for Criminal, Communication, and Investigation Opportunities*,
[105] See: http://theriskyshift.com/2012/12/a-somali-soap-opera-al-shabaabs-split-with-omar-hammami/
[106] Kington, T. (2012). Twitter Hoaxer Comes Clean and Says: I did it to Expose Weak Media. *The Guardian.* http://www.guardian.co.uk/technology/2012/mar/30/twitter-hoaxer-tommaso-de-benedetti; Their, D. (2012). How This Guy Lied His Way Into MSNBC, ABC News, The New York Times and More. *Forbes.* http://www.forbes.com/sites/davidthier/2012/07/18/how-this-guy-lied-his-way-into-msnbc-abc-news-the-new-york-times-and-more/
[107] Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science, 7*(3), 203-220. doi: 10.1177/1745691612442904, p20
[108] See: http://www.huffingtonpost.co.uk/jamie-bartlett/did-the-edl-really-tweet-that_b_2313942.html
[109] Cabinet Office*, A Strong Britain in an Age of Uncertainty: The National Security Strategy*, Oct 2010
[110] The Gang Resistance and Training (GREAT) programme, introduced by the Phoenix Police Department in 2001, involved police-led sessions incorporated into the school curriculum to promote conflict-resolution skills, cultural sensitivity and demonstrate the negative consequences of gang life. This scheme, despite being competently run and well managed, demonstrated statistically negligible results not sustained over the long term. Ebensen F., Osgood, D.W. (1999), Gang Resistance Education and Training (G.R.E.A.T): Results from the national evaluation, Journal of Research in Crime and Delinquency 36. Often cited as a success story, Boston's Operation Ceasefire and the Little Village Gang Violence Reduction Project (GVRP) a number of different non-police actors cooperated in the into counseling and occupational advice programmes, and other community engagement projects.
[111] Ibid
[112] HMIC, *The Rules of Engagement: Report into the August 2011 Disorders*, 2011, p. 32.
[113] See: http://www.facewatchid.co.uk/
[114] Lynda Peters (2012) *Utilising Social Media to Further the Nationwide Suspicious Activity Reporting Initiative*, Masters Degree Disseration, Calhoun University
[115] Bach, C., Winckler, M, Gatellier, B "Challengs for the gamification of incident reporting systems", *Fun and Games 2012*
[116] Bartlett, J & Littler, M (2011) *Inside the English Defence League*, Demos
[117] Interview, [anonymous source]
[118] *Leveson Enquiry Evidence Day 53 afternoon* pp 20-21
[119] Criminal Code, section 183, http://laws.justice.gc.ca/eng/acts/C-46/page-87.html#h-61 (last accessed 27.03.2013). According to section 183 of the Code, 'private communications' are defined as: "any oral communication, or any telecommunication, that is made by an originator who is in Canada or is intended by the originator to be received by a person who is in Canada and that is made under circumstances in which it is reasonable for the originator to expect that it will not be intercepted by any person other than the person intended by the originator to receive it, and includes any radio-based telephone communication that is treated electronically or otherwise for the purpose of preventing intelligible reception by any person other than the person intended by the originator to receive it."
See *Katz v. United States* - 389 US 347 (1967), http://supreme.justia.com/cases/federal/us/389/347/case.html (last accessed 27.03.2013)
[121] Supreme Court of the United States, *Opinion on United States v. Antoine Jones*, 2011, http://www.**supremecourt**.gov/opinions/11pdf/10-1259.pdf (last accessed 27.03.2013)
[122] McCann Truth Central, *The Truth about Privacy: Canada and Beyond*, 2012, http://www.adstandards.com/en/MediaAndEvents/TruthAboutPrivacy.pdf (last accessed 27.03.2013); Toronto Sun, *Canadians oppose internet spy law: poll*, 2011, http://www.torontosun.com/2011/08/25/canadians-oppose-internet-spy-law-poll (last accessed 28.03.2013); Bartlett, J, *The Data Dialogue*, Demos, 2010. This is based on a representative population level poll of circa 5,000 people. Also see Bartlett, J & Miller, C,

Demos CASM submission to the Joint Committee on the Draft Communications Data Bill, Demos, 2012.
[123] McCann Truth Central, *The Truth about Privacy: Canada and Beyond*, 2012, http://www.adstandards.com/en/MediaAndEvents/TruthAboutPrivacy.pdf (last accessed 27.03.2013)
[124] Toronto Sun, *Canadians oppose internet spy law: poll,* 2011, http://www.torontosun.com/2011/08/25/canadians-oppose-internet-spy-law-poll (last accessed 28.03.2013)
[125] European Commission, *Data Protection in the European Union: Citizens' Perceptions. Analytical Report*, Eurobarometer 255, Feb 2008, http://ec.europa.eu/public_opinion/flash/fl_225_en.pdf (last accessed 27.03.2013)
[126] Accenture (2012) *Are Police Maximizing Technology to Fight Crime and Engage Citizens?* http://www.accenture.com/policecitizensurvey (Accessed 27.03.2013)
[127] The six principles are: 1) research should be designed, reviewed and undertaken to ensure integrity, quality and transparency; 2) research staff and participants must normally be informed fully about the purpose, methods and intended possible uses of the research, what their participation in the
research entails and what risks, if any, are involved; 3) the confidentiality of information supplied by research participants and the anonymity of respondents must be respected; 4) research participants must take part voluntarily, free from any coercion; 5) harm to research participants and researchers must be avoided in all instances; 6) t he independence of research must be clear, and any conflicts of interest or partiality must be explicit. ESRC, Framework for Research Ethics, latest version: September 2012, http://www.esrc.ac.uk/_images/Framework-for-Research-Ethics_tcm8-4586.pdf (last accessed 27.03.2013)
[128] AoIR, *Ethical Decision-Making and internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0),* 2012, p2
[129] *Ibid*
[130] *Ibid*
[131] British Psychological Society, *Conducting Research on the internet: Guidelines for ethical practice in psychological research online*, 2007
[132] ESRC, Framework for Research Ethics, latest version: September 2012, http://www.esrc.ac.uk/_images/Framework-for-Research-Ethics_tcm8-4586.pdf (Accessed 27.03.2013)
[133] *Century 21 Canada Limited Partnership v. Rogers Communications Inc,* 2011BCSC 1196 (CanLII), http://www.canlii.org/en/bc/bcsc/doc/2011/2011bcsc1196/2011bcsc1196.html (Accessed 27.03.2013)
[134] Twitter Terms of Service, http://www.twitter.com/tos (Accessed 27.03.2013)
[135] Twitter Privacy Policy, http://www.twitter.com/privacy (Accessed 27.03.2013)