THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Encountering #Feminism on Twitter

OPEN ACCESS

Encountering #Feminism on Twitter: Reflections on a research collaboration between social scientists and computer scientists

Steve Kirkwood & Viv Cree (The University of Edinburgh)

Daniel Winterstein, Alex Nuttgens, Jenni Sneddon (SoDash)

Abstract

The growth of social media presents an unparalleled opportunity for the study of social change. However, the speed and scale of this growth presents challenges for social scientists, particularly those whose methodologies tend to rely on the qualitative analysis of data that are gathered first-hand. Alongside the growth of social media, companies have emerged which have developed tools for interrogating 'big data', although often unconnected from social scientists. It is self-evident that collaboration between social scientists and social media analysis companies offers the potential for developing methods for analysing social change on large scales, bringing together their respective expertise in technological innovations and knowledge of social science. What is less well-known is how such a partnership might work in practice. This article presents an example of such a collaboration, highlighting the opportunities and challenges that arose in the context of an exploration of feminism on Twitter. As will be shown, machine-learning technologies allow the analysis of data on a scale that would be impossible for human analysts, yet such approaches also heighten challenges regarding the study of social change and communication.

Keywords

Collaboration, feminism, social media, Twitter, 'big data', machine-learning, qualitative analysis.

Introduction

The growth of social media and web 2.0 technologies have provided new ways of communicating in the digital age. Tinati, Halford, Carr and Pope (2014: 665) describe 'an extraordinary proliferation of user-generated content on the World Wide Web' over the last decade, as platforms such as Facebook, Twitter, YouTube and Wikipedia have 'captured the popular imagination and become embedded in the daily practices of people, businesses and governments around the world'. Tinati *et al.* see this as evidence of a shift away from 'social as society' (society bounded by nation states) to 'society as mobility' (dynamic flows of people, objects, images and information); in this changed context, networks no longer simply reflect society, but instead, they shape or produce society (Urry, 2000). Analysis of social media therefore offers great opportunities to researchers wishing to understand how information flows and how social roles and networks are produced and reproduced here and now.

Twitter offers one such example. Twitter began in 2006 as a micro-blogging website, which allows people to 'tweet' their messages in 140 characters. In contrast to other social media platforms, Twitter is mostly open and public: tweets can be recovered through Twitter's search tools, unless a direct message is sent to an individual privately. It is also non-reciprocal and directed: users 'follow' whoever they wish unless the person they are following actively blocks them. They do so without expectation of interchange; they may not know the person they have chosen to follow, and that person may opt not to follow them. Users 'see' the tweets of those whom they are following through an individualised timeline; they can also see the tweets of someone they are not following by accessing that person's profile and, hence, the record of their sent tweets. One person's tweets can be retweeted by someone else for everyone in their own network, thus facilitating what has been called 'interactive multicasting' (Murthy, 2012: 1062). Twitter (2017) reported 330 million monthly active users worldwide in the third quarter of 2017.

This article reports on an attempt to use Twitter for the purposes of social science research, focusing on the topic of feminism. The research came about, in large part, accidentally. One of the researchers, a long-term advocate of feminism, was working on a research study of feminism in social work (her academic discipline) and she had recently joined Twitter in order to expand the conversations she was having about feminism and about social issues more generally. In March 2014, Twitter issued an open call, inviting researchers from across the world to bid for projects that would use the full potential of Twitter's massive data source for academic research. This seemed too good an opportunity to miss; it offered the possibility of opening out what had started as a rather narrow interest in feminism and social work to a much broader exploration of the existence and impact of feminism more generally, through an analysis of Twitter. The first researcher invited a male colleague to join her, someone whom she knew was both sympathetic to feminism and active on Twitter; his expertise in qualitative data analysis, and especially discourse analysis, would be extremely useful to such a research project.

She also approached two computer scientists with whom she had worked previously on an admissions' modelling exercise, and learned that they were now running a company offering social media analysis predominantly to the private sector. The team came together to devise an outline for a project entitled, 'The Battle for Control of an Agenda - Exploring Attitudes to Feminism on Twitter'. Our application was, ultimately, unsuccessful, so we sought (and received) a small grant from a university research committee, and went on to conduct a similar, though inevitably smaller-scale, project. This paper offers a reflection on the research process that unfolded. While the project did throw up data that were both interesting and, ultimately rather disappointing for the feminist cause, the lessons learned were much more importantly about the methodological challenges in conducting this study.

Using Twitter for research

Twitter as a data source
It is argued that there are a range of advantages, as well as limitations, in using Twitter data over more conventional forms of data and also over other social media sources for research purposes (see Nagler & Tucker 2015). Advantages are that Twitter data is largely publically available, and may be collected through the Twitter Application Programming Interface (API) for research purposes, albeit with certain limitations on how it may be collected and used. Given Twitter's large user-base, it provides a way of gaining data on a large scale, with strong potential for generalising to wider populations. The data-set is naturalistic (Potter & Hepburn, 2005) in the sense that it has not been created for research purposes, meaning it is not influenced by some of the data collection tools (e.g., interviews or questionnaires) used in conventional research, giving it greater ecological validity. In the words of Nagler and Tucker (2015: 86): 'We can think of Twitter as having completely open-ended responses, in which the analyst has not even posed a question.' Tweets also have time stamps, meaning it is possible to track them with great specificity or in real time. The shortness of tweets lends them to being coded as discrete units, which may assist the coding, analysis and management of the data. Also, because they are accessible, researchers are able to check the results from previous research and attempt replication.

However, the use of Twitter as a source of data comes with a range of limitations. Although large numbers of people from around the world use Twitter, they are not representative of the general population. UK Twitter users are younger and more concentrated in professional occupations than the general population (Sloan *et al.*, 2015). Globally, the most common language used on Twitter is English and the highest concentration of users is based in the USA (Sloan *et al.*, 2013). The shortness of the tweets also means it may be difficult to code and analyse them in meaningful ways, particularly as the 'context' of the tweet may be difficult to identify, especially at scale. Moreover, researchers do

not have access to all tweets (the 'firehose') but rather only a small proportion via the API (the 'sprinkler'); this is a general issue for social media research, where the social media companies have full access to the data, but third parties do not, which inevitably limits the way in which research can be carried out and the generalisability of findings (Zelenkauskaite & Bucy, 2016).

Edwards *et al*. (2013) assert that rather than regarding social media data analysis as a surrogate for traditional methods of inquiry, we need to consider how it might augment, and be augmented by, traditional social research methods. They argue that such analysis is:

> '… distinctive in capturing naturally occurring or 'user-generated' data at the level of populations in real or near-real-time. Consequently, it offers the hitherto unrealisable possibility of studying social processes as they unfold at the level of populations as contrasted with their official construction through the use of 'terrestrial' research instruments and curated data-sets' (2013: 245).

This is not to ignore the difficulties in using social media data. On the contrary, it suggests 'a re-orientation of social science around new kinds of objects, different kinds of populations and alternative techniques of analysis' (Edwards *et al*., 2013: 248) alongside more conventional research methods.

Social scientists' and computer scientists' uses of social media data
Tinati *et al*. (2014) highlight that there is substantial research going on using social media data, however, much of it is undertaken by computer scientists. In comparison, relatively few social scientists are analysing social media data such as Twitter (although this is growing, particularly among postgraduate students), and those who are tend to rely on traditional social science methods (e.g., content analysis and the analysis of small qualitative samples) rather than using methods that utilise the strengths of social media, notably the scale and dynamic nature of so-called 'big data', as well as the scope to approach the data as representing 'networks' rather than isolated statements or individuals. Felt (2016) reviewed communication journal articles and found that the vast majority of social scientists were using conventional methods for analysing social media data, with relatively few making use of data analytics, such as tools for data visualisation. She argued that social scientists need to make better use of these tools, which enhance data analysis, while also drawing on qualitative methods and social science theory as an antidote to atheoretical data analytics undertaken by corporations. Similarly, Zelenkauskaite and Bucy (2016) have highlighted the growing 'scholarly divide' between many social scientists and those who have the technical skills and access to analyse and fully exploit 'big data'.

At the risk of oversimplification, social scientists, computer scientists and commercial social media analysts are likely to have different (although not necessarily mutually exclusive) research interests,

goals and approaches. For social scientists, the interest in social media analysis is both in terms of its nature as a form of social life and in terms of how it represents, facilitates and interacts with other aspects of the social world. As explained by Tinati *et al*. (2014), computer scientists may have greater interest in the technical aspects of social media, such as data processing, aggregation and storage, but their research includes social dimensions, such as examining friend networks, political views and communication flows. Commercial social media analysts are likely to be directed by corporate interests, including customer engagement and public relations management (Fan & Gordon, 2013).

Businesses also have a major stake in the use and analysis of social media data. The types of analysis and uses are wide ranging, including: sentiment analysis to track the public perception of an organisation and its products or services; engaging with consumers to involve them in the co-creation of products; analysis of the context or potential market for a product; evaluating the effect of advertising campaigns; and identifying and responding to consumer complaints (Fan & Gordon, 2013). In response to this, a range of companies have developed to offer professional social media analysis services to corporations and governments. The companies draw on and create technological innovations to respond to the business needs of these corporations. Such companies include BrandWatch, Klout, HootSuite and SoDash and have been responsible for innovations such as data aggregation tools, reporting dashboards, influence-scoring, and the development of sentiment analysis (in parallel to similar work in academia). As such, technology companies could make ideal partners for collaborative research on social media. However, the business needs that they serve differ from the interests of social scientists in several important ways. Most obviously, the corporations will generally have an interest in the public response to and market potential for their services and products, rather than social issues as such (except to the extent that they directly relate to their business concerns). Moreover, the data mining techniques used are often (although not exclusively) atheoretical, being data driven rather than theory driven. Therefore, any collaboration between social scientists and the social media analysis industry is likely to involve social scientists reconsidering, revising or changing their standard approaches to data collection and analysis, while social media analysts may need to alter the way they typically consider their topics of analysis and the way these are theorised. As suggested by Barry *et al*. (2008), such interdisciplinary research has the potential to alter both disciplines, through emphasising the social within social media, broadening the analysis beyond corporate concerns, while also forcing social scientific approaches to respond to the way technological advances have changed social relations.

Research on social media and social change
Commentators express differing views about the usefulness and significance of social media. Some have asserted that sites like Twitter represent a significant 'demotic turn'; ordinary people are able to break news, produce media content and voice opinions and because of this, it is a democratising

technology (e.g. Murthy, 2012). Not only are a wider variety of views able to be expressed and communicated than ever before, but social media may have contributed to a change in the way that we communicate with each other. However, others have highlighted that the democratising and transformational potential of social media may be more an ideal than a reality. For example, Cook and Hasmath (2014: 989), in a study of the 'Slut Walk' Facebook web pages, concluded that while participants on the Facebook pages 'engaged in the freeing environment of online interaction, they nevertheless remained situated in the social context of their everyday lives'. The Slut Walk movement's projects of resistance and inclusiveness were, they argue, undermined by the inability of participants to acknowledge the intersectionality of subjects' identities; by fore-fronting gender, and indeed a particular representation of gender, patriarchal and neoliberal forces were perpetuated.

Whilst there may be disagreement about social media's power to bring about social change, there is increasing academic interest in making use of social media for analysis purposes. One such analysis by Procter *et al*. (2013) used the August 2011 riots in England as its focus of attention. They conclude that social media communications could not be regarded as accurate representations of popular sentiment about the riots and their causes, because of the prevalence of misinformation, pranks, rumour and sarcasm. Others are less pessimistic. For example, Australian researchers Harrington, Highfield and Bruns (2013: 405) argue that Twitter has the potential to catalyse audience discussion and interaction; it has become 'an important backchannel through which such social activity is sustained and made more widely visible'. Procter *et al*. conclude that there are substantial methodological challenges to the use of Twitter as a source of social research, and these, they assert, demonstrate that computational tools can only go so far; 'human expertise is essential for robust interpretation and analysis' (Procter *et al*., 2013: 209).

A common charge against Twitter is that it constitutes an 'echo chamber', meaning that like-minded people merely recycle information and views among themselves, rather than the Twittersphere constituting a space in which people with different backgrounds and attitudes communicate with, and learn from, each other. Barberá *et al*. (2015) explored this issue by analysing the extent to which retweets moved across the political spectrum or tended to stay among people of similar political persuasions. They found that political topics (e.g., the 2012 American presidential election campaign) were more polarised than non-political topics (e.g., the Academy Awards), although some incidents became more polarised over time, such as when a shooting incident shifted to become a debate on gun control.

Social media research ethics
It is important to note that developments in social media and big data research raise important ethical questions. While much social scientific research relies on informed consent from participants, large

scale social media research is unlikely to involve this. As explained by Eynon, Fry and Schroeder (2011), there is a general lack of guidance regarding online research ethics, and while publically available information may be regarded as 'fair game', it raises the issue of 'privacy in public'. That is, to what extent could social media users expect their online comments to be used and cited in research? Moreover, to what extent should researchers work to maintain confidentiality and anonymity even where comments are available and findable online? Moreno, Goniu, Moreno and Diekema (2013) suggest that the nature of the content needs to be considered, in terms of the extent to which it constitutes 'private' information, and moreover that direct quotes should not be used where it would be possible for other people to search for and find the individual being quoted. Clearly, the ethical dimensions of social media research are complex and evolving, and cannot be discussed in full here, but care and consideration are required in the undertaking and presenting of research.

The project

Feminism online

Considerable attention has been given to feminism in academic literature over the last fifteen years or so. While much of the literature from the 2000s bemoaned the abeyance, if not incipient demise, of feminism (e.g., Bagguley 2002; Mackay 2008; Nash 2002), more recent studies have highlighted that although feminism is different today, it has in no way disappeared. On the contrary, social media has opened up significant new spaces for feminist debates and resistance; it is even suggested by some commentators that 'fourth wave' feminism (as it has been termed) was born on social media (Banyard 2010; Leupold n.d.; Solomon 2009), following on the legacy of three previous 'waves' of feminist activity (Phillips & Cree, 2014). This new social media-based feminism has been characterised as intolerant of 'isms' and inclusive of diverse sexualities and cultures; it reflects the popularity of intersectionality as a theoretical frame for analysis and has created a 'call-out' culture in which sexism or misogyny are challenged (Munro, 2013: 3).

Research questions

Our research was exploratory, both in terms of the focus of the study and the methods that were used. Our primary research aim was to examine how feminism and debates relevant to feminism were manifested on social media, specifically on Twitter. Our more specific research questions were: (1) what is the nature and extent of feminist activity on Twitter? (2) is it possible to identify and categorise activity on social media within topic clusters, particularly related to the different waves of feminism (as well as anti-feminism)? and (3), if so, what can we learn from an analysis of these categories regarding the nature of contemporary feminism? A further aim of the project was to gain insight into the different disciplinary approaches and analytic methods, particularly in terms of the collaboration between social scientists and computer scientists.

Methods

The methods involve training an AI classifier to auto-tag tweets. We did this by manually tagging a small segment of the data-set (0.4%) and giving this to the AI, from which it learned statistical patterns that allowed it to tag the remaining 99.6% of the data-set, and at the same time, test the accuracy of the results. Technical details of the process are presented in the appendix.

Our first task was to identify a set of key words and phrases that might indicate that a statement was being made about feminism. This process took a number of iterations, as we refined our list. We wanted to include explicit references to feminism ('feminist', 'feminism'), topics related to gender equality ('gender equality', 'equal pay'), gender identity ('cis', 'trans person'), issues related to gender relations and the treatment of women ('violence against women', 'everyday sexism', 'rape'), antagonistic terms related to feminism ('feminazi', 'hate men'), and topics relevant to feminism ('sexwork', 'pornography', 'fgm', 'abortion'). We included slang terms that might be used online ('tranny'). The eventual list was as follows: feminism; feminist; feminazi; everydaysexism; sexism; sexist; sexwork; pornography; rape; rape; rapist; misogyny; sexualisation; intersectional; exploitation; "violence against women"; "violence to women"; "violence against men"; "everyday sexism"; "gender equality"; "gender inequality"; "women's lib"; "womens lib"; "zero tolerance"; "hate men"; "male gaze"; "equal pay"; abortion; "domestic violence"; "domestic abuse"; fgm; cis; "trans girl"; "trans man"; "trans person"; "trans people"; trannie; tranny; objectification; "mens rights"; "men's rights"; "mensrights".

Armed with this list of key words, the data analysis team members then filtered the one million-plus messages a day, gathered between 1 December 2012 and 31 May 2014. This is known as 1% Twitter 'sprinkler' data. We applied a scaling factor of 100 to obtain estimates of total volume (that is, we extrapolated from our 1% sample by multiplying this by 100 to estimate the 'true' scale of the tweets). The sample data is created in an unbiased way and with a clear relation between the sample and the full dataset. Although it is only 1%, this alone represents a large volume of data (about 1 million messages per day in total, from which we then distilled the ones relevant to our topic). A survey requires only about 1,000 samples to give reliable results (i.e. to accurately estimate proportions with high confidence), and we have much more data than that. So we believe, this sample allows for accurate analysis to be made. For a part of the sampling period (8 of the 18 months), our data supply from Twitter dropped to a lower rate, from 1% to 0.1% due to connection issues between our server and Twitter's. We do not think this affected the randomness of the sample; we corrected for this in the analysis by applying a higher scaling factor to that period. Our eventual dataset was made up of approximately 500,000 tweets that featured words from our list.

Once our data-set had been created, we set about coding the tweets by tags. After agreeing a tag-set, each of the team members took a day in one month in 2013 (May to September inclusive) and set about coding 100 tweets from that day. Our tag-set was as follows:

- Topic: abortion, children, rape, sexism, violence, employment, female genital mutilation / male circumcision, intersectionality, media portrayal, meta-discussion, pornography, race, sex work, sexuality, transgender rights, off topic
- Position: pro/anti feminism, or neutral
- Wave of feminism: 1st wave vs 2nd wave vs 3rd wave vs 4th wave
- Type of anti-feminist: mens-rights, individualist, misogyny
- Tone of tweet: angry vs calm vs joking/humorous
- Message type: rebuke (tweets that shut conversation down) vs converse (conversational-chatting-engaging) vs explain-inform-proclaim (giving information / expressing an opinion) vs joke vs question.

When we subsequently reviewed the outcome, it was evident that more human tagging was needed for the AI tagging to work well and so one team member coded another 300 tweets. At the same time, we identified a number of key topics that were emerging in the data and that therefore merited further investigation. Two team members did specific searches relating to the topics of rape, abortion and female genital mutilation, identifying another 1000 tweets in this way. Following this, tests for accuracy were carried out, and when aggregating across messages, error-correction was applied to allow for much more accurate overall figures.

Findings

Interestingly, the tweets that emerged in our study were surprisingly diverse. It is fair to say that a very wide range of views about feminism was expressed, with those for, against and neutral towards feminism; and angry and calm, jokey and serious, one-sided and more nuanced tweets. Of course, we cannot claim that this representation is, in any sense, illustrative of the views held in the wider population. On the contrary, it can only indicate who was on Twitter and what they chose to express. As mentioned above, Twitter users tend to be younger and more likely to be based in professional occupations than the general population (Sloan *et al*., 2015); attitudes and engagement regarding feminism vary along age and class dimensions, and online feminism in particular is dominated by younger people, meaning it is likely to differ from the views and experiences of older people (Munro, 2013).

We cannot even know for certain how many people were tweeting about feminism, because people send multiple tweets, and the frequency of sending varies. It is possible to estimate the number of

people, based on the per-person probability-of-capture over the period (Winterstein, 2011), but we did not feel that this was essential information for this project. We can, however, make two observations about our data-set. Firstly, we were able to identify the countries of origin of the tweets. As shown in Figure 1, most tweets about feminism came from the US and UK, with smaller numbers from Australia and Canada. This is not, of course, unexpected, because we used English keywords to compile the dataset, plus the US and the UK are the largest and 4th largest users of Twitter.

Figure 1: World map of proportion of tweets by country of origin.

Secondly, it was clear that there were spikes in the amount of tweeting about feminism over our selected period, as shown in Figure 2. These tallied with what was happening in the social and political world beyond Twitter – specific campaigns, media projects, news-stories that flared up for a time and then settled back down again. So the spike in mid-2013 can be explained by the launch on 27th April of a petition by Caroline Criado-Perez to demand that the Bank of England reconsider its decision to remove the only woman (Elizabeth Fry) from the £5 note; the filibuster on 25th June by Texas senator Wendy Davis to stop a law to close down abortion clinics; the speech on 12th July by Malala Yousafzai at the United Nations in New York, where she argued for young women's right to education; and a blog by Feministing on 22nd July, which railed against sexual assault and rape culture. Feministing filed a complaint against Yale University in connection with a specific incidence of rape, and US President Barack Obama subsequently formed a taskforce to tackle what was presented as an epidemic of sexual violence on college campuses.

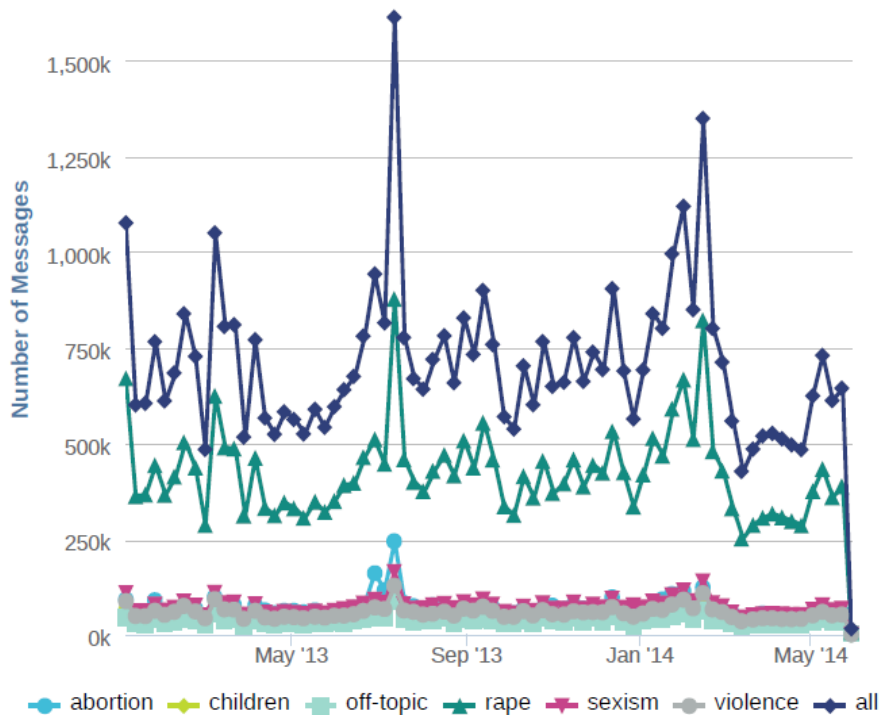Figure 2: Feminism tweets by topic over time.

Figure 2 also demonstrates what is the single most striking, and in our shared view, depressing finding of our project, that is, the concentration of tweets on rape, by far and away the biggest topic of conversation that we found in our study, with abortion coming a long way behind in second place. Rape is something that is referred to and discussed on Twitter in a variety of ways. Some of this includes discussion and comment on the nature of sexual violence and appropriate justice responses to it. Other aspects demonstrate a casual and minimising attitude towards sexual violence, as well as direct threats and abusive behaviour. We encountered a high level of casual misogyny on Twitter, and 'jokes' about rape (for example, 'I'm going to rape my fridge this weekend'), as well as many explicitly anti-women 'jokes', were commonplace. This mirrors a study by Bartlett *et al.* (2014) on misogyny among UK Twitter users which found that, during a six-week period, there were approximately 2.7 million tweets using the word 'rape'. Their analysis suggested that about 40% of uses were 'serious' or regarding news, 29% of uses were metaphorical or as 'jokes' (like our example above), 12% were threatening or abusive (the remaining 19% were categorised as 'other').

Although we were able to identify topics of concern to feminism, and feminist spikes in particular, beyond this, coding in relation to both tone and message type proved difficult. For example, an indicator of tone might be the words used, and exclamation marks and question marks may provide cues which indicate that something else is going on – but what? It could be signifying whole-hearted agreement or even the exact opposite, if the tweeter is expressing cynicism or sarcasm. Similarly, message types proved to be illusive. What had seemed clear in discussion amongst the project team members became less certain when faced with a specific tweet. For example, what do we make of a

tweet that says, 'I love a feminist bitch'? Exactly what is the tone and message being conveyed here? And if we could not agree, what hope was there for training a machine (the AI classifier) to do likewise?

Coding challenges were not, however, related only to tone and message type. Tagging by wave of feminism proved impossible, not least because the team members could not be certain what was meant by the different waves. It also, however, demonstrates a more fundamental issue: that feminists may hold attitudes that might be regarded as sympathetic to different approaches to feminism at the same time. We will explore this more fully in the discussion section below. We did have some success in asking the classifier to make sense of the speaker's position in relation to feminism (pro/anti/neutral) to a reasonable degree. When predicting for individual messages, different camps could be distinguished with reasonable accuracy: pro-feminist: 78% correct (that is, agreeing with the human tagger), anti-feminist: 66% correct.

Discussion and conclusion

An important observation to make at the outset is that the wider world of Twitter is nothing like our personal Twitter feeds! Perhaps this should have been obvious before we looked at our data, but maybe the online and public nature of Twitter masks the extent to which we select who we follow and therefore the fact we see only a tiny slice of tweets that are strongly influenced by our personal preferences and social or professional networks. Beyond this, it was evident that feminist activity on Twitter is diverse. Our results reflected the dominance of the English-speaking world on Twitter, and at the same time, demonstrated spikes in pro-feminist dialogue at times of relevant 'real world' events, showing times when Twitter was being used as a tool to engage public opinion and through this, to create social change. We also identified the less life-affirming side of Twitter, with its casual sexism, misogyny and alarming attitudes towards rape and sexual violence.

We have stated that the difference between pro, neutral and anti-feminist views was very clear on Twitter, confirming Barberá *et al*.'s (2015) assertion that debate is often polarised on Twitter. It was not possible, however, to go beyond this to distinguish between different 'waves' of feminism. This supports Knappe and Lang's (2014: 3) claim that feminist waves are 'not neatly delineated temporal or developmental stages but rather stand for overlapping and intersecting periods of activism'. So, for example, although feminist researchers may be writing about the emergence of a fourth wave of feminism, organisations representing second and third wave feminisms still predominate in the public domain. Many women's organisations are, they argue, hybrids, using approaches and tactics that are familiar to either the second or third waves, or even both. Moreover, Pruchniewska (2016) suggests that the notion of 'waves' of feminism may be both limited and divisive, so that it becomes more productive to focus on feminist practices. One such feminist practice to emerge in our study was the

discussion of rape, highlighting not only cultural attitudes towards sexual violence, including various attempts to influence the criminal justice response, but also the ways that online abuse is carried out.

Our study raised as many methodological issues as it did content-specific ones. While it was easy to capture large amounts of data, the challenge was how to code and analyse this, beyond simple descriptive statistics or inspecting small samples. The use of AI showed some potential, although this was limited by the inherent challenges of coding. Tagging, like all coding in research, is inherently subjective. We developed a set of tags over several workshops, refining the tags by testing the use of them on the data. Several of our concepts, so clear in discussion, were difficult to apply to the actual message data, which drove a simplifying of our analysis, as well as an awareness of the shaded diversity of the discourse. In our case, we did not need to run the AI to establish that it is difficult, if not impossible, to code tweets by 'wave' of feminism; this was evident from the experience of the coders themselves. Actually, the discussions we had among ourselves regarding the nature of feminism and gender relations were fascinating and showed how complex this topic is. Further research should ensure that sufficient time is allowed for the researchers to discuss and clarify the concepts and tags before applying them, and once they have been applied, to discuss their application to see if they make sense. If and once tags have been agreed, then they should have written definitions.

A further limitation was the amount of data we tagged. To give really robust statistical results on topics, message-types, and other tagsets aside from pro-/anti-feminist position, we would like considerably more tagging. With more tagged data, the AI would be better able to recognise each case, plus the error-correction would give tight confidence levels. The mathematics of this is similar to confidence intervals for opinion surveys. The confusion matrix gives percentages based on the sample tagged for each tag – if those samples were around 1,000 per tag, then we would get tight confidence intervals. A limitation of the AI classifier is that it will only apply one tag from each tagset. So it cannot properly account for a message which should be tagged as "rape" and "abortion". One way to increase the amount of tagging is to use crowdsourcing. Although this could introduce further error into the process, if the guidance is clear it could increase the amount of tagging dramatically, and this approach has been shown to be effective (Burnap & Williams, 2015). Moreover, it would be possible to use inter-rater reliability tests to check the extent to which different coders agree on their tagging, and remove or adjust the results where reliability is particularly low for specific tags or coders (Sloan *et al.*, 2015). To help ensure the reliability of the process, it would be a good idea to use random sampling from the data, which is something we did not do. It would also help to use a process to remove bots (e.g., Chu *et al*., 2010).

In terms of taking forward a research agenda for exploring feminism online, there are a number of ways forward. For instance, a future project could use a more restricted set of search terms to get a better sense of who is talking about topics more explicitly related to feminism (e.g., 'feminism', 'feminist', 'mensrights', 'genderequality'). Researchers could identify accounts that describe themselves as feminist or anti-feminist and check to what extent the coding of their tags matches their description. Following Barberá *et al*. (2015), researchers could explore the extent to which tweeting and following does or does not circulate across the feminist / anti-feminist divide (and, indeed, the extent to which such a 'divide' may be said to exist). Data mining techniques could be used to explore what geographical and demographic information tells us about pro- or anti-feminist tweets and accounts (see Sloan *et al*. 2013, 2015). The tweets could also be analysed in relation to sentiment to explore people's attitudes towards key topics. It would also help to identify key 'real world' events relevant to feminism (e.g., controversial rape trials, the introduction of policies relevant to gender equality) and map tweet frequencies over time to examine the extent to which discussion peaks around key events (e.g., see Barberá *et al.,* 2015; Vis & Goriunova, 2015). Such analysis should explore how debates change over time. Discussion online can actually reduce polarisation of views over time (Barberá, 2014) and the public response to issues is responsive to certain (social) media events (Vis & Goriunova, 2015).

As for the potential for partnership between social scientists and those working in the social media analysis industry, we believe that each has a lot to offer the other, although such collaboration is not without its challenges. Social scientists have much to learn about the nature of social media and the ways in which it can be analysed, drawing on technological innovations. This is likely to require questioning, revising or even abandoning conventional research methods. And computer scientists have much to learn from social science. Perhaps the most persistent finding in our project was that in spite of a shared commitment to feminist values and a desire to challenge gender oppression, we struggled as a research team to understand one another. If we are to truly learn from each other in the future, we will have to be willing to open up all our assumptions for scrutiny, and to appreciate that the very language that we use needs to be interrogated. In doing so, we may even learn things that we did not know we didn't know.

Appendix

Description of the AI classifier process

We used a "Bayesian classifier", which learns patterns for each tag, then when presented with a new message, sees which patterns fit best. More specifically and technically, we fitted generative models related to the commonly used n-gram technique. The AI starts with a certain strength of prior belief that each tag is equally likely to produce a word or phrase, then it learns to prefer one tag over another from observed frequencies. After training, when it examines an untagged message, it considers how words and phrases fit the model for each tag. This produces a posterior distribution of what tag to assign the message. The assignment is done by considering the relative likelihood of each tag. For this, we group the tags into "tagsets" of mutually exclusive tags. A given message would receive at most one tag from each tagset (the AI was also allowed to pass if very unsure). For example, a message could be tagged "pro-feminism" + "employment" + "question". As described earlier, designing the tagsets to be representative and mutually exclusive was one of the challenges of this project.

Having tagged the messages, the system then counts the results, producing an aggregate report. Our AI's aggregate reporting includes an error-correction step. It analyses the type of errors it makes during training (using what is called a confusion matrix, an example of which is shown below). The observed accuracy is combined with a prior belief that all types of mistake are possible, to give an expected level for each specific kind of error – and hence an error-correction calculation. This prior is known as a Dirichlet prior, and involves adding a fixed 'pseudo-count' to each observed count. This mitigates us from undue confidence in a categorisation and has the effect of "blurring" the final results. Error correction used a pseudo-count of 10 for each tag/tag combination, which is both lower than ideal, and sufficiently high compared to our training data (e.g. for the topic tagset with 7 tags this amounted to almost 500 pseudo-counts, versus our manual training data for that tagset of 740 actual counts for that tagset) that it applies a heavy blur.

Example confusion matrix

In this illustrative mock example, there were 100 messages about cats and 50 about dogs, the AI was mostly right when it did say "cat", but it missed many cats and was considerably less reliable when it said "dog". Knowing this pattern of errors lets us correct for them on aggregate.

human tag ↓ / AI tag →	"cat"	"dog"	Sensitivity
"cat"	70	30	70%
"dog"	25	25	50%
Accuracy	84%	55%	

The algorithm is commercially sensitive; researchers wishing to replicate the study should contact the third author for further details. Our dataset is available for download here:

http://www.winterwell.com/feminism.php

References

BAGGULEY, P. (2002) 'Contemporary British feminism: A social movement in abeyance?' *Social Movement Studies*, Vol. 1, Issue 2, p. 169–185. [doi.org/10.1080/14742830220000010664]

BANYARD, K. (2010) *The equality illusion.* London: Faber and Faber.

BARBERÁ, P. (2014). 'How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S.' LXXIII Congress of the Midwest Political Science Association, 44.

BARBERÁ, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). 'Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?' *Psychological Science*, Vol 26, Issue 10, p. 1531–42. [doi.org/10.1177/0956797615594620]

BARRY, A., Born, G., and Weszkalnys, G. (2008). 'Logics of interdisciplinarity', *Economy and Society*, Vol 37, Issue 1, p. 20–49. [doi.org/10.1080/03085140701760841]

BARTLETT, J., Norrie, R., Patel, S., Rumpel, R., and Wibberley, S. (2014). *Misogyny on Twitter.* London. Retrieved from <http://www.demos.co.uk/files/MISOGYNY_ON_TWITTER.pdf?1399567516>

BURNAP, P., and Williams, M. L. (2015). 'Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making', *Policy and Internet,* Vol 7, Issue 2, 223–242. [doi.org/10.1002/poi3.85]

CHU, Z., Gianvecchio, S., Wang, H., and Jajodia, S. (2010). 'Who is Tweeting on Twitter: Human, Bot, or Cyborg?' Proceedings of the 26th Annual Computer Security Applications Conference on - ACSAC '10, 21. [doi.org/10.1145/1920261.1920265]

COOK, J., and Hasmath, R. (2014). 'The discursive construction and performance of gendered identity on social media', *Current Sociology,* Vol 62, Issue 7, p. 975–993. [doi.org/10.1177/0011392114550008]

EDWARDS, A., Housley, W., Williams, M., Sloan, L., and Williams, M. (2013). 'Digital social research, social media and the sociological imagination: surrogacy, augmentation and re-orientation', *International Journal of Social Research Methodology*, Vol 16, Issue 3, p. 245–260. [doi.org/10.1080/13645579.2013.774185]

EYNON, R., Fry, J, and Schroeder, R. (2011). 'The Ethics of Internet Research'. In N. Fielding, R. M. Lee and G. Blank (Eds), *The SAGE Handbook of Online Research Methods* (pp. 22-14). London: Sage. [dx.doi.org/10.4135/9780857020055]

FAN, W., and Gordon, M. D. (2013). 'Unveiling the Power of Social Media Analytics', Communications of the ACM, 12 (JUNE 2014), p. 1–26. [doi.org/10.1145/2602574]

FELT, M. (2016). 'Social media and the social sciences: How researchers employ Big Data analytics', *Big Data and Society*, Vol 3, Issue 1, p. 1–15. [doi.org/10.1177/2053951716645828]

HARRINGTON, S., Highfield, T. and Bruns, A. (2013) More than a backchannel: Twitter and television', Participations. *Journal of Audience & Reception Studies*, Vol 10 Issue 1, p. 1-5.

KNAPPE, H., and Lang, S. (2014). 'Between whisper and voice: Online women's movement outreach in the UK and Germany', *European Journal of Women's Studies*, Vol 21, Issue 4, p. 361–381. [doi.org/10.1177/1350506814541643]

LEUPOLD, J. (n.d.). *Fourth Wave Feminism*. Retrieved from <http://journalism.nyu.edu/publishing/archives/portfolio/leupold/>

MACKAY, F. (2008). 'The state of women's movement/s in Britain: Ambiguity, complexity and challenges from the periphery'. In S. Grey and M. Sawer (Eds.), *Women's Movements: Flourishing or in abeyance?* (pp. 17–32). London: Routledge.

MORENO, M. A., Goniu, N., Moreno, P. S. and Diekema, D. (2013). 'Ethics of Social Media Research: Common Concerns and Practical Considerations'. *Cyberpsychology, Behavior, and Social Networking*, Vol 16, Issue 9, p. 708–713. [DOI:10.1089/cyber.2012.0334]

MUNRO, E. (2013). *Feminism: A fourth wave?* Retrieved from <http://www.psa.ac.uk/insight-plus/feminism-fourth-wave>

MURTHY, D. (2012). 'Towards a Sociological Understanding of Social Media: theorizing Twitter', *Sociology*, Vol 46, Issue 6, p. 1059–1073. [doi.org/10.1177/0038038511422553]

NAGLER, J., and Tucker, J. A. (2015). 'Drawing Inferences and Testing Theories with Big Data', *PS: Political Science and Politics*, Vol 48, Issue 1, p. 84–88. [doi.org/10.1017/S1049096514001796]

NASH, K. (2002). 'A Movement Moves . . . Is There a Women's Movement in England Today?', *European Journal of Women's Studies*, Vol 9, Issue 3, p. 311–328. [doi.org/10.1177/1350506802009003377]

PHILLIPS, R. and Cree, V. E. (2014). 'What does the 'Fourth Wave' Mean for Teaching Feminism in Twenty-First Century Social Work?', *Social Work Education*, Vol 33, Issue 7, p. 930–943.

POTTER, J., and Hepburn, A. (2005). 'Qualitative interviews in psychology: problems and possibilities', *Qualitative Research in Psychology*, Vol 2, Issue 4, p. 281–307. [doi.org/10.1191/1478088705qp045oa]

PROCTER, R., Vis, F., and Voss, A. (2013). 'Reading the riots on Twitter: methodological innovation for the analysis of big data', *International Journal of Social Research Methodology*, Vol 16, Issue 3, p. 197–214. [doi.org/10.1080/13645579.2013.774172]

PRUCHNIEWSKA, U. M. (2016). 'Working across difference in the digital era: riding the waves to feminist solidarity', *Feminist Media Studies*. [doi.org/10.1080/14680777.2016.1190045]

SLOAN, L., Morgan, J., Burnap, P., and Williams, M. (2015). 'Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data', *PLoS ONE*, Vol 10, Issue 3. [doi.org/10.1371/journal.pone.0115545]

SLOAN, L., Morgan, J., Housley, W., Williams, M., Edwards, A., Burnap, P., and Rana, O. (2013). 'Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter', *Sociological Research Online*, Vol 18, Issue 3.

SOLOMON, D. (2009, November 13). 'Fourth-wave feminism', *The New York Times Magazine*. Retrieved from <http://www.nytimes.com/2009/11/15/magazine/15fob-q4-t.html?_r=1>

TINATI, R., Halford, S., Carr, L., and Pope, C. (2014). 'Big Data: Methodological Challenges and Approaches for Sociological Analysis', *Sociology,* Vol 48, Issue 4, p. 663–681. [doi.org/10.1177/0038038513511561]

TWITTER (2017). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 3rd quarter 2017 (in millions). *Statista*. Retrieved from <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

URRY, J. (2000). *Sociology beyond Societies.* London: Routledge.

VIS, F., and Goriunova, O. (2015). *The Iconic Image on Social Media: A Rapid Research Response to the Death of Aylan Kurdi*. Retrieved from <https://www.scribd.com/doc/293236608/Visual-Social-Media-Lab-The-Iconic-Image-on-Social-Media-a-rapid-research-response-to-the-death-of-Aylan-Kurdi>

WINTERSTEIN, D. (2011). 'Mark and recapture estimation on Twitter'. Retrieved from <http://platypusinnovation.blogspot.co.uk/2011/09/mark-and-recapture-estimation-on-twitter.html>

ZELENKAUSKAITE, A. and Bucy, E. P. (2016). 'A scholarly divide: Social media, big data, and unattainable scholarship'. *First Monday*, Vol 21, Issue 5. [http://dx.doi.org/10.5210/fm.v21i5.6358]