Islamophobia on Twitter: March to July 2016

Carl Miller
Josh Smith
Jack Dale
Centre for the Analysis of Social Media, Demos

DEMOS

RESULTS

The Centre for the Analysis of Social Media (CASM) at Demos is conducting continuous research on hateful, xenophobic, anti-disability, anti-Semitic and anti-Islamic ideas and expressions on Twitter. This is part of a broad effort to understand the scale, scope and nature of uses of social media that are possibly socially problematic and damaging.

This short paper details recent results for the use of Twitter to share expressions which are identified as Islamophobic, derogatory and hateful. It covers a broad stretch of time, from the 22nd February 2016 to the time of writing, August 4th, but focuses especially on activity over the month of July 2016. For a discussion of how these Tweets were collected and analysed, see the methodology section of this report.

July 2016

Over July, we identified 215,247 Tweets, sent in English and from around the world, as highly likely to be hateful, derogatory, and anti-Islamic.[1] On average, this is 289 per hour, or 6943 per day. This is the highest monthly average since measurements began at the end of February. However, the rate of anti-Islamic activity on Twitter significantly changed over the month. Twitter is, in general, a real-time, reactive and event-specific platform, and most of the anti-Islamic activity identified was likewise linked to an event that had recently happened. The most significant increase was in the immediate wake of the terrorist attack in Nice, on July 14th, with another appreciable increase in the rate of anti-islamic expressions in the aftermath of the killing of Jacques Hamel in Normandy. The five most significant spikes are analysed in greater depth, below. This is to try to uncover the triggers, drivers and dynamics of anti-Islamic hatred online.

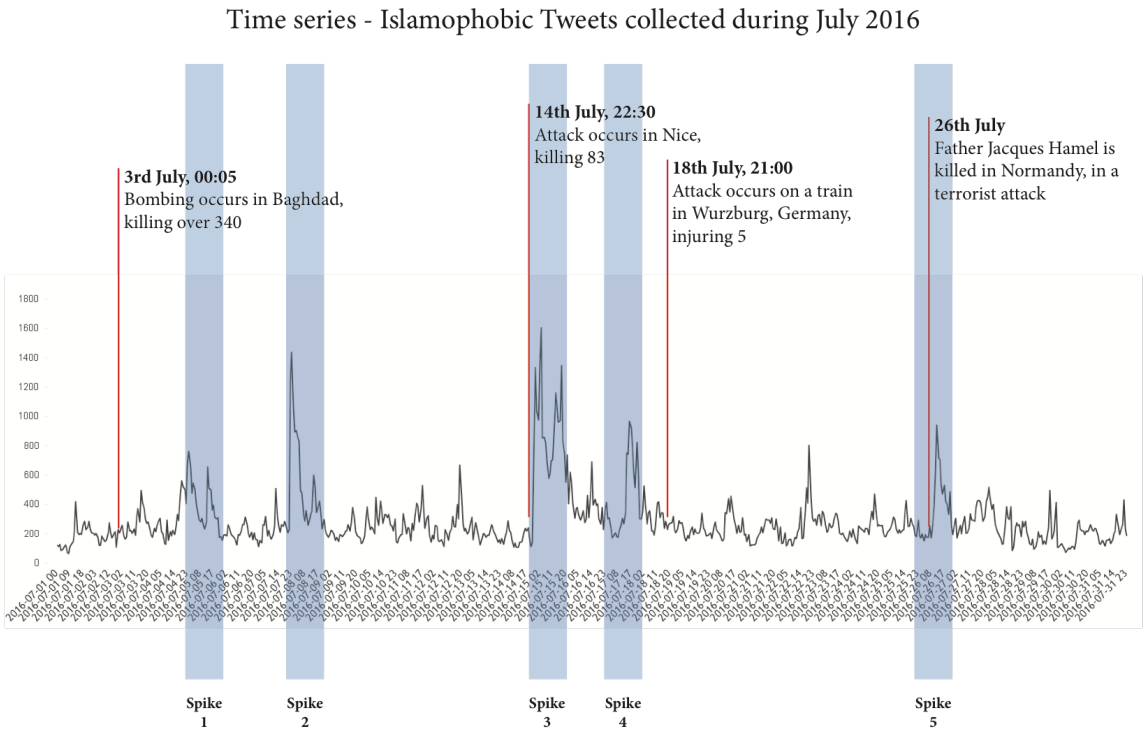


Figure 1 – Islamophobic Tweets collected during July 2016

[1] See methodology for more information on how this research was conducted.

Spike 1 – July 5th

We identified 9,220 Islamophobic Tweets on 5 July. It is difficult to identify one singular event that triggered this rise in Islamophobic language. One possible explanation is that this was 4 days after the 12-hour siege by IS militants in a café in Bangladesh. American, Italian, Indian, Japanese and Bangladeshi victims were among the 22 people killed in this attack. Furthermore, this day marked the end of Ramadan before the start of Eid al-Fitr, perhaps intensifying a global focus on Islam.

Examples of tweets from this day include: *"Nobody can stop Muslims committing jihad attacks any more than they can stop Buddhists meditating or Mormons knocking on people's doors"*; *"Morocco deletes a whole section of the Koran from school curriculum as it's full of jihad incitement and violence. The Religion of peace"*; and *"I fucking hate pakis."*

Spike 2 – July 8th

11,320 Islamophobic tweets were sent on 8 July. Again, it is difficult to attribute this rise to one specific event, though this was the day after the shootings in Dallas, U.S., in which Micah Xavier Johnson shot and killed 5 police officers, wounded 7 others and wounded 2 civilians.

Many tweets appeared to try and link this event to Islam. An example includes: *"Obama is a damn Raghead explains a lot"*.

Spike 3 – July 15th

21,190 tweets sent on 15 July were identified as Islamophobic. This was the day after the attack in Nice, in which an armed IS militant drove a truck through crowds of people celebrating Bastille Day; 84 people were killed and many more injured.

Tweets sent on this day focused on the attack: *"Sorry to hear about france- These muzzies just dont quit"*; "and *"Stop saying 'the majority are peace-loving'. Until the majority denounce every jihadi & turn them in, we are safer believing the evidence"*.

Spike 4 – July 17th

10,610 tweets were sent on 17 July. This was the day after an attempted military coup in Turkey failed; a faction within the Turkish Armed Forces organised themselves under the 'Peace at Home Council'. The Council cited the erosion of secularism as one reason behind the attempted coup. As such, some tweets commented on this: *"That's the end of Turkey. Another country ruined by Islam and its terrorist culture. Same shit happened to Persia. Now its Islamic Republic"*. Other tweets were more general: *"France's Islamic population is at 9.6%. 10% is usually when Jihad begins. They're starting early because the French are so weak"*; *"ALL this because of the muzzie in the White House"*.

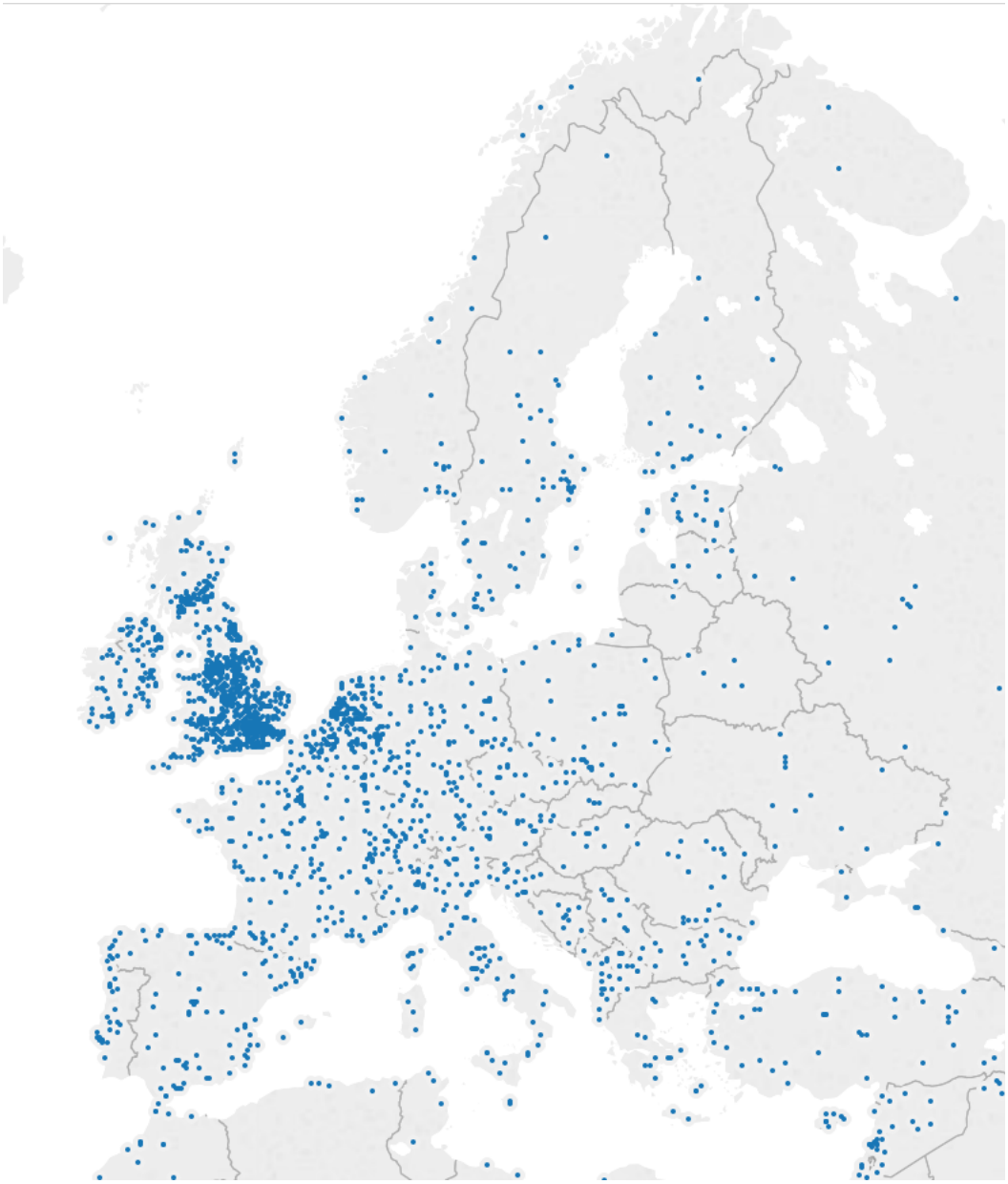
Spike 5 – July 26th

8,950 tweets were sent on 26 July. This was the day of the Normandy church attack, in which IS militants killed Father Jacques Hamel, and seriously wounded another.

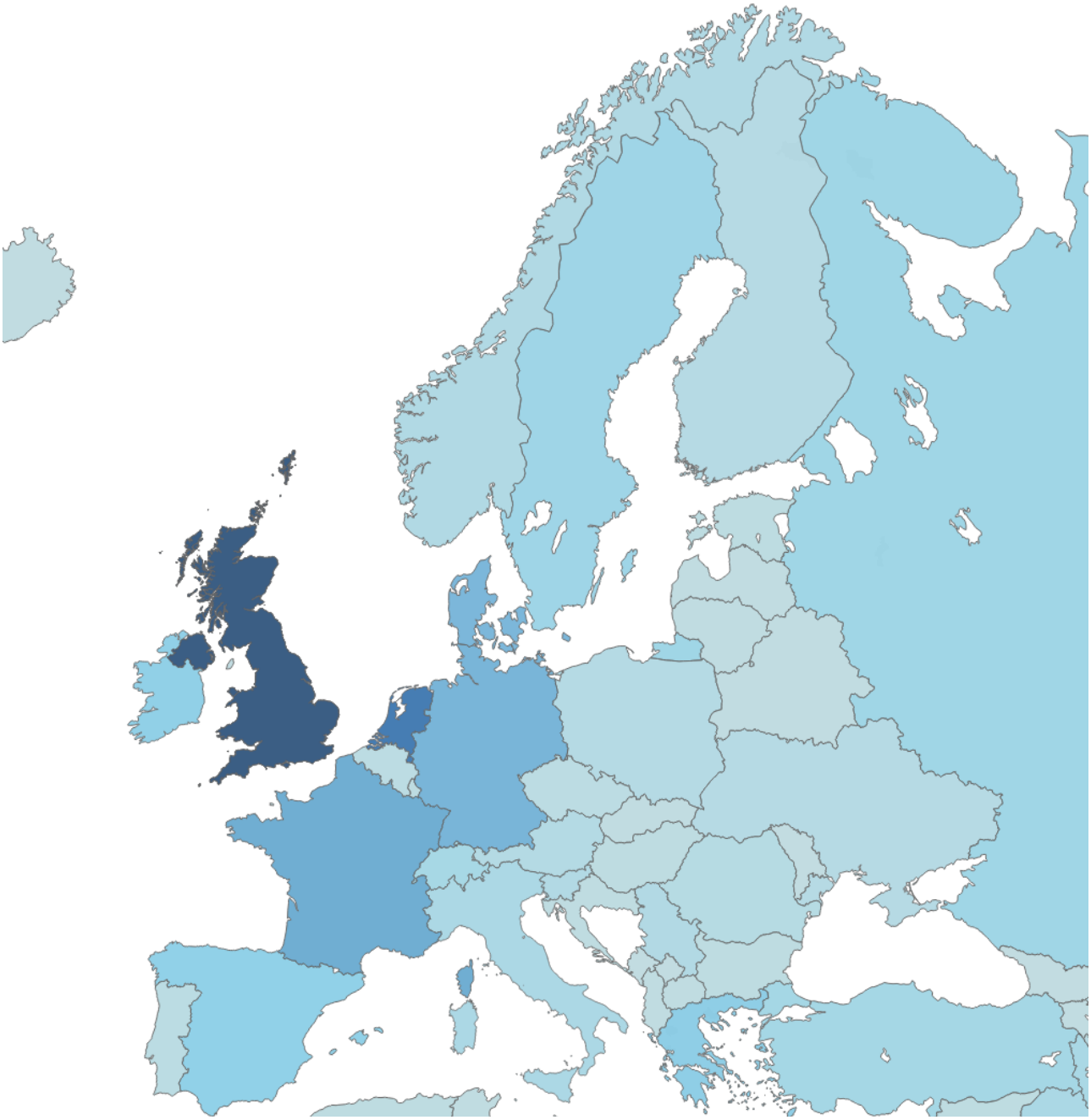
Tweets sent on this day comment on this attack: *"Normandy is reason 1488 that you should elect Marine Le Pen! > close borders > deport murderous Islam & Muzzies > deport the rioting negroes"*; *"So some sleazy scum committed jihad in the name of #Islam this time in Normandy but hey let's keep telling Muslims we love them"*; and *"Priest killed in #Normandy today by a Radical Islamic Terrorist yet Hillary says that Islam is peaceful! 1274 attacks this year=peaceful? Ok."*

GEOGRAPHY

The analysis was conducted only for anti-Islamic expressions in the English language. Consequently, the vast majority of the Tweets that could be located to Europe came from the United Kingdom. However, as the maps below illustrate, anti-Islamic Tweets were sent from every European Union member state, with other concentrations in the Netherlands, France and Germany.

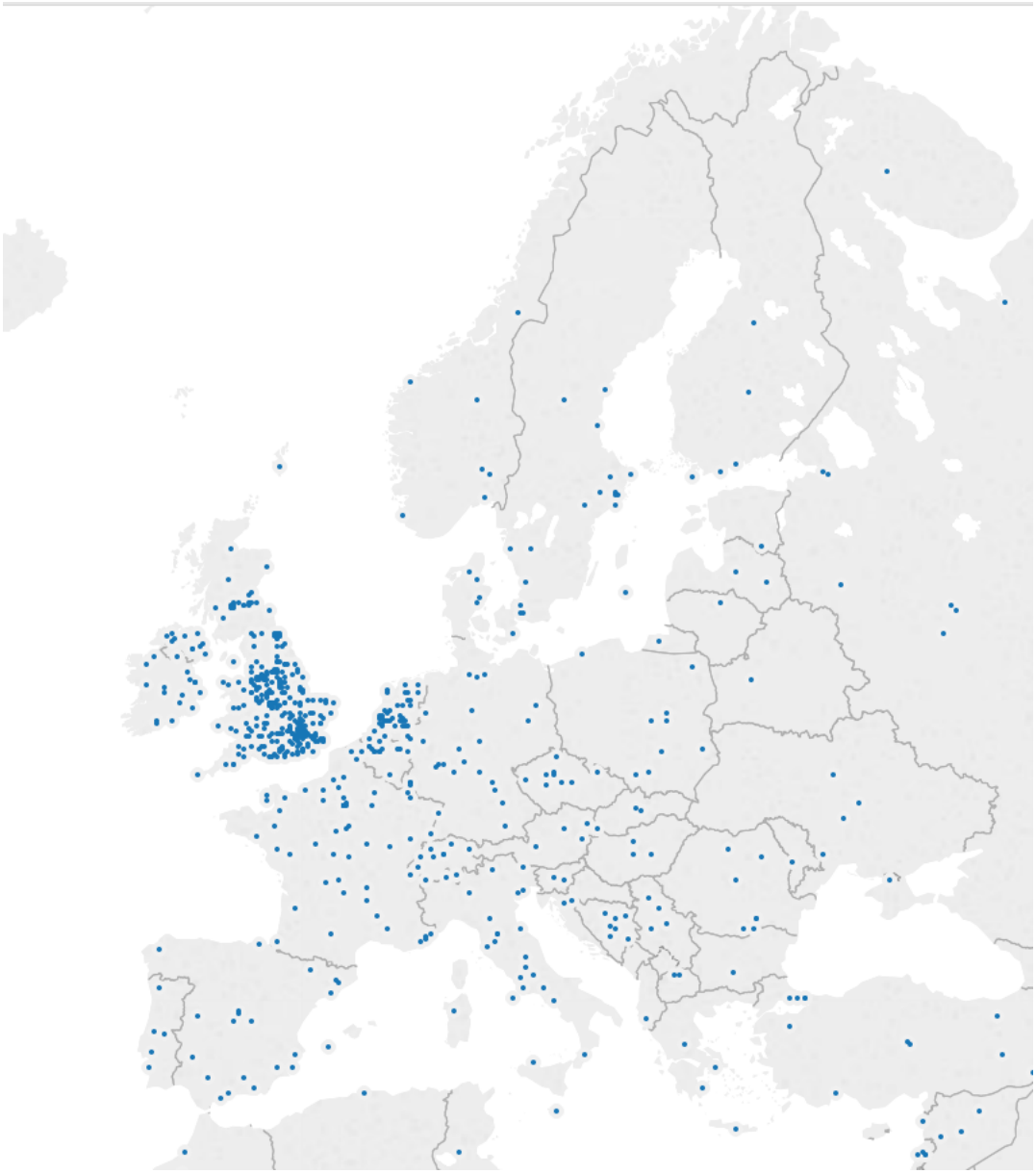


Geo-located anti-Islamic Tweets over July 2016

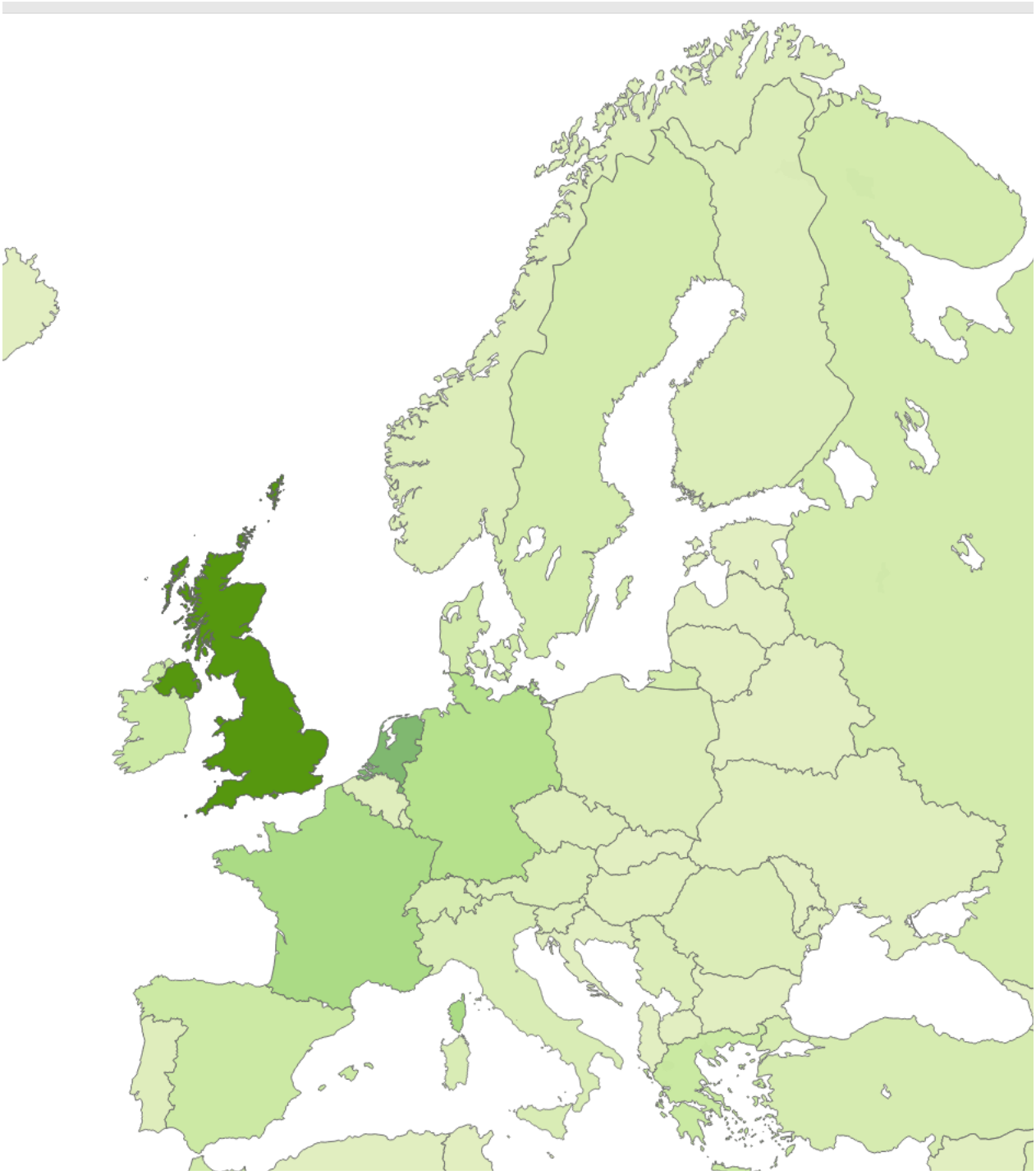


Country concentrations of geo-located anti-Islamic Tweets

Geographical concentrations of anti-islamic expressions on Twitter also varied across the month. The tweets sent in reaction to the Nice attack (spike 3, above), as shown below, had higher concentrations of anti-Islamic expressions in Holland and France than the reaction to the Turkish coup attempt (spike 4).

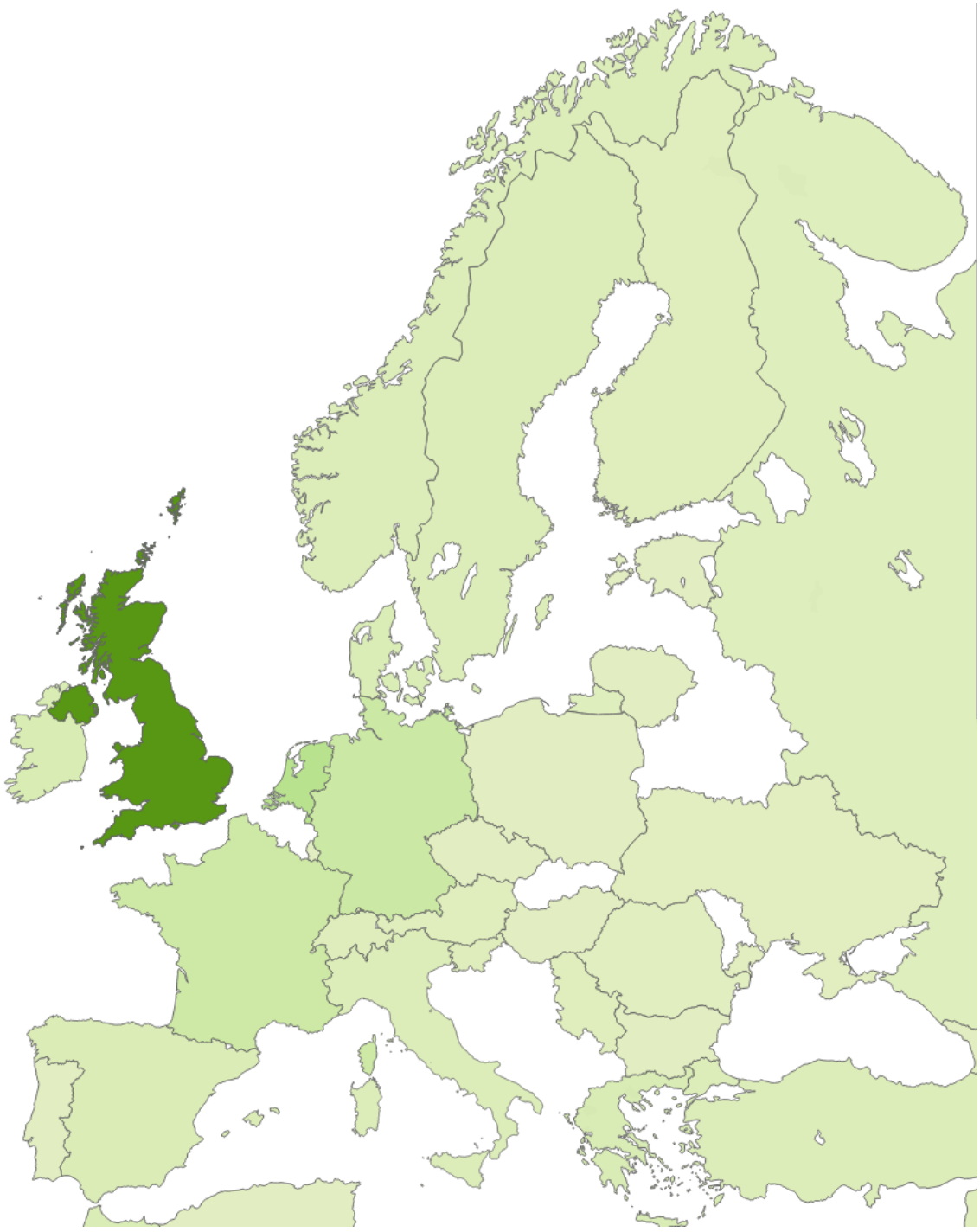


Geo-located anti-Islamic Tweets for 15th July



Country concentrations of geo-located anti-Islamic Tweets for July 15th

Geo-located anti-Islamic Tweets for 17th July



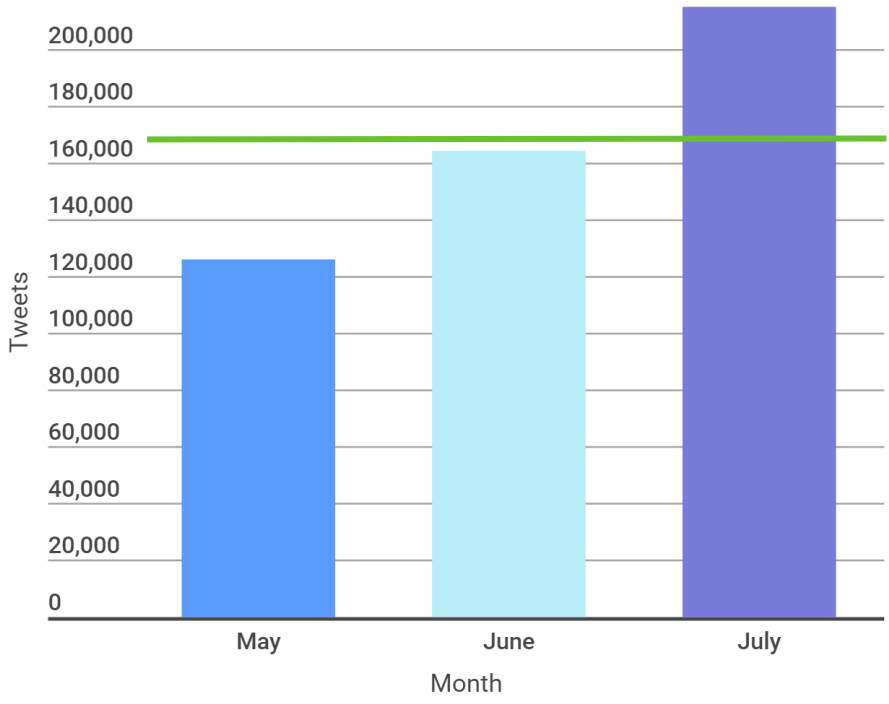
Country concentrations of geo-located anti-Islamic Tweets for 17th July

VOLUME

From the beginning of March to the end of July 2016, an average of 4972 Tweets were identified a day.[2] The rate dropped sharply between March (the month of the terrorist attacks in Brussels, and the subject of earlier Demos work in this area) and April, and has since then been increasing month-on-month.[3]

Month	Islamophobic Tweets sent per day (average)	% increase/decrease on previous month
March	5,024	N/A
April	2,512	-50%
May	3,985	+37%
June	5,480	+27%
July	6,943	+21%

July, with an average of 6,943 anti-Islamic Tweets per day, or 215,247 across the month, has the highest rate of anti-Islamic Tweets of any month analysed, and considerably above the monthly average (168,595) during this time.

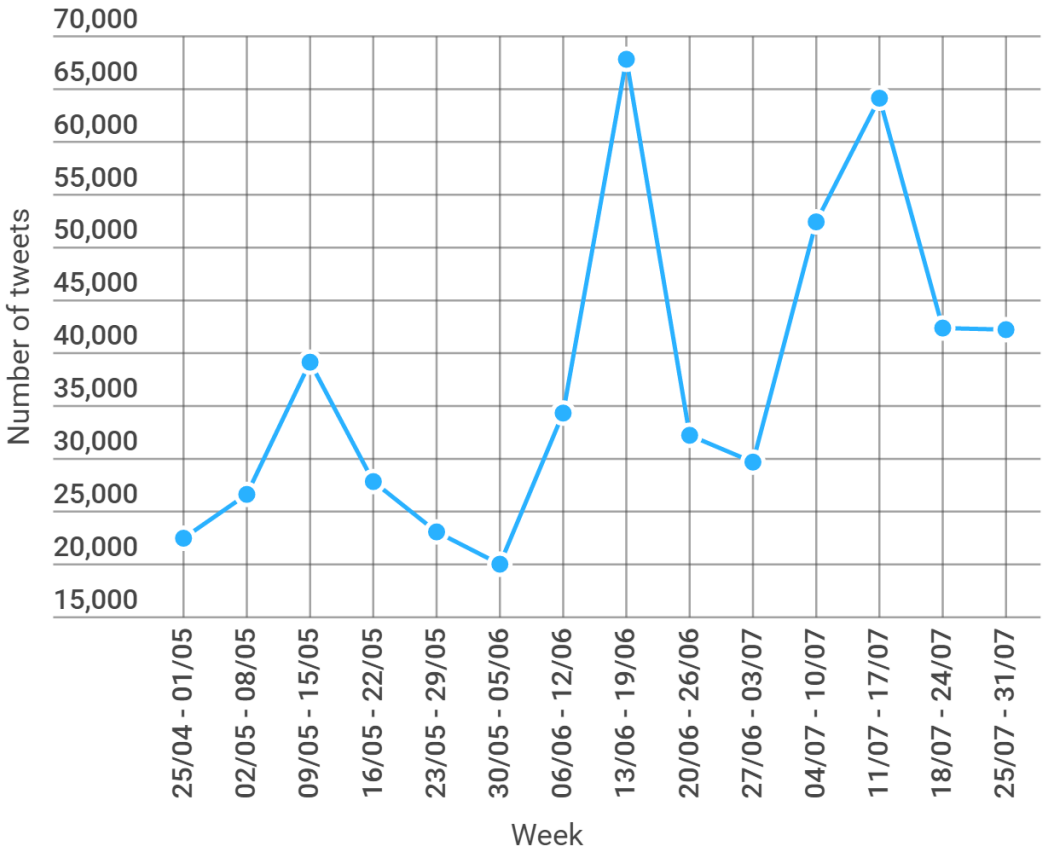


Islamophobic tweets sent each month, from May to July

[2] N.B. During the initial pilot of this system, only 19 days were analysed in March and April, and averages were based on the days measured.

[3] For Demos' analysis of the online reaction to the Brussels attacks, see http://www.demos.co.uk/project/hate-speech-after-brexit/

Over the most recent period, the highest volume of Islamophobic Tweets were sent from 11th to 17th July (64,143), when both the Nice attacks and the attempted military coup in Turkey occurred. This week was one of the biggest spikes in Islamophobia throughout the dataset, second only to 13th to 19th June.



Islamophobic tweets sent each week, from 25th April to 31st July

ISLAMOPHOBIC TWEETS FROM THE UK: MAY TO JULY

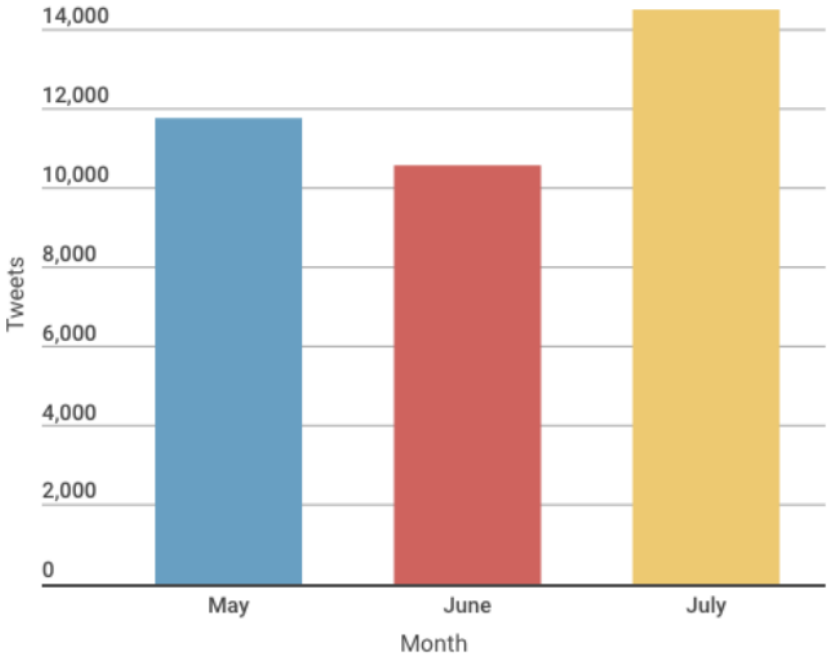
The report also analysed Islamophobic Tweets that were likely sent from the UK. It is important to note that not all Tweets can be geographically placed. A small amount of tweets have definitive information about where they were sent from. These are geo-tags: precise longitude and latitude coordinates that indicate very precisely where the tweet was posted. Only users who proactively turn on the geo-location facility on their smart phone will include this information. A larger number of tweets can be algorithmically located based on geo-location meta-data attached to the Tweet. These include (in addition to the longitudinal- latitudinal data contained above), the 'location field' – where users report where they are from, and time zone. On tests of this method of geolocating Tweets, it has been found to be between 80 and 90 per cent accurate for those Tweets it could locate, and be able to locate between 40 and 70 per cent of Tweets.[4]

From the beginning of March to the end of July 2016, an average of 367 Islamophohbic Tweets from the UK were identified a day. Unlike the global data, the rate of Islamophobic Tweets decreased slightly between May and June. However, consistent with the global picture, the rate then increased sharply between June and July.

Month	Islamophobic tweets sent per day (average)	% increase/decrease on previous month	Total month
May	380	N/A	11,766
June	351	-8%	10,557
July	468	+33%	14,512

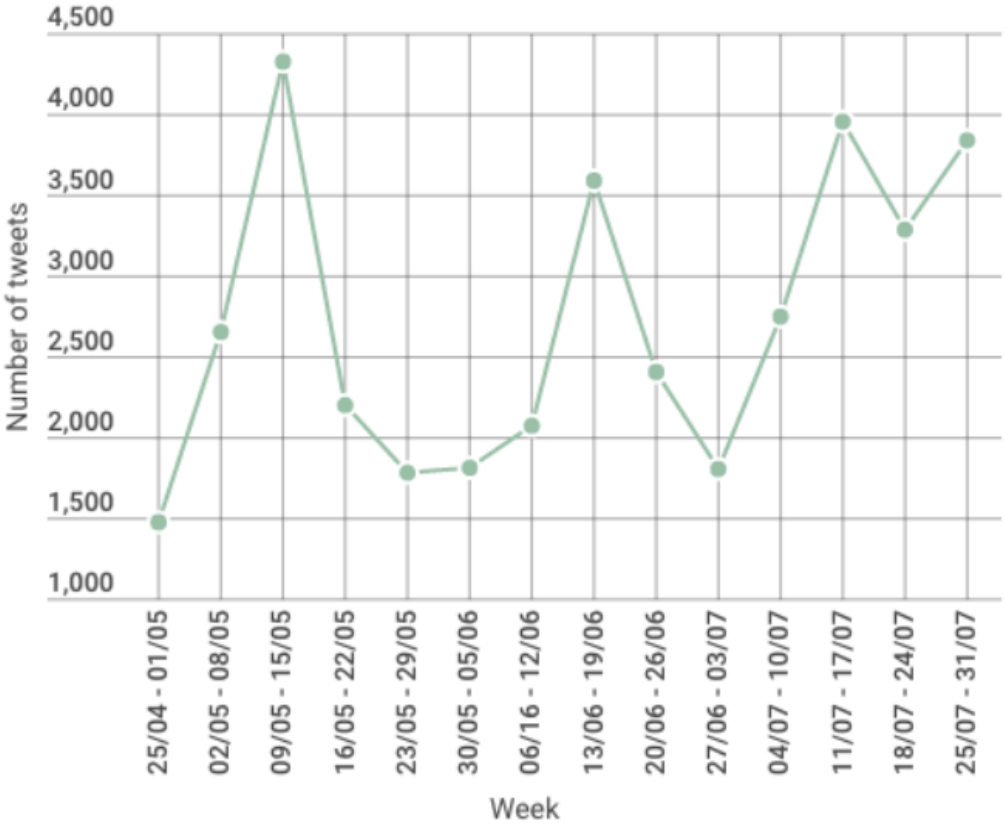
[4] For more information on this, see the Demos paper *The Road to Representivity,* page 30. http://www.demos.co.uk/files/Road_to_representivity_final.pdf?1441811336

July, with an average of 468 anti-Islamic tweets per day, or *14,512* across the month, has the highest rate of anti-Islamic tweets of any of the three months, and above the monthly average of 12,278 for the three months.



Islamophobic tweets sent each month, from May to July

Over the most recent period (July), the highest volume of Islamophobic tweets were sent from 11th to 17th July (3,958) when both the Nice attacks and the attempted coup in Turkey occurred. This week was one of biggest spikes in Islamophobia throughout the dataset, second only to 2nd – 8th May.



Islamophobic tweets sent each week, from 25th April to 31st July

ETHICS

At Demos we believe it is important that the principle of internet freedom should be maintained; and that it should be a place where people feel they can speak their mind openly and freely. However, racist, xenophobic, Islamophobic and misogynistic abuse can curtail freedom, and the capacity to speak and act freely online, as much as it can be an expression of it. It is important, as society confronts the ways that social media acts as a new platform for the expression and dissemination of these of kinds of views, to understand as best as possible the scale, scope, nature and severity of these kinds of practices: when they happen, who they happen to, and why. This is what this research hopes to contribute to.

CASM has conducted extensive work on the ethics and public acceptability of social media research.[5] An ethical framework has been applied to this project, such that:

- The research only uses publicly available data, viewable and visible to any Twitter user;
- The research conducted is aggregated and anonymous: the research does not identify any specific user or users, but to understand the overall scale and nature of Islamophobic abuse on Twitter;
- Where quotations are used as examples and elaborations, they have been altered to maintain the overall meaning, but to prevent the retrospective identification of any Twitter user on the basis of the quotation;
- There is no suggestion of any illegality of any of the content measured: the purpose of the research was not to look for content that was illegal, and it does not suggest that the content that was found was illegal. This research is not seeking to inform how laws should be enforced on social media. This research, and Demos' broader research agenda, seeks instead to inform the broader question of how people from different races, religions, sexualities and genders are spoken about on social media, and the extent that people from different backgrounds face abuse and hostility.

[5] See, for instance, Demos' recent paper with Ipsos MORI #socialethics: A Guide to Embedding Ethics in Social Media Research , https://www.ipsos-mori.com/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf

OVERALL METHODOLOGY

Twitter data is often challenging to analyse. Data drawn from social media are often too large to fully analyse manually, and also often not amenable to the conventional research methods of social science. The research team used a technology platform called Method52, developed by CASM technologists based at the Text Analytics Group at the University of Sussex.[6] It is designed to allow non-technical researchers to analyse very large datasets like Twitter.

Defining Islamophobia

This paper is predicated on the training of a machine to be able to distinguish between an expression that is Islamophobic and one that isn't. An Islamophobic expression was defined as the illegitimate and prejudicial dislike of Muslims because of their faith. However, Islamophobia can take on a very large number of different forms, and its identification, especially within Twitter research, was often challenging. Ultimately, this research comes down to the judgement of the researchers involved. Four main qualitative categories of Islamophobia, on the judgement of the researchers conducting the analysis, were identified:

- 'Islam is the enemy': The idea that it is a fundamental injunction of Islam for all of its followers to be engaged in a violent struggle against non-Muslims and the West;
- The conflation of Muslim populations with sexual violence and a proclivity towards rape;
- Especially in the wake of terrorist attacks, the apportioning of blame for the attack not on the terrorists themselves, or on Islamist militancy, but on the Muslim population generally;
- General abuse, and the general use of anti-Islamic slurs and derogatory descriptions of Muslims.

Data Collection
Method52 was used to directly collect Tweets from Twitter's Stream and Search 'Application Programming Interfaces' (or APIs). They allow all Tweets to be collected that contain one of a number of specified keywords. The keywords used in the various collections used in this research are detailed in the annex.

Data Analysis
Method52 allows researchers to train algorithms to split apart ('to classify') Tweets into categories, according to the meaning of the Tweet, and on the basis of the text they contain. To do this, it uses a technology called natural language processing. Natural language processing is a branch of artificial intelligence research, and combines approaches developed in the fields of computer science, applied mathematics, and linguistics. An analyst 'marks up' which category he or she considers a tweet to fall into, and this 'teaches' the algorithm to spot patterns in the language use associated with each category chosen. The algorithm looks for statistical correlations between the language used and the categories assigned to determine the extent to which words and bigrams are indicative of the pre-defined categories. Details about how these algorithms were used, and how well they worked, are provided below.[7]

The Accuracy of Algorithms
To measure the accuracy of algorithms into the categories chosen by the analyst, we used a 'gold standard' approach. For each, around 100 Tweets were randomly selected from the relevant dataset to form a gold standard test set for each classifier. These were manually coded into the categories defined above. These Tweets were then removed from the main dataset and so were not used to train the classifier.

As the analyst trained the classifier, the software reported back on how accurate the classifier was at categorising the gold standard, as compared to the analyst's decisions. On the basis of this comparison, classifier performance statistics – 'recall', 'precision', and 'F-score' are created and appraised by a

[6] This group is led by Professor David Weir and Dr Jeremy Reffin. More information is available about their work at: http://users.sussex.ac.uk/~davidw/styled-3/

[7] For a more detailed description of this methodology, see the Demos paper *Vox Digitas*,

human analyst. Each measures the ability of the classifier to make the same decisions as a human in a different way:

Overall accuracy: This represents the percentage likelihood of any randomly selected Tweet within the dataset being placed into the appropriate category by the algorithm. It is based on three other measures (below).

Recall: The number of correct selections that the classifier makes as a proportion of the total correct selections it could have made. If there were 10 relevant Tweets in a dataset, and a relevancy classifier successfully picks 8 of them, it has a recall score of 80 per cent.

Precision: This is the number of correct selections the classifiers makes as a proportion of all the selections it has made. If a relevancy classifier selects 10 Tweets as relevant, and 8 of them actually are indeed relevant, it has a precision score of 80 per cent.

F-Score: All classifiers are a trade-off between recall and precision. Classifiers with a high recall score tend to be less precise, and vice versa. The 'overall' score reconciles precision and recall to create one, overall measurement of performance for each decision branch of the classifier.

N.B. the values for each algorithm (called a classifier) are presented within the detailed methodology of this report. The values are expressed as value up to 1: a value of 0.76, for instance, indicates a 76% accuracy.

CAVEATS

The research of large social media datasets is a reasonably new undertaking. It is important to set out a series of caveats related to the research methodology that the results must be understood in the light of:

- **The algorithms used are not perfect:** throughout the report, some of the data will be mis-classified. The technology used to analyse Tweets is inherently probabilistic, and none of the algorithms trained and used to produce the findings for this paper were 100% accurate. The accuracy of all algorithms used in the report are clearly set out in this report.

- **Some data will be missed:** Acquiring Tweets on the basis of the keywords that they contain presents two possible problems. First, the initial dataset may contain Tweets that are irrelevant to the thing being studied. Secondly, it may miss Tweets that are relevant to the thing being studied. Researchers worked to construct as comprehensive a list of keywords as possible (these are detailed in the report, below), however it is likely some were missed, and the numbers presented in this report are likely a subset of the total.

- **Twitter is not a representative window into British society**: Twitter is not evenly used by all parts of British society. It tends to be used by groups that are younger, more socio-economically privileged and more urban. Additionally, the poorest, most marginalised and most vulnerable groups of society are least represented on Twitter; an issue especially important when studying the prevalence of xenophobia, Islamophobia and the reporting of hate incidents.[8]

- Overall, this research is intended to be an **indicative**, first-take of the reaction on Twitter to these important events. It is not presented as either exhaustive or definitive; and it is very much hoped that it will stimulate further research on this vital topic in the future.

[8] For a longer discussion of this issue, see the Demos paper *The Road to Representivity*

DETAILED METHODOLOGY

Identifying Tweets that were hateful, derogatory and anti-Islamic was a formidable analytical challenge. First, all Tweets were collected that contained one of an extensive list of terms that could be used in an anti-Islamic way (see annex). This collection began on February 29th and continued until the 2nd August. It returned a very large number of Tweets over this period, over 34,000,000. The very large majority of these Tweets were not anti-Islamic or hateful. A series of algorithms were built to respond to the different challenges that this dataset posed in order to identify the anti-Islamic subset within the larger body of data. Each was designed to remove Tweets which were not Islamophobic from the dataset:

- A large number of Tweets contained the word 'Paki'.[9] A classifier was used to separate derogatory uses of this word from non-derogatory uses.
- A large number of Tweets also contained the word 'terrorist'. Of course, many Tweets containing this word were in no way derogatory or anti-Islamic. Two classifiers were built to analyse tweets containing these words:
  - First, a classifier was trained to separate Tweets referring to Islamist terrorism from other forms of terrorism.
  - Second, of the Tweets referring to Islamist terrorism, a classifier to distinguish views broadly attacking Muslim communities in the context of terrorism, from those broadly defending Muslim communities.
- A classifier was trained to separate all other Tweets in the dataset into those that were derogatory and anti-Islamic from those which were not.
- Last, the Tweets that, based on the above, (a) used the term 'Paki' in a derogatory way, (b) that used the term 'terrorist' to broadly attack Muslims or Muslim communities, (c) that used the other possible slur terms in the collection in a way that was anti-Islamic were combined. These were then filtered to include only Tweets sent from the UK. This resulted in the final total of Islamophobic Tweets.

The accuracy of these algorithms are as follows:[10]

[9] N.B. whilst this word refers to an ethnic rather than religious group, it was found that it was often used interchangeably to refer to Muslim communities

[10] Due to the large number of classifiers used, the accuracy was checked by taking a random sample of Tweets that - according to the system of algorithms described above - were classified as derogatory anti-Islamic. 75 of these 100 were identified by an analyst as derogatory and anti-Islamic.

Category	Precision	Recall	F-Score	Accuracy
Hate	0.680	0.745	0.711	0.715
Other	0.753	0.689	0.719	0.715

Tab. 9. Reliability scores for classifier identfiying Tweets using 'Paki' in a derogatory way

Category	Precision	Recall	F-Score	Accuracy
Islamist Terrorism	0.870	0.762	0.812	0.858
Non-Islamist Terrorism	0.852	0.923	0.886	0.858

Tab. 10. Reliability scores for classifier separating references to Islamist and non-Islamist terrorism

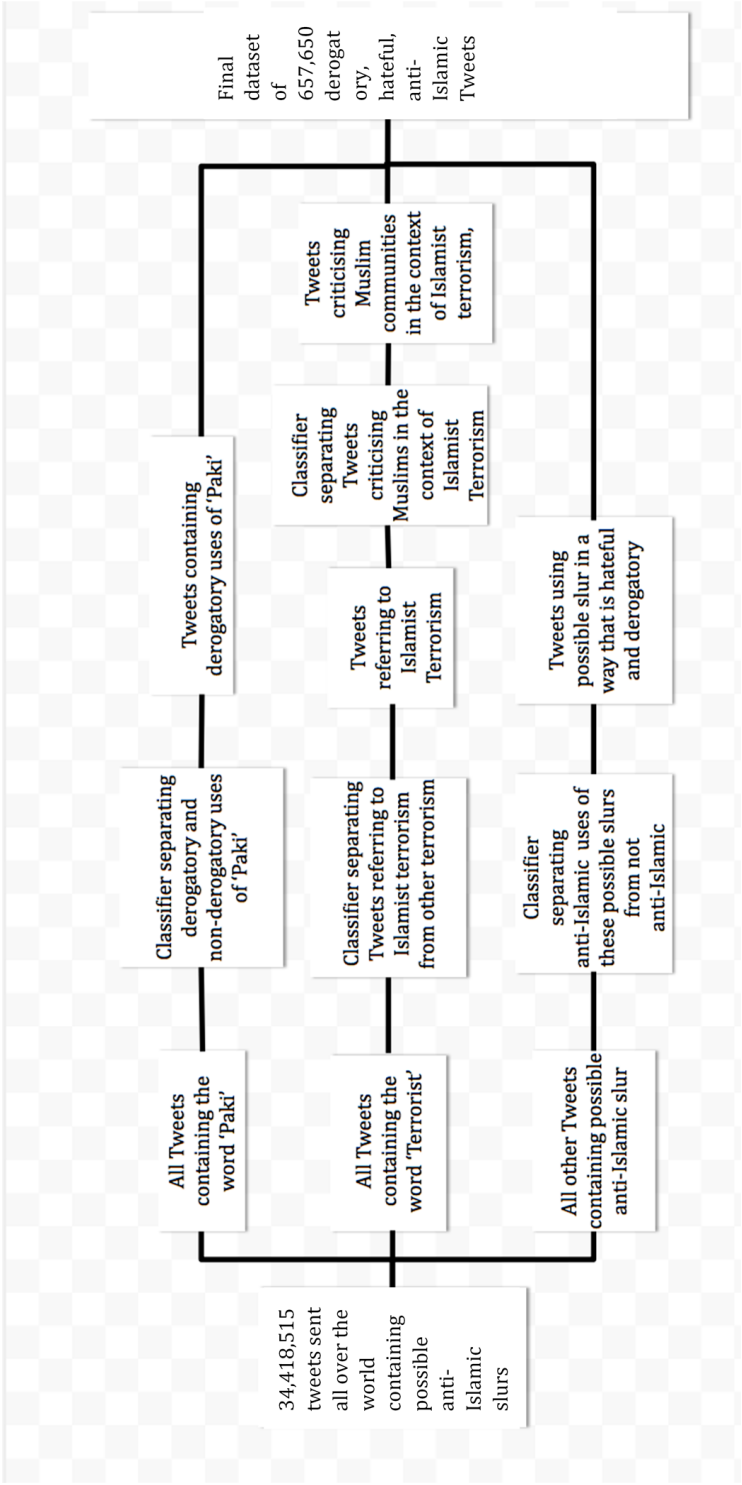
Category	Precision	Recall	F-Score	Accuracy
Attacking	0.645	0.727	0.648	0.693
Defending	0.718	0.830	0.770	0.693
Other	0.800	0.291	0.427	0.693

Tab. 11. Reliability scores for classifier separating Tweets attacking or defending Muslims in the context of a reference to terrorism

Category	Precision	Recall	F-Score	Accuracy
Anti-Islamic	0.533	0.702	0.606	0.740
Not anti-Islamic	0.864	0.755	0.806	0.740

Tab. 12. Reliability scores for classifier separating Tweets containing other possible slurs into those that were hateful and anti-Islamic from those that were not

These algorithms were connected together into an 'architecture', shown below. Each Tweet collected passed through the architecture on the basis of how it was classified. Overall, this system of algorithms succeeded in filtering the very large (over 34,000,000) number of Tweets into a much smaller (657,650) subset that were much more likely to be hateful, derogatory and anti-Islamic.



Annex - Data Collection Keywords

The annex contains the keywords used to collect Tweets analysed throughout this report.

1. Words/Hashtags used to collect Tweets that could be derogatory and anti-Islamic

- Jihad
- Jihadi
- Sand Flea
- Terrorist
- hijab
- Camel Fucker
- Carpet Pilot
- Clitless
- Derka Derka
- Diaper-Head
- Diaper Head
- Dune Coon
- Dune Nigger
- Durka-durka
- Jig-Abdul
- Muzzie
- Q-Tip Head
- Rab
- Racoon
- Rag-head
- Rug Pilot
- Rug-Rider
- Sand Monkey
- Sand Moolie
- Sand Nigger
- Sand Rat
- Slurpee Nigger
- Towel-head
- Muslim Paedos
- Muslim pigs
- Muslim scum
- Muslim terrorists
- Muzrats
- muzzies
- Paki
- Pakis
- Pisslam
- raghead
- ragheads
- Towel head
- FuckMuslims
- WhiteGenocide
- Pegida
- EDL
- BNP
- Rapefugee
- Rapeugee
- mudshark
- kuffar
- kaffir