

HATE M E T E R

Hate speech tool for monitoring, analysing
and tackling Anti-Muslim hatred online
REC Action Grant (REC-DISC-AG-2016-04)
24 months (01.02.2018 - 31.01.2020)



This document was funded by the European Union's
Rights, Equality and Citizenship Programme (2014-2020)

D18

November 2019

Training module A for academics and research organisations



D18 – Training Module A for academics and research organisations

WP5	Training, dissemination and sustainability events
Due Date:	30/11/2019
Submission Date:	30/11/2019
Responsible Partner:	UT1-Capitole
Version:	1.0
Status:	Final
Author(s):	Jérôme Ferret and Mario Laurent (UT1-Capitole), Andrea Di Nicola, Daniela Andreatta, Gabriele Baratto and Elisa Martini (UNITRENTO), Marco Guerini and Sara Tonelli (FBK), Georgios A. Antonopoulos and Parisa Diba (TEES)
Reviewer(s):	Jérôme Ferret and Mario Laurent (UT1-Capitole), Andrea Di Nicola, Daniela Andreatta, Gabriele Baratto and Elisa Martini (UNITRENTO), Serena Bressan, Marco Guerini and Sara Tonelli (FBK), Georgios Antonopoulos and Parisa Diba (TEES), Isis Koral (CCIF), Francesca Cesarotti (Amnesty Italy), Bill Howe (Stop Hate UK)
Deliverable Type:	R
Dissemination Level:	CO

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

The content of this document represents the views of the authors only and is their sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

Glossary

CAP	Computer Assisted Persuasion
CCIF	Collectif contre l'islamophobie en France
LEA	Law Enforcement Agency (plural: LEAs)
MS	Member State (plural: MSs)
NGO	Non-governmental organisation (plural: NGOs)
NLP	Natural Language Processing
SD	Standard Deviation
WP	Work Package (plural: WPs)

Table of contents

EXECUTIVE SUMMARY	4
1. INTRODUCTION	6
1.1 What is Islamophobia?.....	6
1.2 Project Hatemeter	7
2. HATEMETER PLATFORM	8
2.1 Data Analysis	8
2.1.1 Recent trends Functionality	9
2.1.2 Hashtag trends Functionality	10
2.1.3 Hate speakers Functionality	12
2.2 Computer-Assisted Persuasion (CAP).....	12
2.2.1 Alerts Functionality	12
2.2.2 Counter-narratives Functionality.....	13
3. HATE SPEECH AGAINST MUSLIMS ON SOCIAL MEDIA: EVIDENCE AND ANALYSIS.....	16
3.1 Italy	16
3.1.1 Islamophobia in Italy	16
3.1.2 Evidence of Islamophobia in Italy through the Hatemeter Platform.....	17
3.2 France	27
3.1.1 Islamophobia in France.....	27
3.1.2 Evidence of Islamophobia in France through the Hatemeter Platform	29
3.3 UK.....	36
3.3.1 Islamophobia in the United Kingdom	36
3.3.2 Evidence of Islamophobia in the UK through the Hatemeter Platform	36
4. SUGGESTIONS AND INSIGHTS ON THE USE OF THE PLATFORM AND OF THE HATEMETER METHODOLOGY FOR ACADEMICS AND RESEARCHERS	51
REFERENCES	53

Executive summary

This document is the **Deliverable “D18 – Training Module A”** of the European project **“Hatometer - Hate speech tool for monitoring, analysing and tackling anti-Muslim hatred online”** (hereinafter also referred to as “Hatometer”, project reference: 764583), which aims at systematising, augmenting and sharing knowledge on anti-Muslim hatred online, and at increasing the efficiency and effectiveness of non-governmental organisations (NGOs) in preventing and tackling Islamophobia at the European Union (EU) level.

This deliverable has been designed to be used by academics and research organizations outside the Hatometer project as a **manual** to understand the **Hatometer Platform’s** (also referred to as the Hatometer Tool) main goals, how it works and what it can achieve. Together with Training Module B for stakeholders, the present document is one of the main deliverables of the **“Activity 5.1 - Networking, capacity building and training activities for target stakeholder groups”** within the Hatometer **WP5 “Training, dissemination and sustainability events”**.

Hatometer addresses a **strategic challenge towards NGOs’ needs to tackle anti-Muslim hate speech online**, by developing and testing an **ICT tool (i.e., Hatometer Platform)** which **automatically monitors and analyses Internet and social media data on the phenomenon and produces computer-assisted responses and hints to support counter-narratives and awareness raising campaigns**. In this regard, the Hatometer Platform uses a combination of natural language processing (NLP), machine learning, and big data analytics/visualisation to: a) identify and systematise in real-time actual “red flags” of anti-Muslim hate speech and/or possible related threats online (**Real-time identification**); b) understand and assess the sets of features and patterns associated with trends of Islamophobia online (**In-depth understanding**); c) develop an effective tactical/strategic planning against anti-Muslim hatred online through the adoption of an innovative Computer Assisted Persuasion (CAP) approach (**Tactical/strategic response**); d) produce an accurate counter-narrative framework for preventing and tackling Islamophobia online, and building knowledge-based and tailored awareness raising campaigns (**Counter-Narratives Production**).

As such, this **Training Module A** outlines **practical uses** and also **provides examples to find and analyse hate speech against Muslims on Twitter** from the three social sciences and humanity research teams involved in the Hatometer project, namely University of Trento in Italy, University Toulouse 1 Capitole in France and Teesside University in the United Kingdom, using the Hatometer Platform developed by FBK and tested by the three NGOs involved in the Partnership, namely Amnesty International – Sezione Italiana (AMN - IT), StophHate UK (STOPHATE - UK), and Collectif Contre l’Islamophobie en France (CCIF - FR). In these countries, the magnitude of the phenomenon is significant, but no systematic responses have been implemented. This document is tailored to the scientific and technical requirements and needs of specific key actors and designed with a view to durably integrate them into their daily working practices. The Training Module A will be employed during the **training seminar for academics and researchers in Toulouse, France on the 17th December 2019** and made available through the **Hatometer website** (www.hatometer.eu).

This document is organised as follows. The first section **“Introduction”** starts by giving an idea of the problem addressed within the project Hatometer and, consequently, it briefly describes **what Islamophobia is** (subsection 1.1). The second subsection (1.2) reports a **general presentation** of the project and its aims. The second section of the document is a technical description of the Hatometer Platform and of its various features. It presents the **data analysis** that can be performed (subsection 2.1, and more specifically: 2.1.1 Recent trends Functionality, 2.1.2 Hashtag trends Functionality, and 2.1.3 Hate speakers Functionality) and the **Computer-Assisted Persuasion tool** (subsection 2.2, and more specifically: 2.2.1 Alerts Functionality, and 2.2.2 Counter-narratives Functionality).

The third section summarises **evidence of online Islamophobia** in the three countries involved in the project: namely, Italy (subsection 3.1), France (subsection 3.2) and the United Kingdom (subsection 3.3). Each subsection starts with a brief description of the **context and background** of online Islamophobia and anti-Muslim hatred in the country and then specifically presents **evidence** of hate speech collected through the Hatemeter Platform, as well as the **analysis** of the information gathered thanks to the visualisation tools available in the Platform. The fourth and last section proposes some **suggestions and insights** on the use of the Platform and of the Hatemeter methodology for academics and researchers in further works.

1. Introduction

1.1 What is Islamophobia?

Islamophobia is defined as “all acts of discrimination or violence against institutions or individuals because of their real or supposed affiliation to Islam” (CCIF, 2019), evinced as **feelings of anxiety** or **perceptions of fear** and **hatred**. Additionally, Islamophobia does not merely entail anxious awareness or perceptions rooted in apprehension and contempt, but also discriminatory attitudes and hostile practices through which it is manifested and expressed, such as harassment, verbal and physical abuse as well as hate crimes, perpetrated in both **offline** and **online contexts**.

The European Islamophobia Report (Bayrakli and Hafez, 2019a) recorded that Muslims are among the **primary victims** of the rise of far-right extremism in Europe. Below some reported examples retrieved from the Report. In Austria, the Office for Documenting Islamophobia and Anti-Muslim Racism recorded a 74% increase of anti-Muslim racist acts in its 2018 report. In France, the Collectif contre l'islamophobie en France (CCIF) recorded a 52% increase.¹ In the UK, the number of cases recorded in official statistics rose by 17% in 2017-18 and religion-specific incidents multiplied by 40%. In Italy, there has been an increase of hate tweets in social media in 2018, in respect to 2017. Finally, in the Netherlands, the Anti-discrimination Agencies announced that in 91% of religious discrimination cases reported to the authorities, the target was Muslims.

In the last decade, Islamophobia has gained momentum through the **Internet**, which has enabled the spread of polemics, anti-Islamic and anti-Muslim discourses to a worldwide audience (Larsson, 2007), along with new media technologies including social media platforms and global digital networks (Horsti, 2017). With the advent of the Internet, online or cyber Islamophobia has seen a **large increase**, with spaces on the Internet now becoming a platform for the spreading of its rhetoric, in which xenophobic viewpoints and racist attitudes towards Muslims being easily disseminated into public debate (Ekman, 2015). Online Islamophobia takes place primarily through **blogs** and **social media**, as well as through traditional media outlets seen online (Aguilera-Carnerero and Azeez, 2016).

In 2018, The Collective Against Islamophobia in Belgium underlined that **29% of reported Islamophobic incidents in 2018 pertain to Islamophobia in the media and online** (Bayrakli and Hafez, 2019b). However, as Faytre (2019: 18) points out, “Islamophobic controversies often started from social media before getting debated in mainstream media and triggering reactions among politicians.”

According to Oboler (2016), anti-Muslim hate, much like many other forms of hate, is unlikely to remain purely virtual, with online Islamophobia likely to incite religious hatred and xenophobia leading to **real world crimes** and a **rise in political extremism** from both the far-right and the radicalisation of Muslim youth in response to such messages of exclusion. The outcome is a vicious circle that is particularly difficult to break. Thus, as Larsson (2007) points out, it is important to question to what extent the Internet is being used to spread and foster anti-Muslim and anti-Islamic opinions in contemporary society.

¹ It is necessary to underline that, according to the French government (Gouvernement.fr, 2019), the official number of anti-Muslim acts is the lowest since 2010. This statement recalled the CNCDH report (2018) on racism, antisemitism and xenophobia, which affirms that the number of anti-Muslim hatred acts is lower than the previous years because a lot of acts are not reported to the police. On the contrary, the CCIF's number reported in the text indicates an increase. The difference can be due to two possible explanations: i) CNCDH and CCIF can rely on different definition of Islamophobic acts and consequently register differently the information; ii) the phenomenon can suffer of under-reporting problems in regard to police's number, whereas people can feel more comfortable to report these types of acts to CCIF.

1.2 Project Hatemeter

Project “Hatemeter - Hate speech tool for monitoring, analysing and tackling anti-Muslim hatred online” aims at **systematising, augmenting and sharing knowledge on anti-Muslim hatred online**, and at increasing the efficiency and effectiveness of NGOs in **preventing and tackling Islamophobia** at the EU level, by developing and testing an **ICT tool (i.e., Hatemeter Platform)** that **automatically monitors and analyses Internet and social media data** on the phenomenon, and **produces computer-assisted responses and hints** to support **counter-narratives** and **awareness raising campaigns**.

More specifically, backed by a strong interdisciplinary effort (criminology, social sciences, computer sciences, statistics, law), the Hatemeter Platform uses a combination of **natural language processing (NLP), machine learning, and big data analytics/visualization** to:

- A. identify and systematise in real-time actual “red flags” of anti-Muslim hate speech and/or possible related threats online (**Real-time Identification**);
- B. understand and assess the sets of features and patterns associated with trends of Islamophobia online (**In-depth Understanding**);
- C. develop an effective tactical/strategic planning against anti-Muslim hatred online through the adoption of the innovative Computer Assisted Persuasion (CAP) approach (**Tactical/Strategic Response**);
- D. produce an accurate counter-narrative framework for preventing and tackling Islamophobia online and building knowledge-based and tailored awareness raising campaigns (**Counter-Narratives Production**).

The Hatemeter Platform has been **piloted and tested in three NGOs of EU Member States (MSs)** where the magnitude of the problem is considerable but no systematic responses have been implemented (**France, Italy and the United Kingdom**), thus enabling Project Hatemeter to address several objectives of the **Annual Colloquium on Fundamental Rights "Tolerance and respect: preventing and combating anti-Semitic and anti-Muslim hatred in Europe"** and the **European Agenda on Security (2015)**, as well as the **priorities** of the REC call of proposals.

In order to strengthen **cooperation** between key actors and to ensure the **widest circulation** and **long term impact** of project results on future research streams and operational strategies, the project favours **capacity building and training** and the **sustainability and transferability** of the Hatemeter Platform among **other target stakeholder groups** (e.g., LEAs, journalists/media, etc.) across the EU and for **other forms of hate speech**, through the building of the “**EU laboratory on Internet and social media for countering online anti-Muslim hate speech**” (i.e., **Hatemeter Lab**).

The present document is one of the main deliverables of the “**Activity 5.1 - Networking, capacity building and training activities for target stakeholder groups**” within the Hatemeter **WP5 "Training, dissemination and sustainability events"**, and precisely, the “**D18 - Training Module A**”. This **Training Module** is meant to be a **manual for academics and research organizations** outside the Hatemeter project to understand the main goals of the Hatemeter Platform (or Hatemeter Tool), how it works and what it can achieve.

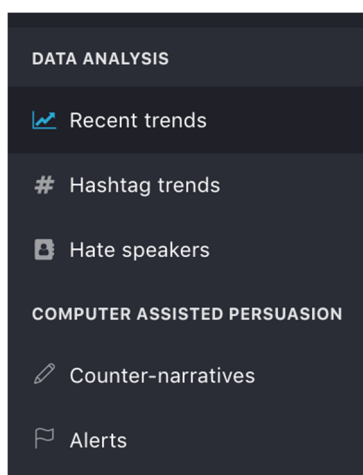
2. Hatemeter Platform

The HATEMETER Platform is a **web-based tool designed to support researchers and stakeholders** (e.g. NGO operators, researchers, law enforcement agencies) in analysing and countering anti-Muslim speech online. **Twitter** and **YouTube** are continuously monitored and, when keywords or hashtags related to anti-Muslim discourse are detected, the corresponding messages are retrieved and stored in the project database. Here we **focus on the analytics applied** to tweets to better understand Islamophobic messages, which are displayed in the Platform following different perspectives. Indeed, since the analyses related to YouTube have not been tested during the pilot phases, they have not been included in the training kit, although they are visible in the final release of the Platform (v.3). From a technical point of view, the Platform relies on a relational database and a tomcat application server. The interface is based on existing javascript libraries such as C3.js (<https://c3js.org>), D3.js (<https://d3js.org>) and Sigma.js (<http://sigmajs.org>).

The Platform is password-protected, and each account gives access to data in a **specific language** (Italian, English or French) and a version of the underlying database. The analytic tools, instead, remain the same across the three versions, since they have been designed so to be language-independent.

The **Platform functionalities** are divided into three main groups, displayed on the left of the homepage view: 1) DATA ANALYSIS, 2) COMPUTER ASSISTED PERSUASION, and 3) PROJECT HATEMETER (see Figure 1). The first item includes all analyses involving Islamophobic hashtags and keywords and the network of users spreading. The second item displays the Platform which, taking a hate message in input, automatically generates possible counter-narratives. The third item includes general information on the project and the link to the website. The first two groups are described in more detail below (section 2.1, 2.2).

Figure 1 - Platform menu with main functionalities



SOURCE: Screenshot from the Hatemeter Platform

2.1 Data Analysis

The **“DATA ANALYSIS”** item presents three views: *“Recent trends”*, *“Hashtag trends”* and *“Hate speakers”*. There is a main difference between the first item and the others: *“Recent trends”* performs **real-time monitoring** and therefore calls Twitter APIs on the fly. The others, instead, rely on the **HATEMETER database**, i.e. the outcome of Islamophobic discourse monitoring which has been in operation since October 2018. This means that, while the analyses on past data are stable, those in real-

time may be subject to changes whenever Twitter’s Application Programming Interface (API) policy undergoes some revision.

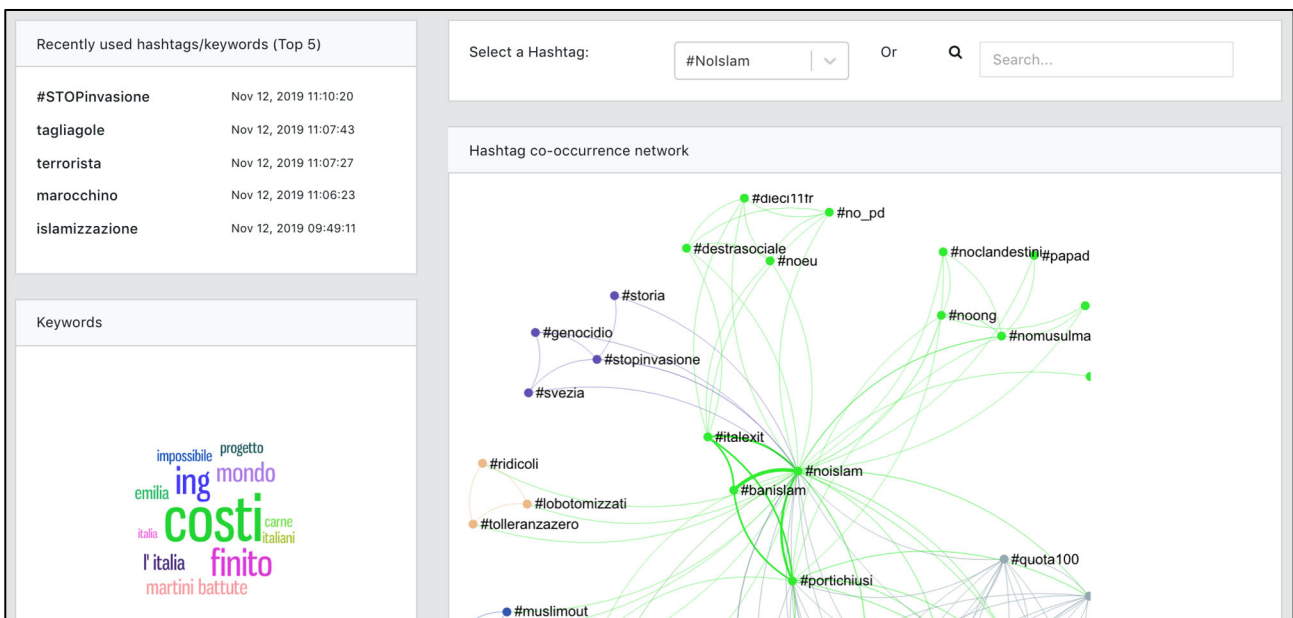
For each language, a pool of academics and activists defined in the early stages of the project, a set of **hashtags and keywords** that are directly associated with Islamophobic messages (e.g. #STOPIslam, #Muslimshit², #BanIslam). These hashtags and keywords were used as query terms to access Twitter APIs on a regular basis (twice a week) and collect all messages containing the query term. The collected tweets were analysed using **text processing tools** to extract the most relevant information related to anti-Muslim hatred online. These could be the **metadata** connected to the messages (i.e., user, date, frequency), the **content popularity** (number of replies, i.e. answers to a message or tweet, and retweets, i.e. broadcasting a tweet or message posted by another person), and the **network** in which the discourse is spread (i.e. nodes that had most interactions involving the hashtags or keywords of interest).

The information distilled and structured in the previous steps are made available to final users through an advanced visualisation Platform. This provides functionalities for the **visual exploration and analysis of the data**, enabling content monitoring, synchronic and diachronic comparisons, close and distant reading, data clustering, network analysis, etc. Pictorial and graphical format are used as much as possible so as to make the tool language and country-independent.

2.1.1 The “Recent trends” Functionality

Under DATA ANALYSIS, the “Recent trends” view allows users to **monitor current Twitter activities around pre-defined Islamophobic keywords and hashtags** or looking for terms defined on-the-fly by the user. After selecting a hashtag or writing a term in the field, each box displays different related information: in the “Recently used hashtags/keywords”, the system ranks the most recently used Islamophobic terms with the corresponding date and time when they were last posted on Twitter.

Figure 2 - View of the “Recent trends” tab

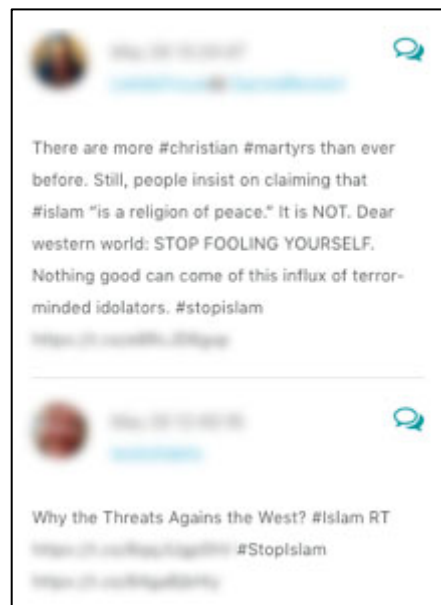


SOURCE: Screenshot from the Hatometer Platform

² Original language: Musulmerda

For example, Figure 2 shows what would be displayed for the Italian pilot on 12 November 2019: the five Islamophobic hashtags and terms that have been most recently tweeted; the network of hashtags being currently used together with #NoIslam (selected by the user); and the tag cloud extracted from the recent tweets containing #NoIslam and extracted with Keyphrase Digger, a keyword extractor tool (Moretti et al., 2015). In the hashtag network, **each node is a hashtag co-occurring with the central one**, and the arcs connecting nodes represent co-occurrence. **The thicker the arc, the more frequently two hashtags have been used together**. Finally, on the right it is possible to see the list of tweets, ranked by date, containing the hashtag or keyword selected by the user (Figure 3). This information is displayed to provide a **more fine-grained view of the Islamophobic content** currently circulating online. By clicking on one of these tweets, it is possible to open the message of interest inside Twitter, so to check for example the replies or the number of retweets. Each of these tweets displayed inside the feed contains also the link to the user who originally posted it, as well as that of the user who re-tweeted it. These messages are also **linked to the counter-narratives tool**, which can be activated by clicking on the speech bubble icon on the right. More details on this tool are reported in Section 2.2.

Figure 3 – Feed of tweets containing a user-defined hashtag or keyword, ranked by date



SOURCE: Screenshot from the Hatemeter Platform – Tweet feed

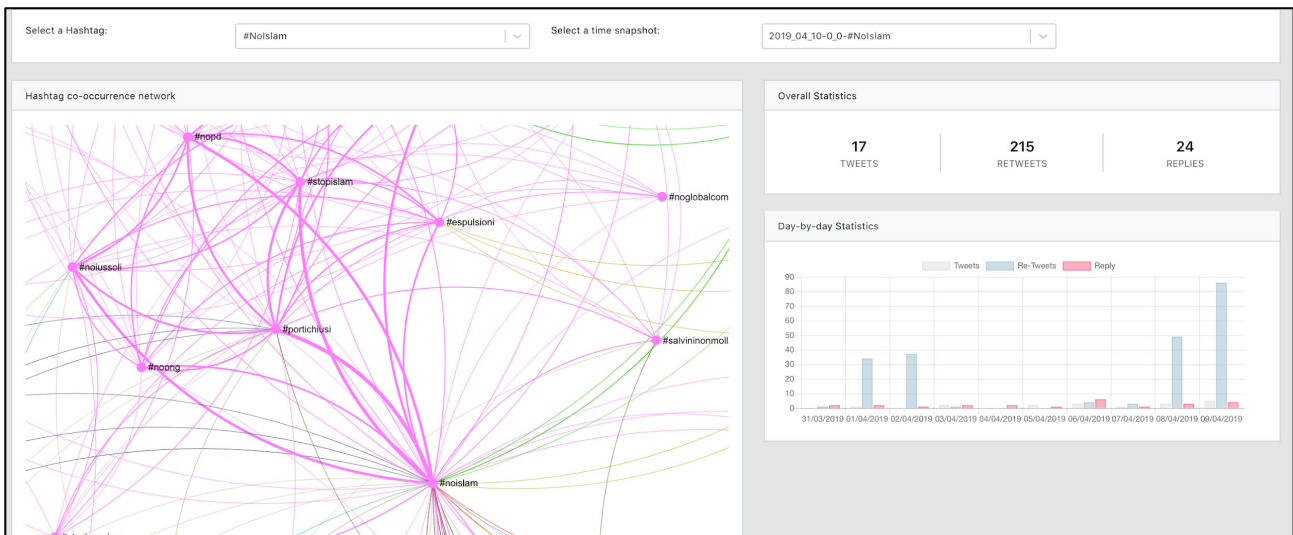
Overall, the information shown in this tab can be **utilised by researchers and NGO operators** to *i)* discover new hashtags to be monitored, and *ii)* acquire an overview of the main topics discussed in Islamophobic tweets.

2.1.2 The “Hashtag trends” Functionality

By clicking on the “Hashtag trends” item under the “DATA ANALYSIS” menu, users can see a **more fine-grained analysis of hashtags**, focused in particular their **trend over time**. In this case, the information is not retrieved on the fly using Twitter APIs, but it is gathered from the underlying database. In particular, users can select a hashtag, and a temporal snapshot of interest, and then the system returns **general statistics** reporting the number of tweets, retweets and replies containing the hashtag of interest in the selected snapshot. In addition, the system displays the **hashtag co-occurrence network** and the **day-by-day statistics** showing the hashtag presence on Twitter over time and enabling the visualisation of the message popularity day-by-day as well as to compare different snapshots.

Figure 4 reports an example: users can select a hashtag (in this case #NoIslam) and a temporal snapshot of interest (10 days ending on April 10th). In the upper right of the figure, it is possible to read the general statistics on these 10 days, reporting the number of tweets, retweets and replies containing the selected hashtag. On the left of the figure, the system shows the hashtag co-occurrence network (similar to the one displayed in Figure 2), whereas in the lower right there are the day-by-day statistics showing the hashtag presence on Twitter over time.

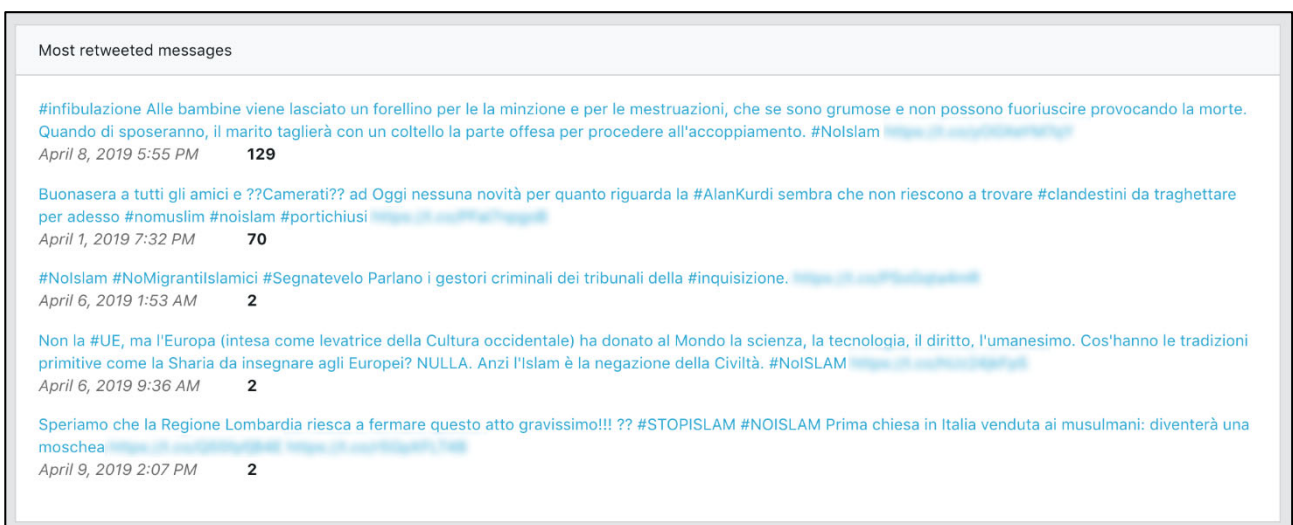
Figure 4 – Feed of tweets containing a user-defined hashtag or keyword, ranked by date



SOURCE: Screenshot from the Hatemeter Platform

Furthermore, a box at the bottom of the tab can show the **list of most retweeted messages among those collected through the query of interest** (see Figure 5). In particular, each tweet is displayed together with the date when it was posted and the number of retweets (ranked in decreasing order from the most retweeted ones). Note that this box is not always present, since for some hashtags and some temporal snapshots no retweets with a frequency > 1 are present.

Figure 5 – View of the most retweeted messages for a given hashtag/keyword in a given time period. The number next to the date shows how many times the message was retweeted.



SOURCE: Screenshot from the Hatemeter Platform

2.1.3 The “Hate speakers” Functionality

By clicking on the “**Hate Speakers**” item under the “DATA ANALYSIS” menu, users of the Platform can display the **analyses related to the** most active **users** in the community spreading Islamophobic messages online. As in the previous tab, data are extracted from the Hatometer database and not collected on the fly.

After selecting a hashtag and a time snapshot, the Platform displays the network of users that posted messages containing the selected hashtag. This is called the “**User co-occurrence network**”, where colours are automatically assigned by the network analysis algorithm to identify communities of users, i.e. those who interact more often with each other through replies or retweets. In the “**Most connected users**” frame, the Platform displays a ranked list of users with the most connections inside the network, i.e. those that are more likely to give visibility to Islamophobic messages. Finally, a tag cloud representing the messages exchanged inside the identified community of users is also created, again using the KD keyword extraction tool.

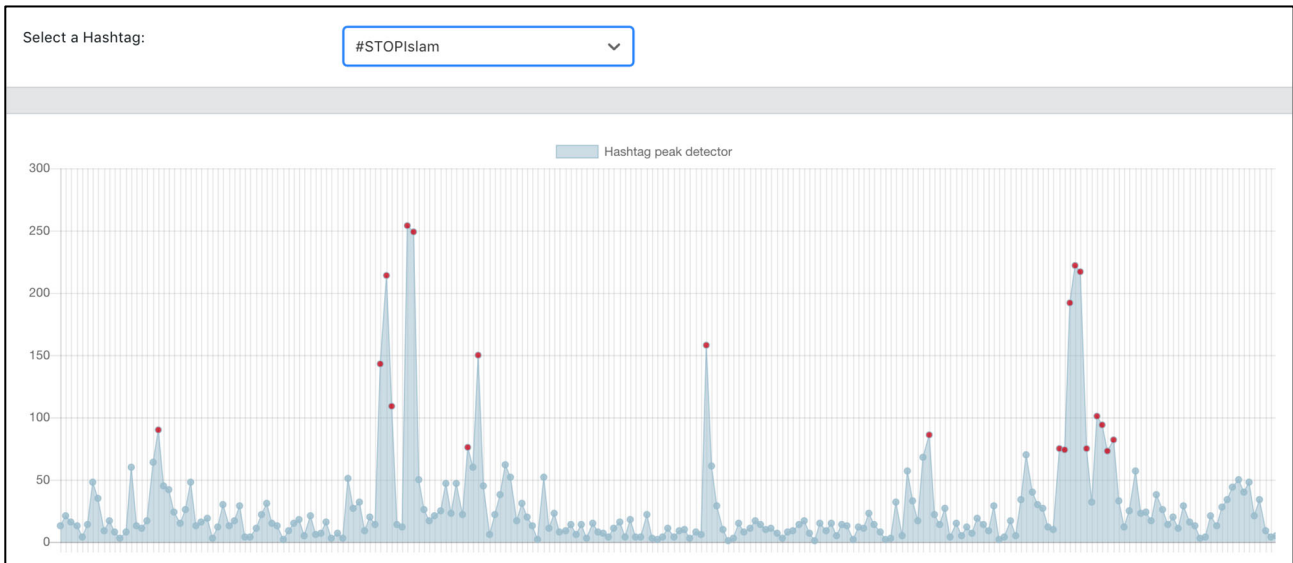
2.2 Computer-Assisted Persuasion (CAP)

The goal of the “Computer-Assisted Persuasion” CAP item of the Hatometer Platform is to provide a **suite of tools for stakeholders to support actual intervention online**, i.e. answering virtual hatred to prevent and combat it. To this end, CAP approach is twofold: i) providing timely, context-sensitive and incident/crime-specific **alerts** based on complex data analytics; ii) producing an accurate **counter-narrative framework** on the basis of a “best experiences” repository containing real positive examples of counter-narratives to Islamophobia.

2.2.1 The “Alerts” Functionality

The “**ALERTS**” view was created to increase awareness on Islamophobic messages at scale, monitoring the trend of hashtags and keywords over time without a focused, time-bound framework, like the one described in the “DATA ANALYSIS” section of the Platform. In this view, users are asked to select one of the pre-defined **hashtags or keywords** monitored since the beginning of the project. The system displays the trend of the selected term over time based on the number of tweets, retweets and replies, in total. The visualisation shows a **peak** when the Islamophobic hashtag or keywords have been particularly present on Twitter, corresponding to a possible alert for operators. This is computed on the fly by the system by taking into account the average frequency of messages plus one standard deviation. This value is dynamically computed for each hashtag or keyword, since some of them may be generally frequent and more present online than others. As an example, we report the analysis for the #STOPIslam hashtag in Figure 6.

Figure 6 – Hashtag trend for the #STOPIslam hashtag. Red dots signal that the hashtag appeared very often on that day (more than average + 1 SD)



SOURCE: Screenshot from the Hatometer Platform

2.2.2 The “Counter-narratives” Functionality

The main functionality for CAP intervention relies on a **chatbot-like application** that, given an Islamophobic short text in input, provides **five automatically generated suggestions** that could be used **to counter the hate speech and de-escalate the argument**. The counter-narrative suggestions are provided in the specific language of the account in use (Italian, English or French). Figure 7 displays a screenshot of the English version of the interface, where the message above is the input and those below are suggestions provided by the chatbot that could be used to build counter-narratives.

Operators using the Platform can set the input according to different strategies: if operators want to reply to one of the Islamophobic messages displayed in the **list of tweets** retrieved by the Twitter API, they can click on the speech bubble icon on the upper right of the message (see Figure 3 above). In this way, the content of the tweet is displayed as input in the counter-narratives tab, while **five possible answers** are automatically generated by the chatbot-like application. As an alternative, operators can also write an **input text**, and call the chatbot to provide possible replies, or copy and paste it from other sources.

Figure 7 – Screenshot of chatbot-like tool for the creation of counter-narratives



SOURCE: Screenshot from the Hatemeter Platform

After checking the answers, operators can choose to **use one of the responses to reply to an Islamophobic message**. Therefore, each answer can also be **edited and modified** (using the ‘pencil’ icon) and saved (using the ‘floppy disk’ icon), before copying and pasting it on Twitter or any other social media to post the reply. Operators can also write their own responses through the interface, if none of the ones proposed by the system are effective. Although having the possibility to post the reply directly on Twitter would be handier for operators, this would force them to connect the Platform to their social media account (or that of the NGO), making it necessary to share personal data to enable this connection. Since we want to avoid this, also as a means to protect operators who often interact online using fake accounts, a different solution was implemented, requiring copying and pasting the replies from the Platform.

The **chatbot-like suggestion tool** has been implemented following a **data-driven approach**. In particular, it relies on a pool of pre-existing “Islamophobic message – counter-argument” pairs that are used by the tool as examples to choose and rank possible replies given an input message. A language specific pool of pre-existing “Islamophobic message – counter-argument” pairs is used for each account. We implemented a tf-idf³ response retrieval mode, which is built by calculating the tf-idf word-document matrix of the message pairs mentioned above. The suggested responses for a new input message are obtained by finding the hate message in the **pool of examples** that is most similar to the input one and presented in the interface the top five most relevant responses.

³ Base concept in NLP. For a technical explanation, see Spärck Jones (1972).

The example pairs needed to build the tool have been created **with the help of NGOs operators** following the **same data collection procedure for each language** (i.e. English, French and Italian). The data collection has been conducted along the following steps:

1. **Hate speech preparation.** For each language, we asked two native speaker experts (NGOs' trainers) to write around 50 prototypical Islamophobic short hate texts.
2. **Counter-narratives collection forms.** We prepared three online forms (one per language) in which users can respond to hate text prepared by NGOs' trainers, operators were asked to compose up to 5 counter-narratives for each hate text.
3. **Counter-narratives collection sessions.** For each language, we performed three data collection sessions on different days. Each session lasted roughly three hours and had a variable number of operators – usually around 20. Operators were gathered in NGOs' premises with a computer and received a brief introduction on the counter-narrative collection task.
4. **Data annotation and augmentation.** After the data collection phase, we employed the services of three non-expert annotators to perform additional work. In particular, they were asked to (i) paraphrase original hate content to augment the number of pairs per language, (ii) annotate hate speech sub-topics and counter-narrative types (iii) translate content from French and Italian to English to have parallel data across languages.

In total, we had **more than 500 hours of data collection with NGOs** involved in the Hatometer Partnership, during which we collected **4,081 hate speech/counter-narrative pairs**; specifically, 1,288 pairs for English, 1,719 pairs for French, and 1,074 pairs for Italian. At least **111 operators** participated in the **9 data collection sessions** carried out face-to-face and remotely. Each counter-narrative needed about **8 minutes** on average to be composed. The paraphrasing of hate messages and the translation of French and Italian pairs to English brought the total number of pairs to more than **14 thousand**.

If operators modify a counter-narrative message or write their own messages and save them through the CAP interface, they can be automatically added to the pool of examples to increase its variability and size, so that **they can be used to suggest additional counter-arguments in the future**.

A **final evaluation experiment** showed that, as expected, using the **Hatometer counter-narrative (CN) suggestion tool** drastically **reduces the time needed to obtain a CN**: with the CN suggestion tool the time needed for composing a new CN or for modifying a suggestion is almost halved as compared to the one without suggestion tool (**8 minutes vs 4.5 on average**).

3. Hate speech against Muslims on social media: evidence and analysis

This section presents **evidence of online Islamophobia** in the three countries involved in the project, namely **Italy, France and the UK**, according to a subdivision in three parts. Each subsection starts with a brief description of the context and background of online Islamophobia and anti-Muslim hatred in the country, and then specifically presents **evidences of hate speech collected through the Hatemeter Platform**, as well as the **analysis of the information gathered** thanks to the visualisation tools available in the Platform. The three subsections are the results of the work of the three research teams involved in the Hatemeter project, namely University of Trento (Italy), University of Toulouse 1 Capitole (France) and Teesside University (United Kingdom).

3.1 Italy

3.1.1 Islamophobia in Italy

The situation of online hate speech against Muslim communities in Italy presents a distressing picture (see deliverable D7, “Guidelines on the socio-technical requirements of the Hatemeter Platform”⁴). The target groups for online hatred in Italy tend to be **migrants from all nationalities**: hatred is associated with the idea that immigration policies are economically and socially unbearable. Islamophobia becomes particularly noticeable insofar as usually migrants are linked with Muslims and Muslims are then linked with terrorists.

As in other countries, the idea of public discussion has changed with the birth of the Internet and social media. There is an increasing connection between **alternative information websites** (e.g. blogs, informal webpages not connected to a specific institution/journal), **social networks** and **traditional mass media** (especially newspapers): social networks facilitate quick and easy movements of information on hate news between alternative information websites and traditional mass media and vice-versa. ‘Hate preachers’ tend to be individuals, rather than groups and the use of social networks has replaced their use of websites and blogs (Giovannetti and Minicucci, 2015).

Fake news and **inflammatory statements** against Muslims are **spread** on the Internet and through social media platforms (Giancalone 2017). The Internet offers immediacy, pervasiveness, amplification, replicability, social validation and persistence of certain messages. At the same time, social media platforms offer a polycentric proliferation of hate speech and promote the diffusion of demagogic and propagandist messages. Importantly, the **online and the offline worlds** are increasingly **connected** and the impact that one generates on the other is often underestimated (this is the so-called ‘prejudice of the digital dualism’) (Giovannetti and Minicucci, 2015). A recent study on Islamophobia in Italy reports an increase in discriminatory articles in the newspapers and in hate speech against Islam by **Internet-based neo-fascist** and **Catholic fundamentalist groups** (Alietti and Padovan, 2018). According to another study, Muslims are the **fourth most targeted** group on Twitter (Vox, 2019). Among the criticisms and attacks of Islam, there are many that conceptualise it as a violent, absolutist, anti-democratic religion that is against and incompatible with Western values (Malchiodi, 2016). Moreover, the United Nations mentions the existence of dangerous prejudice against immigrants in Italy, especially as originating from politics and the media (Osservatorio sulle Discriminazioni, 2010). Hate speech is soaring together with xenophobia, Islamophobia, anti-Semitism and racism, as a result of both **terrorist attacks** and the **migration flows** (Bortone and Cerquozzi, 2017).

⁴ The contributors of the deliverable are: Andrea Di Nicola, Stefano Bonino and Elisa Martini (UNITRENTO), Jérôme Ferret, Mario Laurent and Jen Schradie (UT1-Capitole), Georgios Antonopoulos and Parisa Diba (TEES).

3.1.2 Evidence of Islamophobia in Italy through the Hatometer Platform

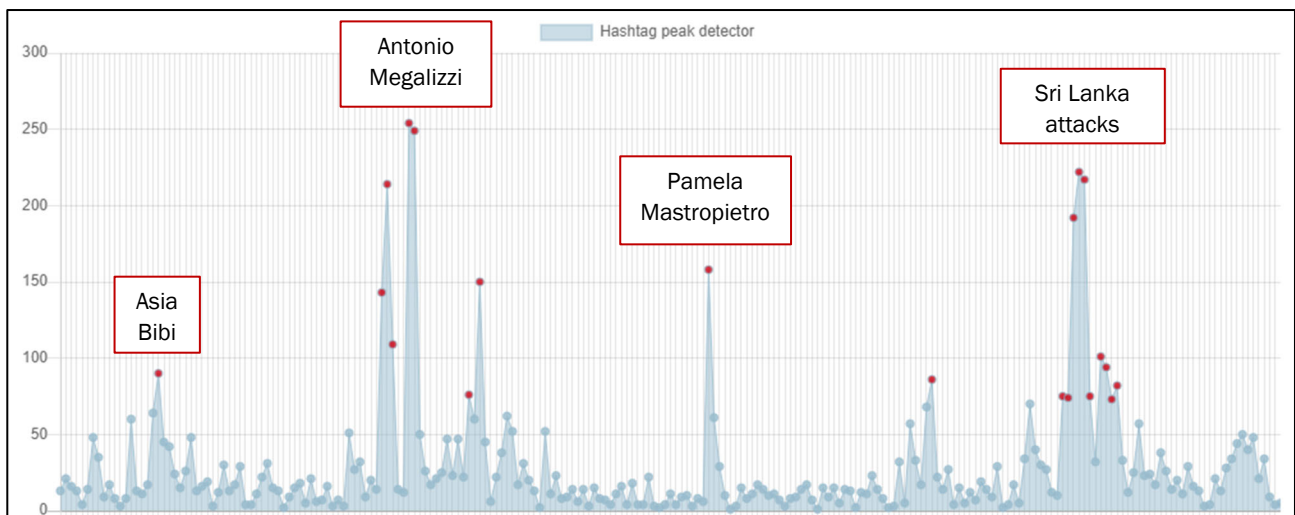
The preliminary results of the Hatometer project confirm the **negative picture** established through the literature review on Islamophobia in Italy (see section 3.1.1). Hate speech against Muslims on social media is **increasing** and closely linked to **Italians' perception of migration and terrorism**. Moreover, as it will be shown, the increase in the use of hashtags indicating hate speech is often connected to political debate and discourse.

Starting from a first list of **keywords and hashtags** entailing anti-Muslim content (for example, #StopIslam, #banislam, #NoMosques, #IslamHorsDEurope), the Hatometer Platform allowed us to monitor them on social media. On Twitter, the Platform has monitored more than 40,000 Italian tweets, 30,000 Italian replies, and 250,000 Italian retweets between September 2018 and September 2019.

Figure 8 shows four primary peaks identified using the hashtag “#STOPIslam”, which correspond to events occurring in Italy and/or abroad:

1. **The release of Asia Bibi** after 8 years, a Pakistani Christian woman convicted of blasphemy by a Pakistani court and sentenced to death by hanging (November 2018);
2. **The death of Antonio Megalizzi**, due to a terrorist attack which occurred in Strasbourg that caused in total five deaths and eleven wounded (December 2018);
3. **The start of the trial on the Pamela Mastropietro investigation case**, an 18-year-old woman from Rome murdered and dismembered by Nigerian drug dealers in Macerata. The murder caused extreme public outrage, anger and anti-immigrant sentiments to the point that in an act of revenge six African immigrants were injured in a drive-by shooting incident by a local resident (February 2019);
4. **The Sri Lanka attacks** on Easter Sunday, when three churches and three luxury hotels in the commercial capital Colombo were targeted in a series of coordinated terrorist suicide bombings. 253 people were killed, including at least 46 foreign nationals (April 2019).

Figure 8 - “Alerts” Functionality, Hashtag peak detector “#STOPIslam” in Italy



SOURCE: University of Trento elaboration – Screenshot from the Hatometer Platform

The Platform facilitates the investigation of the origin of hate speech starting from the aforementioned **four events**. More specifically, the Platform incorporates excellent analytical tools and among them, the hashtag trends Functionality, as described in sections 2.1.2. This item allows users to obtain a more fine-

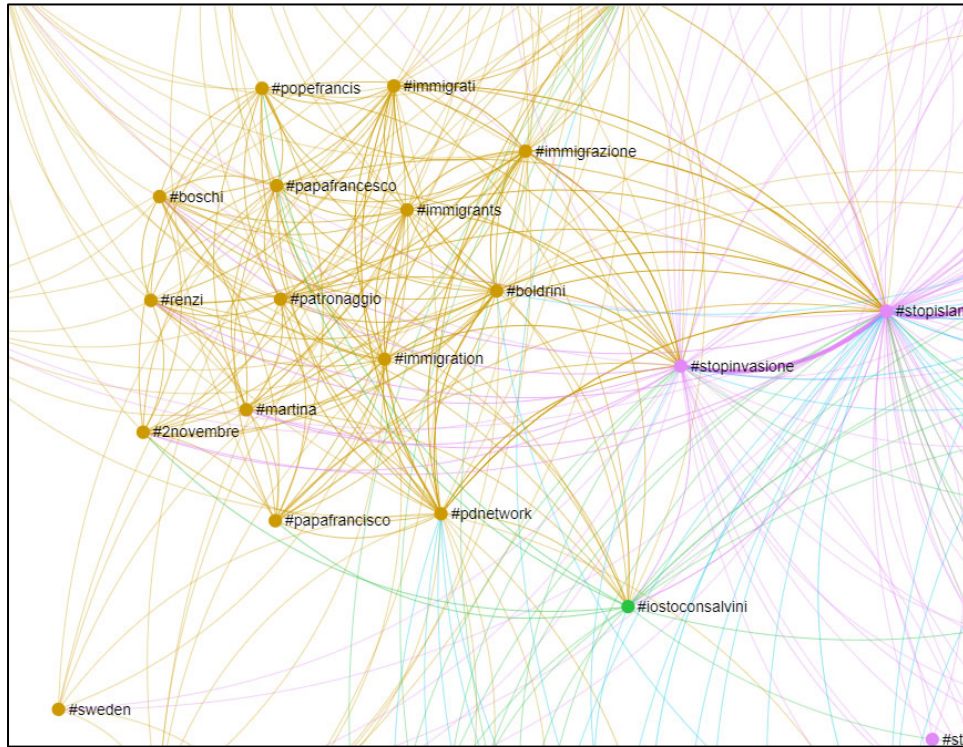
grained analysis of hashtags, focused in particular on their trend *over time*, because the information is gathered from the underlying database. Users can select a hashtag, and a temporal snapshot of interest, and then the system returns several graphic representations (i.e. general statistics on the number of tweets, retweets and replies containing the hashtag of interest, the hashtag co-occurrence network, and the day-by-day statistics showing the hashtag's presence on Twitter over time). Below, **several examples focusing on the Italian context** will be reported, showing how the Platform functionalities can be used to perform different analyses.

The **co-occurrence analysis** presents a map linking a central keyword and other associated words. The distance between #Stopislam (i.e. central word) and another associated word denotes the frequency of co-occurrence. The radial map shows the word association index, by setting the key word in its centre and displaying the other related words around the key word at a distance proportional to their association. That means that the closer the key words are to the word in the centre, the more frequent their co-occurrence with the key word is. Moreover, the Platform allows for the identification of **clusters** that are linked among them, as it is possible to notice from the different colours of the sub-networks that compose the whole network.⁵

Figure 9 shows the **co-occurrence analysis of the peak regarding the release of Asia Bibi**. As it is possible to note, #Stopislam produces a “large meshes” network, with the exception of a cluster where hashtags principally refer to migration issues and, interestingly, the closest hashtag to #Stopislam is precisely #stopinvasione (i.e. “stop invasion”). More specifically, during the period of reference for this first peak (from 26th October to 5th November 2018), there have been 45 tweets, 288 retweets and 48 replies.

⁵ The Platform relies on one of the possible algorithms usually employed to perform community detection. More specifically, the employed algorithm is “Louvain modularity”, often used in social network analysis.

Figure 9 - #STOPIslam in Italy, tweets related to Asia Bibi discharge. Co-occurrence analysis. Focus on a specific cluster



SOURCE: University of Trento elaboration – Screenshot from the Hatometer Platform

By way of example, below are tweets referring to the analysed period, written by accounts that have been subsequently suspended, so it is not possible to show the specific screenshots. In these tweets, the use of dehumanising adjectives is noteworthy (i.e. monkeys and beasts).

Example 1

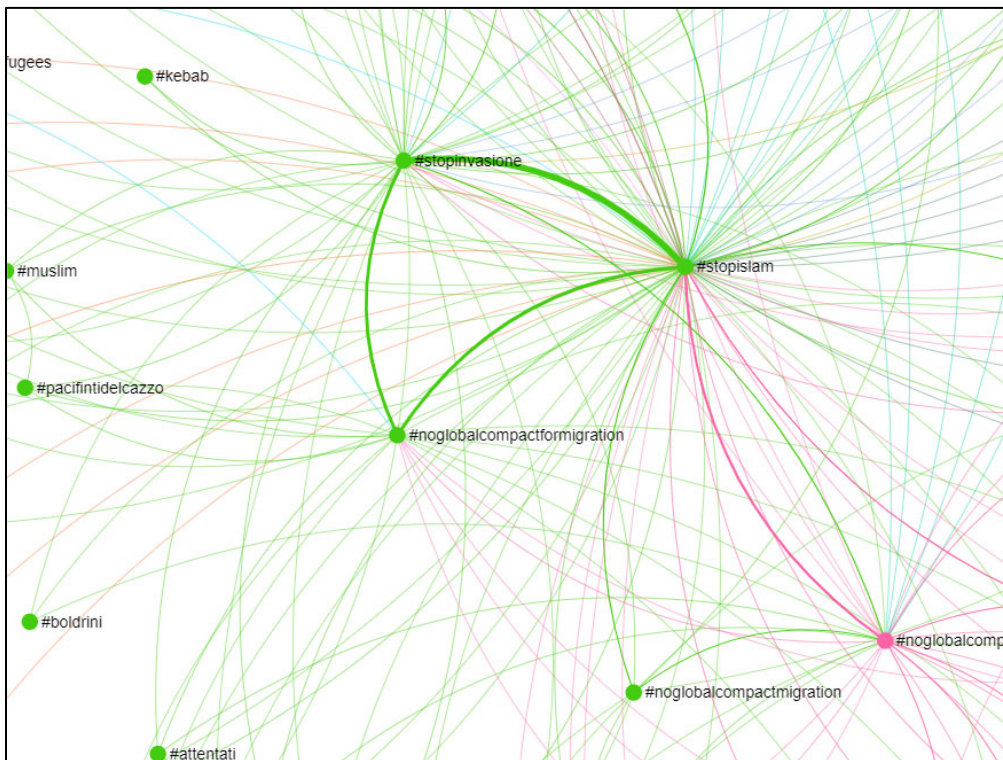
English translation: These monkeys have been burning and ransacking #Pakistan for days, because #AsiaBibi has not been condemned to death...These ones come to Italy as well, what should we expect? Model citizens respecting their neighbour and the law?! LOL #stopinvasione #StopsIslam
Original tweet: Queste scimmie stanno mettendo a ferro e fuoco il #Pakistan da giorni, perché #AsiaBibi non è stata condannata a morte... Questi arrivano anche in Italia, cosa dovremmo mai aspettarci? Cittadini modello rispettosi del prossimo e delle leggi?! LOL #stopinvasione #StopsIslam (November 2, 2018 10:10 PM)

Example 2

English translation: These beasts come in our places and, according to “do-gooding mo*ons”, with LOVE they can become model citizens paying pensions. If you import the Third World, you will become Third World, this is the only truth #stopinvasione #AsiaBibi #stopIslam
Original tweet: Queste bestie arrivano poi da noi e, secondo i cogli*buonisti, con l'AMMORE diventano cittadini modello paga-pensioni. Se importi il terzo mondo, diventerai terzo mondo, questa è l'unica verità #stopinvasione #AsiaBibi #stopIslam (November 1, 2018 12:00 PM)

The **second peak** identified in Figure 8 is connected to the **death of the young Italian journalist Antonio Megalizzi**, due to a terrorist attack occurring in Strasbourg. The co-occurrence analysis (Figure 10) shows that #Stopislam is again associated with the hashtag “Stopinvasion”, but together with #noglobalcompactformigration (i.e. “global compact for migration” is an agreement, prepared under the auspices of the United Nations, that describes itself as covering "all dimensions of international migration in a holistic and comprehensive manner"). As in the previous example, the main network has “large meshes” and it is composed of several different underlying clusters. These are connected respectively to immigration issues, national security and social threat. This peak refers to the period 11th -21st December 2018, during which 47 tweets, 559 retweets and 49 replies were published.

Figure 10 - #STOPislam in Italy tweets related to Strasburg terrorist attack. Co-occurrence analysis. Focus on a specific cluster



SOURCE: University of Trento elaboration – Screenshot from the Hatemeter Platform

Below, are reported some of the tweets employing the hashtag “NOGlobalCompactforMigration” and connected to the aforementioned event.

Figure 11 – First example of tweet about the death of Antonio Megalizzi

English translation: I DO NOT WANT THIS SHI*S IN MY COUNTRY. #NOGlobalCompactforMigration #STOPINVASIONE #StopIslam



SOURCE: University of Trento elaboration – Screenshot from Twitter

Figure 12 - Second example of tweet about the death of Antonio Megalizzi

English translation: They also rape dead bodies. #noglobalcompactmigration #stopIslam #NOGLOBALECOMPACT



SOURCE: University of Trento elaboration – Screenshot from Twitter

Figure 13 - Third example of tweet about the death of Antonio Megalizzi

English translation: Another victim of the religion of peace. #NoGlobalCompact #StopIslam #stopinvasione Strasburg attack, the Italian journalist Antonio Megalizzi is dead



SOURCE: University of Trento elaboration – Screenshot from Twitter

The third identified peak concerns the **opening of legal proceedings in the Pamela Mastropietro's investigation**. Contrary to the other two analysed events, in this case it seems that the association between different hate hashtags (always keeping #StopIslam in the centre) produces a more cohesive network, where the clusters seem more connected between them. During the analysed period (6th – 15th February 2019), 22 tweets employing #StopIslam were published, as well as 172 retweets and 6 replies.

From Figure 14, it is possible to identify two main clusters:

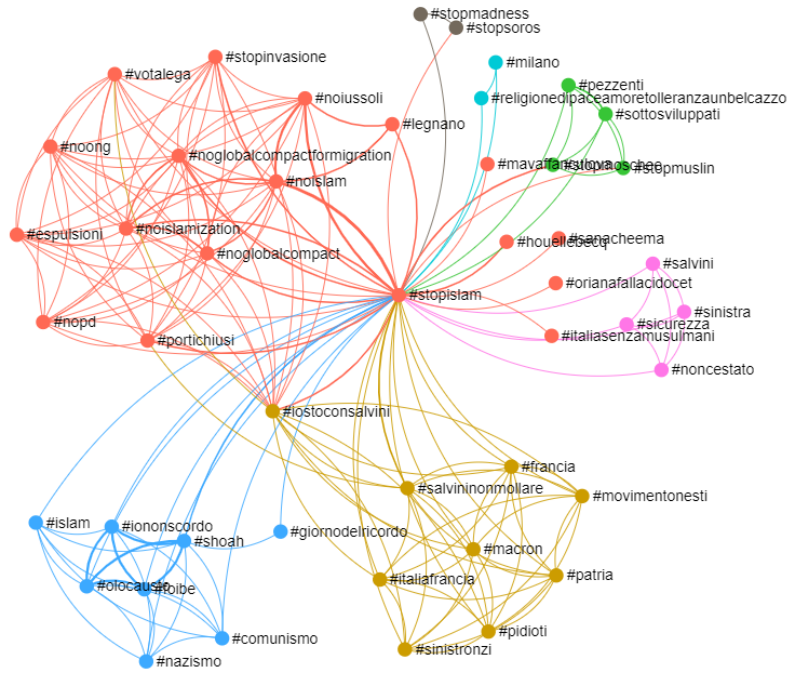
- one related to migration issues (i.e. stopinvasione, Italians first, decreto sicurezza, no ius soli, stop immigrazione, porti chiusi, espulsione, no global compact for migration);⁶
- the other associated with some particular political parties or politicians of the centre-right coalition (i.e. Salvini⁷ non mollare, io sto con Salvini, movimento onesti (movement of honest people)).⁸

⁶ English translation: stopinvasione, Italians first, security decree, no *ius soli*, stop immigration, closed harbours, expulsion, no global compact for migration.

⁷ Matteo Salvini is the former Italian Ministry of Interior.

⁸ English translation: Salvini do not give up, I stay on Salvini's side, movement of honest people.

Figure 14 - #STOPIslam in Italy tweets related to Pamela Mastropietro investigation case. Co-occurrence analysis.



SOURCE: University of Trento elaboration – Screenshot from the Hatemeter Platform

As for the other events, also reported here are some of the tweets connected to the start of the trial on Pamela Mastropietro’s investigation case.

Figure 15 - First example of tweet about the start of the trial on Pamela Mastropietro’s investigation case

English translation: WHAT DO THEY WANT, THESE MUSLIMS? LET THEM GO BACK TO THEIR “WONDERFUL” MUSLIM COUNTRIES. WHO WANTS THEM? WE DO NOT F*CKING NEED THESE BRAIN DEAD F*CKERS STUCK IN THE MIDDLE AGES #F*CK #StopIslam



SOURCE: University of Trento elaboration – Screenshot from Twitter

Figure 16 - Second example of tweet about the start of the trial on Pamela Mastropietro’s investigation case

English translation: Wipe out Muslims together with all “welcomingmo*ons” #stopIslam.



SOURCE: University of Trento elaboration – Screenshot from Twitter

Figure 17 - Third example of tweet about the start of the trial on Pamela Mastropietro’s investigation case

English translation: #StopIslam Muslims get the f*ck out.



SOURCE: University of Trento elaboration – Screenshot from Twitter

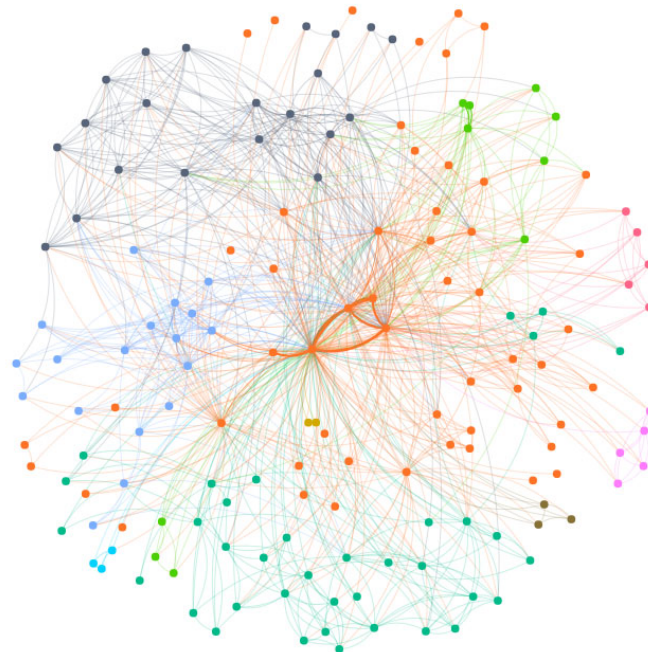
The fourth and last analysed event concerns the **terrorist attack in Sri Lanka that happened on Easter Sunday**. In Italy, as in the rest of the world, this event also led to a significant increase in the number of hate tweets against Muslims communities. For instance, in the examined period (17th – 26th April 2019) there were 85 tweets, 752 retweets and 87 replies. Moreover, as it is possible to notice in Figure 18, the main network appears very dense and the clusters particularly cohesive.

From this first analysis of the keywords appearing in the co-occurrence analysis on the fourth event, it seems that hate speech rotates around three main aspects:⁹

⁹ When users want to look at the whole co-occurrence network, the Platform does not allow visualising the keywords. These are visible only when the image is zoomed. Contrary to the previous analyses, in this case researchers preferred to show the whole network and just list the words in the text.

- 1) Us vs them: *Italy, Italian, our, country, home, Europe, west;*
- 2) Dehumanising adjectives: *parasitic, disgusting, shit;*
- 3) Social Threat: *terrorist, terrorism, attack, closed ports, defence, invasion.*

Figure 18 - #STOPIslam in Italy tweets related to Sri Lanka attacks on Easter Sunday. Co-occurrence analysis.



SOURCE: University of Trento elaboration – Screenshot from the Hatometer Platform

Below are examples of examples of particularly interesting tweets on the fourth event.

Figure 19 - First example of tweet about the start of terrorist attack in Sri Lanka happened on Easter Sunday.

English translation: #Roma MOAROCCAN MUSLIM TRIES TO SLAUGHTER A MAN WEARING A CRUCIFIX AND SHOUTS: "ITALIAN CATHOLIC SH*T" #StopIslam #BanIslam #GETTHEF*CKOUTTAHERE.



SOURCE: University of Trento elaboration – Screenshot from Twitter

Figure 20 – Second example of tweet about the start of terrorist attack in Sri Lanka happened on Easter Sunday.

English translation: A BRAIN-DEAD ITALIAN CONVERT TO ISLAM AND A MOROCCAN WERE READY TO SLAUGHTER ALL OF US IN THE NAME OF ALLAH #ReligionofpeaceandtoleranceABIGSH*T #StopIslam



SOURCE: University of Trento elaboration – Screenshot from Twitter

Figure 21 - Third example of a tweet about the start of terrorist attack in Sri Lanka occurring on Easter Sunday.

English translation: A sura a day keeps Islam away! Quran 47.4 do not look for peace with non-believers, decapitate them when you catch them #stopislam



SOURCE: University of Trento elaboration – Screenshot from Twitter

3.2 France

3.1.1 Islamophobia in France

Deliverable D7, “Guidelines on the socio-technical requirements of the Hatemeter Platform”, has previously included an explanation of the situation of online Islamophobia and anti-Muslim hatred in France, which will be recalled briefly here to provide some context to the analyses.

Anti-Muslim hatred in France is inspired by both **reasons rooted deep in the country's history**, including French colonial history and the Algerian War, and **revived by recent events** involving multiple terrorist attacks claimed by Islamist extremists in France since 2012 (Cediey and Foroni, 2006; CNC DH, 2017). Nevertheless, several other parameters contribute in the construction of hate speech, and here we focus on the **influence of the law** and **influence of the general society** (i.e. politicians and citizens).

Starting from the **legal perspective**, French law condemns, through article 24 of the Law of 29 July 1881 (modified in 2012) and the Penal Code (article R624-398), both **direct hate speech** (insults, threats, etc.) and **incitement to hate speech**, directed against any person because of their religion or ethnic origin. These legal provisions apply on **public** and **private** speech, and cover **all technical means of expression**, including digital content.

In addition to the law mentioned above, in July 2019 a new text of law was voted at *Assemblée Nationale* specifically addressing **online hate speech**.¹⁰ It prohibits a range of different content on social networks (i.e. Facebook, Instagram, Twitter and YouTube), including abusive messages incitement to hate, discriminate or commit violence against people, based on their sex, sexual orientation, gender identity, disability, ethnicity, nationality, "race" or religion. Any French citizen has the right to report unlawful content. If a message or content is clearly illegal, the law would impose that social network has to delete it in less than 24hrs and replace it with a message indicating the removal. The law punishes the lack of cooperation by the social network with one year's imprisonment and a fine of 250,000 euros.

During the debates preceding the adoption of the text by the Assembly, activists expressed some complaints regarding the epistemological issues on Twitter and asked to replace the term **“Islamophobic”** with **“anti-Muslim”**. Indeed, they feared that “Islamophobia” could have been later used to prohibit any criticism of the religion itself rather than to defend the rights of Muslim citizens. This debate once again highlighted the particular tension around the place of Muslims in French society and the epistemological battles over terms qualifying the hatred of Muslims.

Concerning the **influence** that **general society** can exert on Islamophobia, the necessity has arisen to differentiate between **hatred on the grounds of religion** and **racism in term of language designation**, although laws punish and include all forms of discrimination. For several years now, it is possible to note a shift in the way **politicians** discuss and understand racism. Tahata (2018) analysed the discourses at the French Assembly from 1981 to 2018. He found that during the 1980's, the far-right and a part of the right were holding an identity discourse where the "French" did not need to be defined (by virtue of its majority status) and was opposed to the "other". After the 2000s instead, politicians tend to regularly define the French group and associate it with strong moral values, apparently typical of French society and the Republic. A foreigner is viewed not just as “the other”, but as a person with a different culture and different values that are not compatible with French society. This way of viewing foreigners is shared by representatives from all political groups (not only the far-right), particularly because of the complicity of a secular and atheistic society. Froio (2018) stresses the existence of this new opposition between French culture and foreign culture, seen as invasive also among citizens and she distinguishes between

¹⁰ This new law is entitled ‘*Lutte contre la haine sur internet*’, it still needs to be reviewed by the Senate before being applied. For more details, see http://www.assemblee-nationale.fr/dyn/15/dossiers/lutte_contre_haine_internet

two racist positions, namely "**racism**" and "**neo-racism**". The first position is based on biological arguments, while the second is supported by cultural and religious arguments, claiming incompatibility between two civilizations. According to the principles of this second position, the values of the French Republic and secularism can justify the exclusion of Muslims.

The aforementioned different positions on Islamophobia and racist trends are reproduced in **hate speech online**. In France, it is possible to differentiate **three groups of hate speakers**, likely to engage in hate speech directed against Muslims, and **two main** ways to proceed to attack Muslims (according to information gathered from respondents and online observations during the development of the project, see also deliverable D7 "Guidelines on the socio-technical requirements of the Hatemeter Platform"). The three identified groups are:

1. **The far-right groups**, called *Fachosphère*, including political parties, websites communities, activist groups (*Rassemblement National, Génération Identitaire, FdeSouche*, etc.). Their discourse relates many societal problems to immigration and *choc des cultures* (cultural clash). They often demand the expulsion from France people whose culture, it is asserted, cannot bend to their own cultural reference.
2. **Some members of the Catholic conservative right**. Their speeches defy the presence of Muslims in public space on the pretext that they must respect the historical right to govern and to speak in public held by Judeo-Christian white culture.
3. **Some of the secularist claimant-activists**, sometimes affiliated to a left-wing party. Members of this third category follow an extreme interpretation of secularism such that no sign or religious claim should be tolerated in public space. This group has the most sophisticated discourse, in which hate speech, when existent, is subtle. However, they can launch harassment campaigns targeting Muslim personalities in order to silence them, regardless of their role as feminist activists, anti-racists or citizen who presented themselves role models for French Muslims in society.

As for the two main ways to attack Muslims, we have distinguished:

1. **Dissemination of stereotypes** associated with Muslims or people of North African descent and culture, but not targeted toward any individual. Such sentiments effect all Muslims who spend free time on social networks and witness these hateful messages, but it can also convince impressionable third parties that Muslims are indeed the source of many problems affecting society.
2. **Cyberbullying campaigns**, targeting publicly known and recognised Muslim personalities or any person who has been exposed in the media as a Muslim. They include a range of high-profile people in France, from journalists to athletes who can be identified as Muslim (in terms of physical appearance, name, and for women, those wearing a hijab). A connection can be made, considering any visible religious sign in the media as proselytism; proselytism is then associated with extremism; extremism itself associated with support of terrorist groups.

Twitter mechanics also play a role in the form of hate speech. The maximum size of tweets prevents the development of a constructed speech and forces writers to be more concise and often more divisive in their statements, in order to be clearly understood. Furthermore, the emphasis upon gaining followers on the system pushes users to make inflammatory remarks to attract attention and increase their audience. Thus, the exchanges between support and anti-Muslims groups incendiaries. On that note, a linguistic study by Longhi (2017), on the correlation between the content and form of tweets, showed that the effectiveness of a tweet (i.e. the number of times it has been shared) seems to be linked to the controversial nature of its content.

In addition, Twitter moderation policy pushes **anti-Muslim influencers** to adopt more subtle speech, which is often coded, and is difficult to identify and characterize as hateful. The task of identifying and confronting hate speech therefore, becomes a significant challenge. Finally, **hate groups are well-organized** and, rather than just propagate their ideology, they often follow anti-racist NGOs activities to undermine counter-narrative initiatives by interfering in the comments section.

3.1.2 Evidence of Islamophobia in France through the Hatometer Platform

The **data collected for the French part of the Hatometer Platform** was compiled from Twitter, commencing in October 2018. The data collected includes all the tweets employing the following hashtags and keywords:

#EtatIslamique	#IslamAssassin	#IslamHorsdEurope	#IslamDehors
#Islamisation	#Islamophobie	#StopCharia	#StopIslam
#StopIslamisme	#Invasion Musulmane	#Musulmans	

These **hashtags and keywords** were considered initially the most relevant relying on the observation of their use on Twitter and on the help of the NGO partner of the project (i.e. ICCF). However, during the development of the project, new hashtags and keywords has been added to the collection of tweets, including:

#hijab	#GrandRemplacement	#remigration	#laïcité
--------	--------------------	--------------	----------

A total of **151,738 tweets, 268,548 comments and 1,206,347 retweets** have been stored in the database from September 2018 to May 2019.

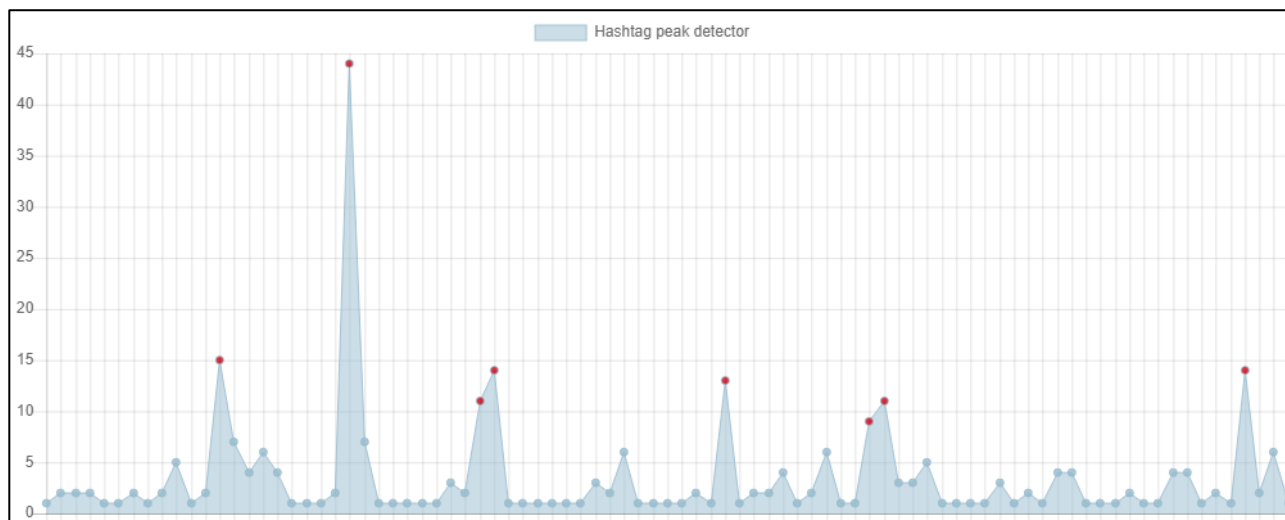
Some of these hashtags and keywords are those employed by hate speakers, whereas others are used by people talking about anti-Muslim hate itself or the place of Muslims in society. Both are useful as they help us to analyse the event in relation with the evolution of the online hate speech.

As mentioned, the visualization tools contained in the Platform allow a direct **qualitative analysis**: first, because it already presents the data in an interpretable form; second, also thanks to the interactivity of the interface, which allows direct navigation from the **graphical representations** in the Platform to the tweets and twitter profiles on the Twitter website.

The Platform's functionalities allow responses to several questions on the topic: *do we see different types of hate speech? How often do they appear? What scale of audience do they reach, i.e. what is the size of the audience for each type of speech?*

Figure 22 shows the first example of a **peak identified employing the “Alert” tool of the Platform** and the hashtag **“#Islamhorsdeurope”** (i.e. Islam out of Europe).

Figure 22 – “Alerts” Functionality, Hashtag peak detector “#Islamhorsdeurope” in France



SOURCE: University Toulouse Capitole elaboration – Screenshot from the Hatemeter Platform

The **alerts** concerning #Islamhorsdeurope follow a rather classic pattern: the term is mentioned between one and four times a day for most of the period considered and there are only a few peaks, corresponding to specific events. Each event can then be identified by using the Platform functions: on the ‘Hashtag trends’ section, by selecting the right time span, the most shared tweet can be read to deduce the event which generated the peak. Here, the **largest peak**, with 44 mentions, corresponds to the day after the date of the **terrorist attack in Strasbourg**, claimed by the Islamic State. Here, the direct association between Islam in Europe and terrorism is clear, but we can see that this hashtag had a relatively small audience. In comparison, the continuous news channels in France, who chose the place in society of women wearing the veil a main topic of their talk show regularly during the month of October 2019, often receive around 200,000 viewers.¹¹

The hashtag #Islamophobie (i.e. Islamophobia) is much more widely used (see Figure 23), probably because not only extremist movements employ it, but also activists denouncing discrimination and people writing ironic tweets complaining about the victimization of Muslims. Over the period from October 2018 to March 2019, we can see two large peaks.

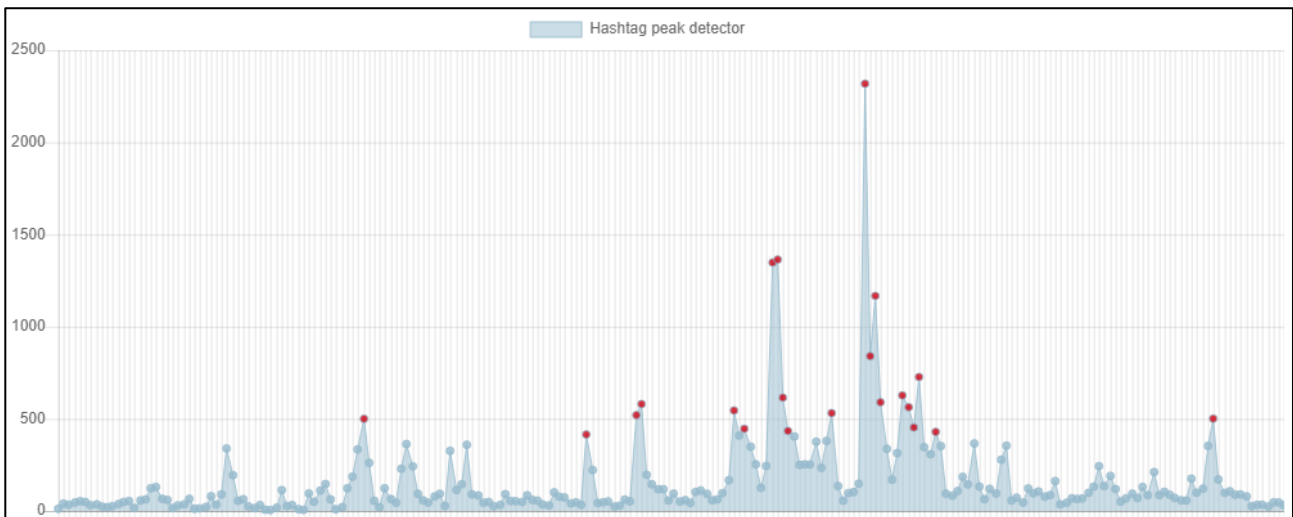
- The first, with two consecutive days of around 1,300 mentions, appears at the time of the **controversy over the sale** by a chain of sports goods stores, of a **“running hijab”**, a veil intended for Muslim women in a material suitable for running. The announcement of the possible placement of this veil in stores triggered a wave of virulent reactions from people opposed to the wearing of the veil in public spaces in France. These subsequently triggered reactions to denounce the stigmatization of veiled women with the hashtag #Islamophobie.
- The second peak comes on the day of the **attacks on Muslims in two mosques in Christchurch**, New Zealand. It is important to note that the perpetrator stated that he was inspired in his hatred

¹¹ The information is retrieved from a website that provides daily information on audience: <http://www.toutelatele.com/face-a-l-info-bilan-d-audience-derapages-de-zemmour-direct-supprime-cnews-gagnante-114499>.

It seems quite important to have an idea of the proportion of audience between Twitter and television programs, to imagine the range of influence they can exert that, in turn, could develop in incitement to hatred.

of Muslims from a trip to France during which he allegedly felt that Muslims had invaded the country. In an accompanying manifesto explaining his act, he refers to the "Great Replacement" theory, which predicts the replacement of the Western population by a population of Muslim culture and which is derived from a book by the far-right French author Renaud Camus.¹² This event therefore had a double impact on Twitter in France, through the horror of the attack itself and through the revelations regarding the terrorist's inspirations.

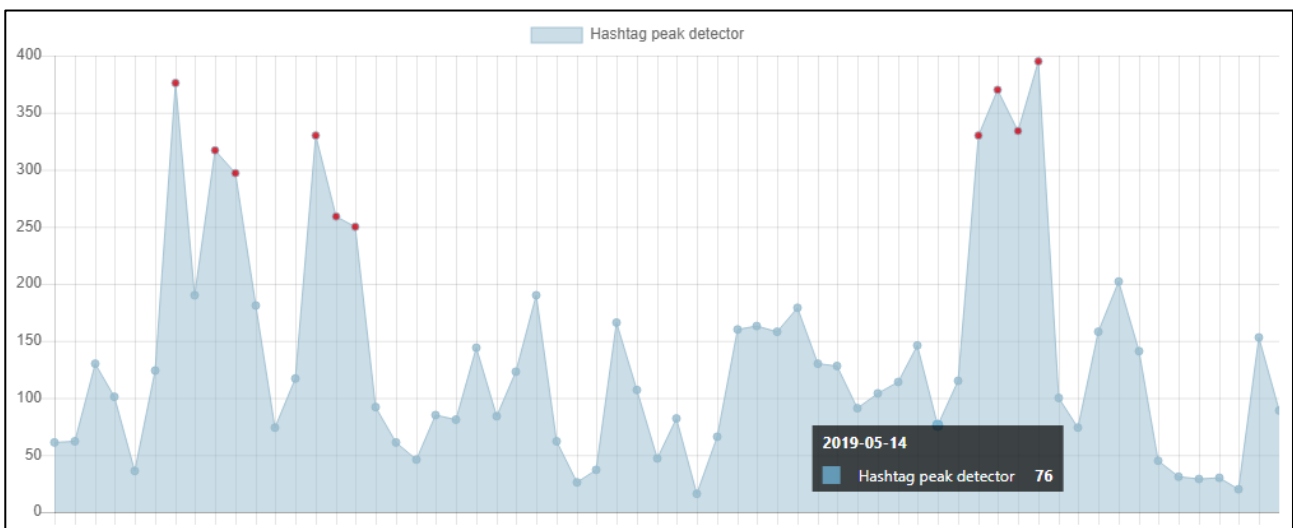
Figure 23 - "Alerts" Functionality, Hashtag peak detector "#Islamophobie" in France



SOURCE: University Toulouse Capitole elaboration – Screenshot from the Hatemeter Platform

Following the Christchurch attack, we added the hashtag #GrandRemplacement in March to the data gathering in our Platform.

Figure 24 - "Alerts" Functionality, Hashtag peak detector "#GrandRemplacement" in France



SOURCE: University Toulouse Capitole elaboration – Screenshot from the Hatemeter Platform

¹² See this article for sources and explanations: https://www.lemonde.fr/les-decodeurs/article/2019/03/15/la-theorie-du-grand-replacement-de-l-ecrivain-renaud-camus-aux-attentats-en-nouvelle-zelande_5436843_4355770.html

Looking at the graph generated for #GrandRemplacement (see Figure 24), we can see that it is used very differently: it is mentioned every day over the entire period of observation, at least twenty times per day but never exceeding 400 mentions in a day. Again, the audience (i.e. twenty mentions) may not seem significant on Twitter, but the fact it is used every day must be taken into account. This hashtag is **not linked to specific media events** and follows the first logic of disseminating hatred explained in the previous section (i.e. Dissemination of stereotypes). It contributes to the **long-term maintenance of certain stereotypes about migrants, including Muslims**, and helps to point the finger at these populations as responsible for societal problems. Initially, this hashtag has been used to convey the fear of an invasion and replacement of the current Western population by a population of migrants of different origin and culture. Muslims are regularly included in the targets of this hashtag. However, Figures 25, 26 and 27 show the employment of #GrandRemplacement for other very different topics.

Figure 25 displays the use of the hashtag according to its **original sense**. The user wishes to warn his followers against the supposed danger for his white Judeo-Christian culture by the increasingly visible presence of other cultures, represented here by a proportion of black children in a primary school class photo of Bagneux, a city near Paris. In this tweet, the author is using irony to reinforce his statement. The ironical tone is supposed to show that the facts are at the same time obvious but hidden by someone or something. This language process recalls the mechanisms used by conspiracy theorists to make their claims more likely.

Figure 25 - First example of tweet employing “#GrandRemplacement”

English translation: The “#great replacement” is a far-right fantasy, here is the evidence number 8,743.



SOURCE: University Toulouse Capitole elaboration – Screenshot from Twitter, through the Hatemeter Platform

The second example on the use of **#GrandRemplacement** (see Figure 26) shows how **far-right activists** try to convince others to join their fight against Muslims by holding them responsible for a problem that cannot leave anyone indifferent. Here, the author of the tweet retweeted an article that reported alarming statistics on the number of acts of violence against women in French society. In the text of his tweet, he specifies that French culture, fundamental rights in France, and the safety of women and children, are in danger because of the ‘great replacement’ and therefore because of people of Muslim culture.

Figure 26 - Second example of tweet employing “#GrandRemplacement”

English translation: With the crossbreeding, the Great Replacement and Islamisation, France is losing its fineness, its cultural legacy, its genetic heritage, its social peace, its social achievements, its hard-won women's rights, its children's rights
 (The text shared within the tweet is entitled ‘Physical or sexual violence by the partner or ex-partner’ and provides statistics on this subject in France)



SOURCE: University Toulouse Capitole elaboration – Screenshot from Twitter, through the Hatemeter Platform

In the next example (see Figure 27), the author asserts that **newcomers** to Belgium (migrants) are responsible for the destruction of a medieval bridge. Although the article quoted under the tweets states that the bridge has been destroyed to let larger ships pass by, the tweet text implies that native Belgians would have protested against the destruction, but that protest is no longer possible with the country's new population.

Figure 27 - Third example of tweet employing “#GrandRemplacement”

English translation: What a pity... of course, that happens in Belgium, a country very affected by the #great replacement. The “new” Belgians do not give a d*mn about old stones.
(The article shared within the tweet is entitled ‘Destruction of the bridge starts’)



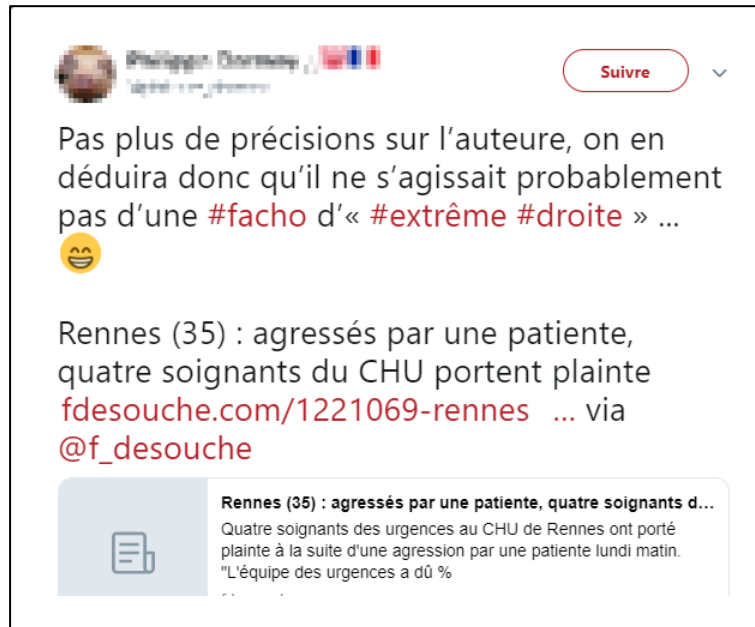
SOURCE: University Toulouse Capitole elaboration – Screenshot from Twitter, through the Hatometer Platform

Figure 28 reports the last tweet example, where the news shared is from the far-right website ‘FdeSouche’. It is a report of aggression against medical professionals in a hospital. The tweet’s author implies that if a crime is committed without any name mentioned in the article, the criminal is probably a North African, otherwise the criminal would have been pointed out as a dangerous far-right activist. This tweet follows a similar logic to the one of the **conspiracy theories**. When a news article reports an act of violence, with information suggesting the criminal could be Muslim or North African, it is very often shared on the ‘*Fachosphère*’ network using ironic hashtags such as:

- #encoreunsuédois (which means ‘it is a Swedish again’, reinforcing the idea all the violent crimes are committed by North African);
- #unechancepourlafrance (which is ironically saying ‘a chance for France’, using a sentence pronounced by former French President François Hollande who was saying that welcoming migrants can be a chance for France).

Figure 28 – Example of far-right tweet

English translation: No details on the aggressor, so we can assume it was probably not a #‘far-right’ #fascist.
 (Text shared: ‘Rennes: assaulted by a patient, four caregivers of the hospital file a complaint’)



SOURCE: University Toulouse Capitole elaboration – Screenshot from Twitter, through the Hatemeter Platform

To conclude this short data analysis by using the Hatemeter Platform, we must specify that the data we collect and their interpretation are highly **dependent on the back and forth between traditional media and social networks**. Data also depends both on the interpretation of laws and the debate on epistemology, which are reactivated after any new mediatized act of discrimination against Muslim citizens.

Indeed, to take into account all the **complexity of the phenomenon and the impact of recent events**, the analysis of anti-Muslim speeches must be carried out over a long period of time and requires a minimum distance. This difficulty seems paradoxical given the unprecedented influx of data available to us, but it is also found in the analysis of other recent upheavals in French society, that are also widely discussed on social networks, challenging traditional reading grids. For example, the social movement of the "gilets jaunes" that does not match the usual political divisions, feminist movements on social networks that shift the usual editorial lines after each new public case of violence or harassment, or contradictory debates on ecological issues that also take place both in the media space and on social networks.

3.3 United Kingdom

3.3.1 Islamophobia in the UK

In the UK context, a report by The Runnymede Trust (1997) concerning **Islamophobia in the UK** identified seven main elements:

- a) Muslim cultures are seen as monolithic;
- b) Islam is perceived as implacably threatening;
- c) Islam's adherents use their faith for political or military advantage;
- d) Muslim criticism of Western cultures and societies is rejected out of hand;
- e) Islamic cultures are substantially different from other cultures;
- f) the fear of Islam is mixed with racist hostility to immigration;
- g) Islamophobia is assumed to be natural and unproblematic.

Richardson (2004: 232) found that the British media engage in the stereotyping of Muslims, illustrating them as being 'cultural[ly] devian[t] ... and as cultural threat', thereby fuelling anti-Muslim sentiment in society. Research conducted by Poole (2006, 101) demonstrates that British news media frame **Muslims as a threat to security in the UK and as a threat to British mainstream values**. Furthermore, the news media often claim that there are fundamental differences between Muslims and non-Muslims that creates tensions in inter-personal relations, and that Muslims are 'increasingly making their presence felt in the public sphere' (Ekman, 2015).

Several studies have been carried out within the British context. Moore *et al.*, (2008), studied the representation of Muslims in the press for almost a decade, from the time period between the years 2000–2008. After the analysis of almost a thousand newspaper articles, the images used in them and of a series of case studies, they concluded that British Muslims appeared in a restricted number of contexts in the media; as a threat (due to their terrorist natural tendencies), as a problem (because their values clash with the ones proper of British culture) or both. The images of Muslims found in the press were mainly of males (rather than females), depicted either in mugshots or being involved in religious practices. The comprehensive work of Baker *et al.*, (2012) reached similar conclusions in its study of the contexts in which the word 'Muslim' appeared in British broadsheets and tabloids from 1998 to 2009. The findings of the qualitative and quantitative analyses were complementary: Muslims were depicted as a homogeneous group associated with conflict who find it very difficult to integrate with the West (i.e., British culture). A study undertaken by Whitaker (2002: 55) in 2002 stated that, in general terms, Muslim representation in the British Press was characterized by four persistent stereotypes: Muslims were intolerant; misogynistic; violent or cruel; and strange or different. Worryingly, Aguilera-Carnero and Azeez (2016) state that current studies show no change in the situation.

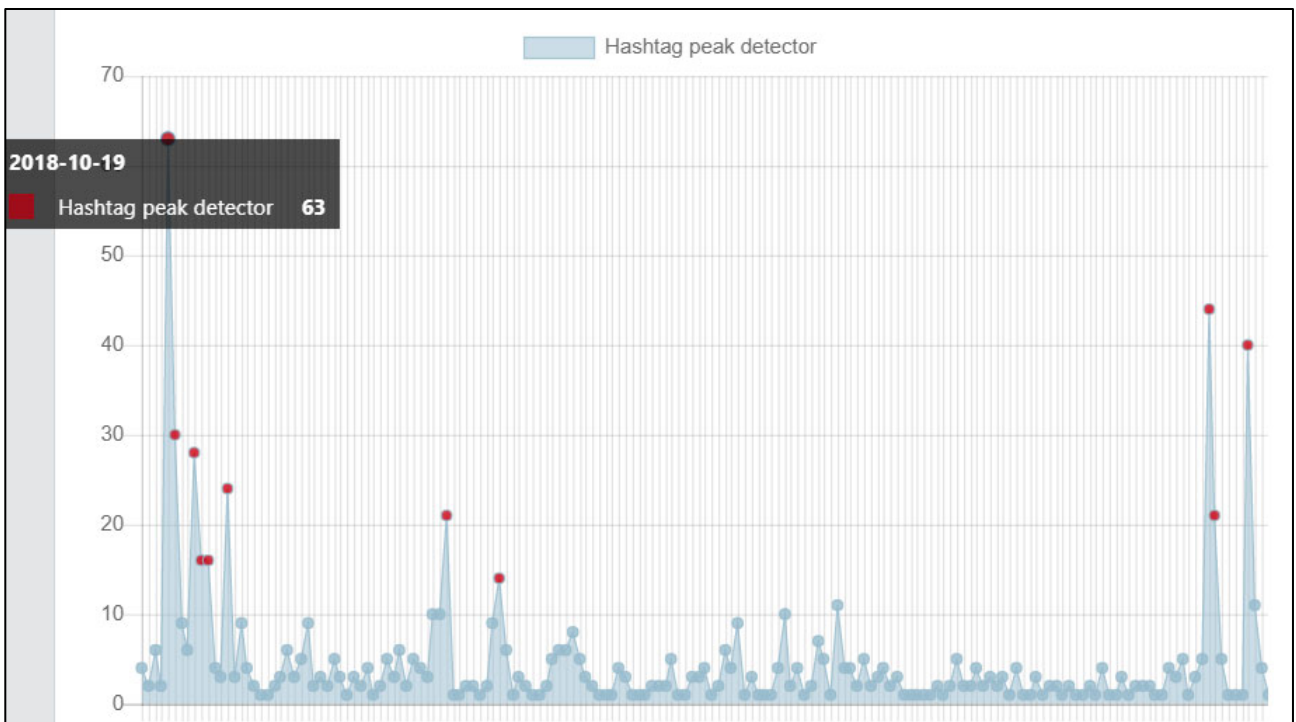
Five main themes have emerged from the preliminary **qualitative research** undertaken in the UK context through a **virtual ethnography** (see D7 "**Guidelines on the socio-technical requirements of the Hatemeter Platform**"), namely: 'Muslims as Invaders to Europe/UK', 'Muslims as Sexual Exploiters and/or Paedophiles', Dehumanisation of Islam and Muslims, 'Islam as a Cult, Not a Religion', and Counter-Narratives.

3.3.2 Evidence of Islamophobia in the UK through the Hatemeter Platform

The analysis in the UK context commences with the utilization of the "**Alerts**" feature of the **Hatemeter Platform**, created to increase awareness of Islamophobic messages at scale and to monitor the trend of hashtags and keywords over time, without a focused, time-bound framework. Monitoring organisations, such as Tell MAMA (Measuring Anti-Muslim Attacks) suggest that levels of Islamophobic discourse online

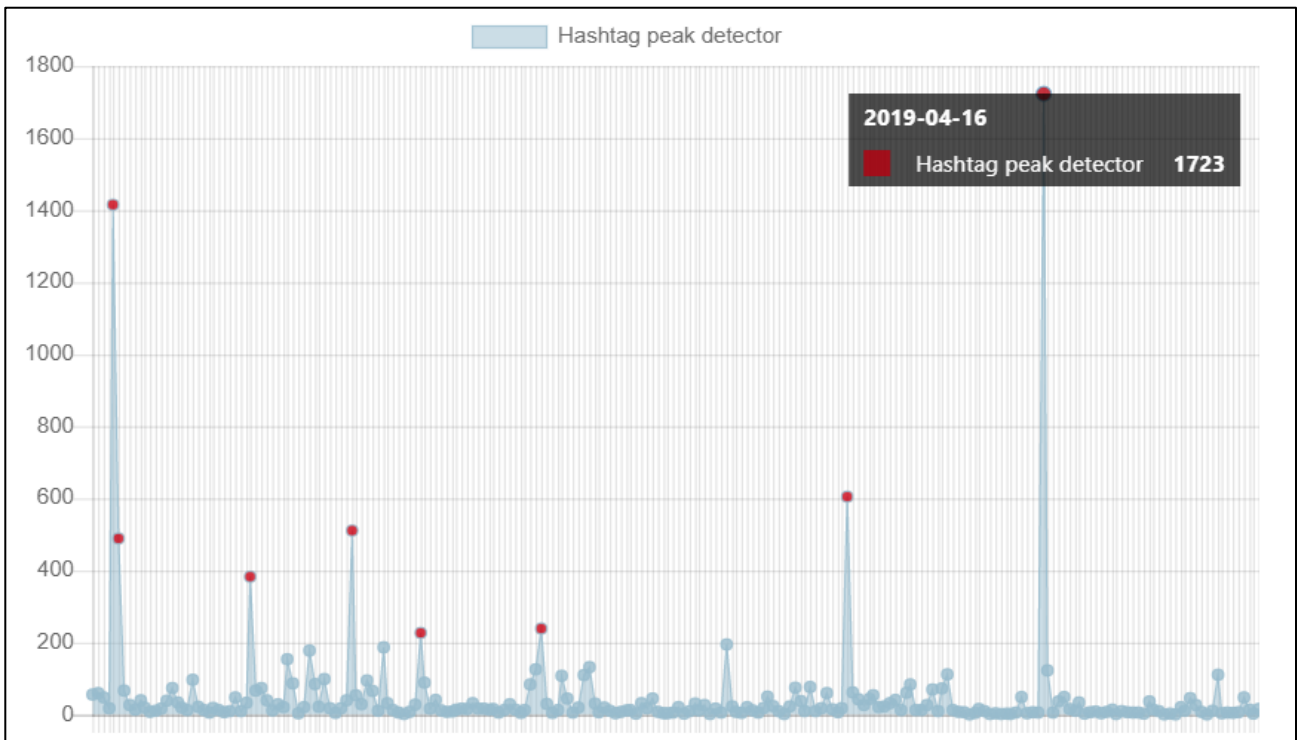
are ongoing and increasing, especially through Twitter (Allen 2014), in which anti-Muslim hashtags are frequently created and used. Here, **hate incidents are principally elevated after ‘trigger events’, including terrorist attacks**, which normally occur between 24-48 hours online (Tell MAMA, 2018). On this issue, the “Alerts” function serves as a useful feature that displays peaks of activity in hashtags, how active they are in real-time, as well as showing high occurrences of online Islamophobia and anti-Muslim hatred. Figures 29 and 30 display the signals peaks resulting from the selection of **#MuslimPaedos** and **#RapeJihad** respectively, which appear to have been particularly present on Twitter, corresponding to a possible alert to investigate further. If a user hovers their cursor over a **peak**, specific information in the form of a date and a “hashtag peak detector” number is revealed.

Figure 29 – “Alerts” Functionality, Hashtag peak detector “#MuslimPaedos” in the UK



SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

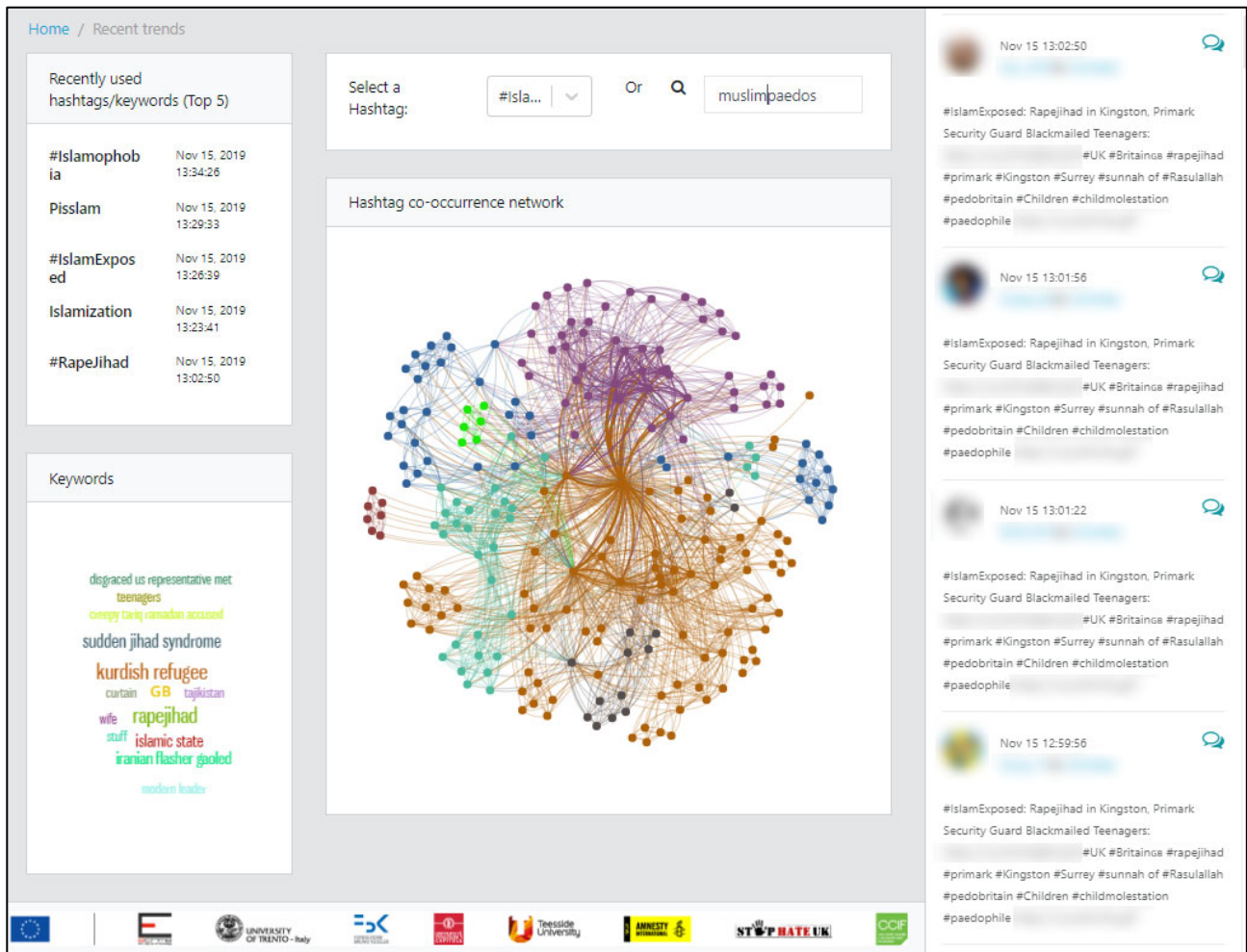
Figure 30 – “Alerts” Functionality, Hashtag peak detector “#Rapejihad” in the UK



SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

After having identified the peaks over time with the “Alert” function, the **“Recent Trends” feature** enables users to see the trends and activities on Twitter within pre-defined hashtags and also specific words, within the 100 most recent tweets. Figure 31 below displays an example with the selected hashtag #IslamExposed, along with the keywords “Muslim Paedos”. In the “Recently used hashtags/keywords”, the system ranks the most recently used Islamophobic terms with the corresponding date and time when they were last posted on Twitter. The Hatemeter Platform also provides a **direct link with hashtags and particular words**, in addition to generating keywords within the tweets, within the **“Hashtag co-occurrence network”** and the **“Keywords”** functions. Here, “Co-Occurrence Network” is the graphic visualisation of the potential relationships between people within written material. If a user selects a hashtag from a list, the **“Hashtag co-occurrence network”** enables the network of hashtags that co-occur with the chosen hashtag to be displayed.

Figure 31 – “Recent Trends” Functionality, utilising “IslamExposed” hashtag and keywords of “Muslim Paedos”



SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

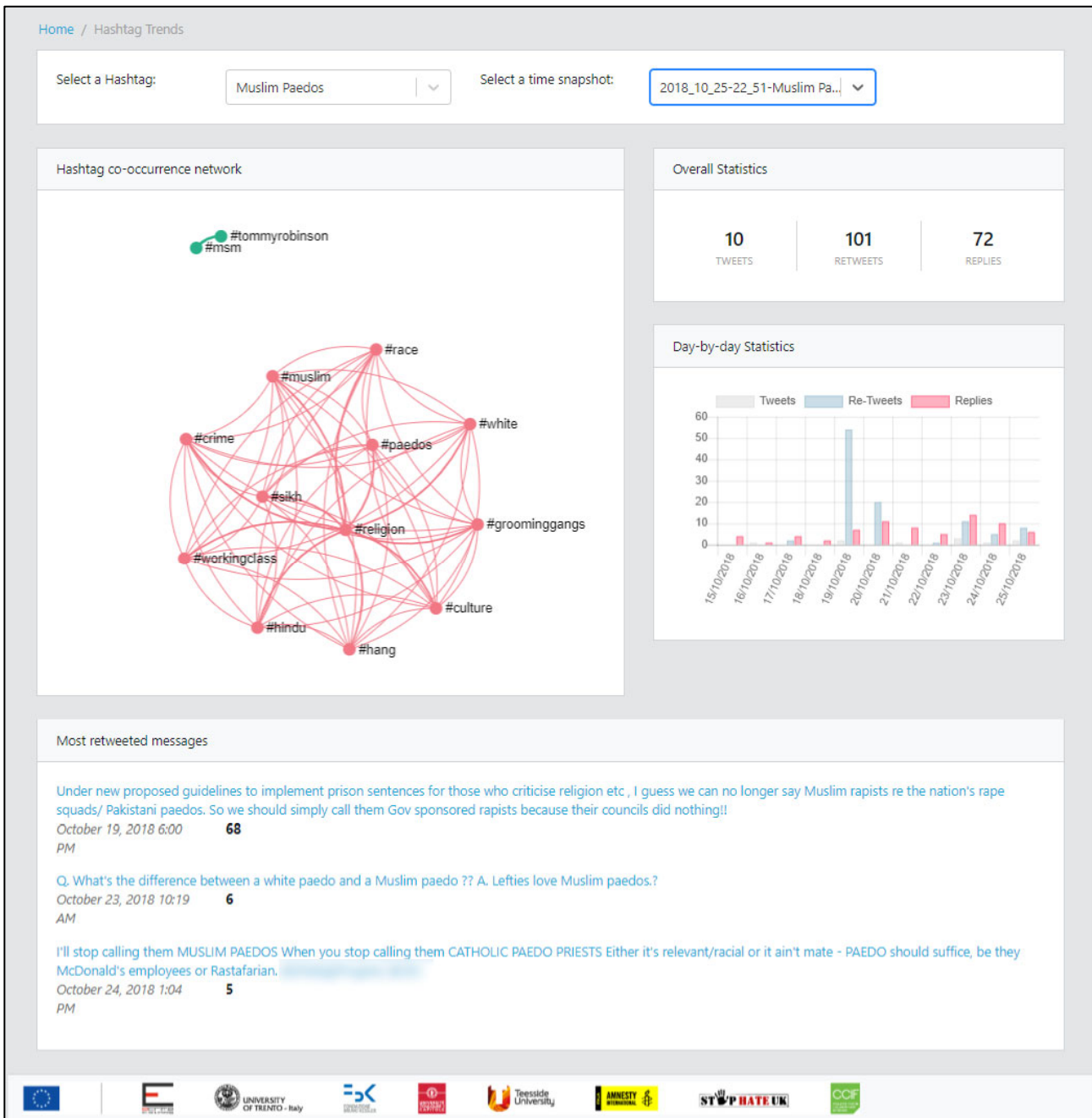
In the above example, it is evident that the “#IslamExposed” hashtag and keywords of “Muslim Paedos” have a high degree of co-occurrence, encompassing a large virtual network of Twitter users. The list of tweets with the aforementioned hashtag and keywords also contain the hashtag “#RapeJihad”, which is also listed in the “Recently used hashtags/keywords” box, strongly indicating that this Islamophobic notion, used to denigrate Muslim men as sexual exploiters and (overwhelmingly) paedophiles, is still extremely relevant in UK society and public consciousness. Here many of the tweets, and re-tweets referred to recent news reports in the UK, involving South Asian and Kurdish males. According to Carter (2017), **such negative perceptions are fuelled by sensationalist media** reporting about child sexual exploitation (CSE) and rape, which in turn fuels the rhetoric of othering that reproduces Muslims as a racialised threat, with Muslim masculinity particularly perceived as dangerous. Indeed, it has been observed that the development of the British Muslim as a racialised threat is a current and on-going process, resulting in their marginalisation (Tufail, 2015), with Twitter being noted as becoming a soapbox for the masses in which popular discourses and disproportionate representations of Muslim are reproduced (Carter, 2017). This information can be utilised by NGO operators, researchers and academics to **discover new hashtags to be monitored** and to **acquire an overview of the main topics discussed in Islamophobic tweets**. As such, particular topics, mentioned persons and places as well as co-occurring hashtags can all be detected, enabling a fine-grained analysis.

On this point, many of the tweets contained a hashtag entitled **#pedobritain**, seemingly referring to the alleged high instances of paedophilic and child sexual exploitative activity in the UK, sanctioned by Islam and perpetrated by Muslims. On searching for this specific hashtag, and examining the generated “*Hashtag co-occurrence network*”, it was discovered that such sentiments from users were indeed the case. As evidenced below, numerous hashtags within the “*Hashtag co-occurrence network*” referred to grooming gangs linked to Muslims, particular customs and practices of the Islamic community (i.e. Sunnah), as well as locations where organised CSE cases were discovered, such as West Yorkshire, including Huddersfield. The generated keywords also demonstrate such narratives, with references to Huddersfield, rape jihad, teenagers and Great Britain (GB). As such, “*Recent Trends*” feature of the Platform enables the detection of newer Islamophobic hashtags, in this instance “**#pedobritain**”, which demonstrates the evolution of specific narratives, and their prevailing nature.

If a user selects one of the tweets on the right-hand side (see Figure 31), it is then possible to **open the message of interest inside Twitter** for further observation and investigation – to check for example the replies or the number of retweets. Each of these tweets displayed inside the Platform contains also the link to the user who originally posted it, and also that of the user who re-tweeted it. These messages are also linked to the counter-narratives tool, which can be activated by clicking on the speech bubble icon (see section 2.2.).

Figure 32 shows an example of the use of the “*Hashtag Trends*” function of the Platform, using the hashtag “Muslim Paedos”, with a time snapshot of October 2018. The feature provides general data, in real-time, reporting how many tweets, re-tweets – with a visual demonstration of the *Most retweeted messages* – and replies that tweets containing this hashtag has received, within various networks, in the chosen snapshot.

Figure 32 – “Hashtag Trends” Functionality, utilising “Muslim Paedos”, October 2018



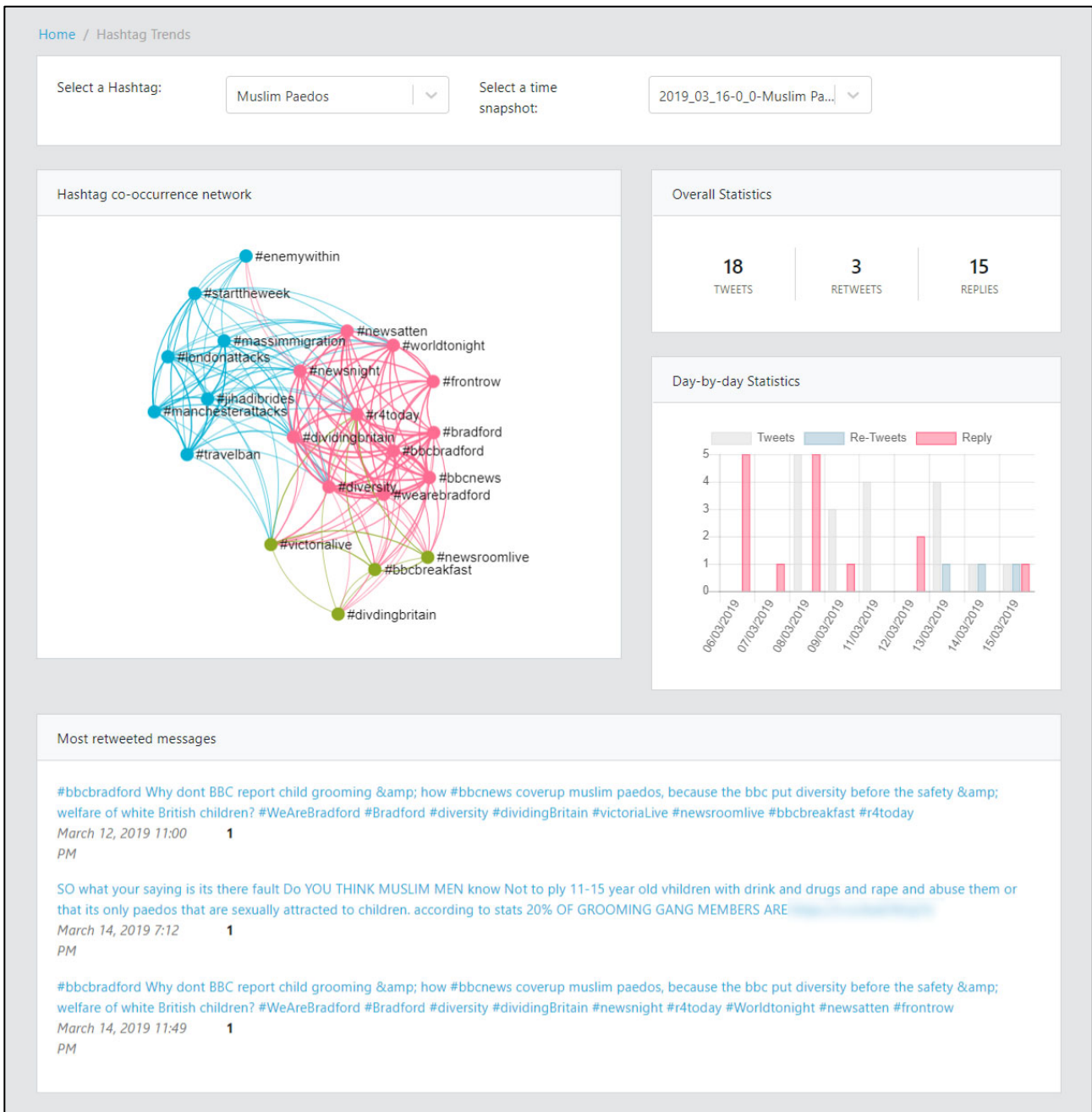
SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

The example of Figure 32 above visually displays a moderate-sized “Hashtag Co-Occurrence Network”, encompassing inter-linked Islamophobic sentiments via hashtags, including Muslims, Islamic culture, criminal activity of paedophilia and by extension, grooming gangs. The extent of the “Muslim Paedos” hashtag is displayed in terms of links with the other networks. The “Overall Statistics” box reveals a sizeable number of posted tweets, re-tweets and replies elicited concerning these narratives. What is interesting to note within the “Day-By-Day Statistics” box – which discerns how active (in presence) hashtags are over time on Twitter – is that on the 19th of October 2018, there was a large spike in re-tweets. On that specific day, several news outlets in the UK, both online sources including BBC News, the Guardian and the Telegraph as well as traditional print, television and radio media, reported that – as

reporting restrictions on the trials had partially been lifted - in Huddersfield, in West Yorkshire in the UK, a grooming gang comprised of over 20 men were convicted of sexual offences against girls. It was reported that the gang was the largest group ever convicted for sex abuse in the UK, in which the sexual offences included trafficking, drugging and raping 18 vulnerable girls, from the ages of 11 to 17. The 20 men were all of Asian – mainly Pakistani – origin, with their ethnic origin and apparently propensity for being part of “rape squads” as “Pakistani paedos” being remarked on in the “Most re-tweeted messages” box. The ringleader of the gang is a British Indian, and a member of the Sikh religion. His religion may be why within the “Hashtag Co-Occurrence Network” in “Muslim Paedos” hashtag, “#Sikh” is also mentioned and displayed, which can provide a nuanced analysis for NGO operators, academics and researchers. The spate of CSE cases in the UK in localities of Middlesbrough, Rochdale, Rotherham, Telford and Huddersfield (among others) has led to the media inciting moral panic over South Asian ‘grooming gangs’ preying on young, underage white girls, contributing to misconceptions that stereotype South Asian men as ‘natural’ perpetrators of these crimes due to culturally-specific notions of hegemonic masculinity (Gill and Harrison, 2015). On this issue, exploring anti-Muslim hate speech on Twitter, Awan (2014) discovered that there were a number of terms that were used to describe Muslims in a negative manner; with “Muslim Paedos” being used up to 30% of times, reflecting and coinciding with recent cases of Asian, and by extension Muslim men convicted of grooming offences against underage girls. Additionally, a number of accounts used and circulated anti-Muslim images and literature as a means to defame and caricature Muslims as dangerous paedophiles. As is highly evident in Figure 32, these notions are still prevalent and persistent, being widely disseminated and commented on, reflecting a recently revealed occurrence in the UK.

In Figure 33, using the same hashtag of “**Muslim Paedos**” and selecting a different time snapshot, in this instance, March 2019, demonstrates that within the “*Overall Statistics*” and “*Day-By-Day Statistic*”s boxes, there were **a higher number of tweets and replies than re-tweets**, signifying that much Islamophobia-related discussion has taken place. In the “*Hashtag Co-Occurrence Network*”, aside from the “Muslim Paedos” hashtag, it is compelling to observe that there are no explicitly Islamophobic hashtags. Instead, hashtags pertain to localities such as the UK city of Bradford, also in West Yorkshire, news outlets such as BBC News, BBC Breakfast and Newsnight, along with seemingly inoffensive terms such as “#diversity”, “#dividingbritain” and “#massimmigration”. However, a cursory glance at the *Most re-tweeted messages* box illustrates numerous messages replete with Islamophobic attitudes and the inclusion of some of the aforementioned hashtags, derogatorily linking Muslim men to paedophilia, rape and child grooming sex offences against young, white girls. Additionally, such tweets also accuse the BBC of deliberately not reporting and covering up alleged child grooming cases perpetrated by Muslim men, for the sake of placing precedence over ethnic diversity, rather than the welfare of indigenous, British children. This analysis is interesting, as users perceive the BBC to be an untrustworthy news source that is more concerned about diversity, whereas in the previous example of Figure 25, the Huddersfield grooming gang story was in fact reported on by BBC News.

Figure 33 – “Hashtag Trends” Functionality, utilising “Muslim Paedos”, March 2019

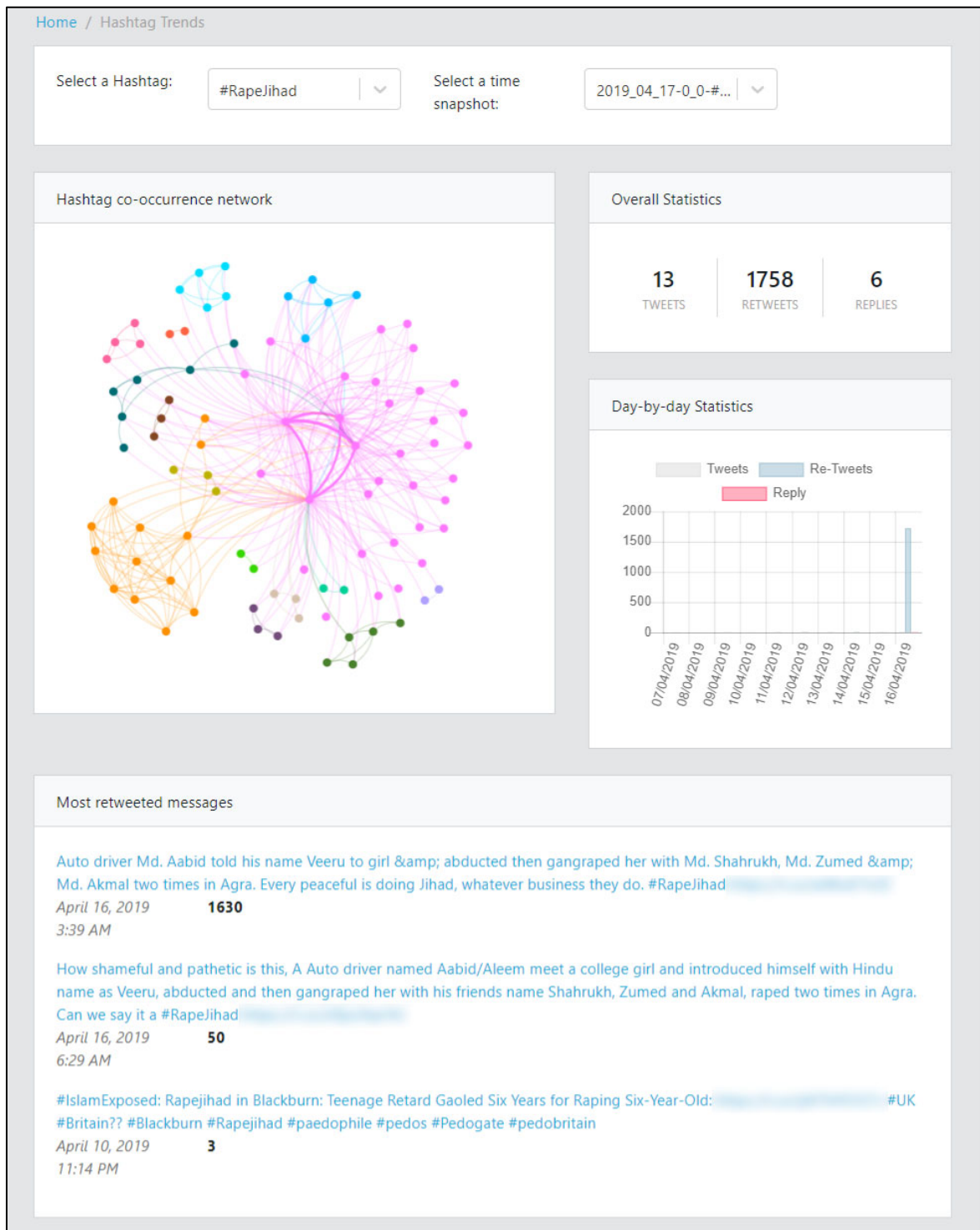


SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

It is also striking to observe the **use of innocuous terms** in the form of hashtags being utilised by users to engage in Islamophobic sentiments online, for example Bradford city and locality-centric hashtags of #bradford, #wearebradford, #bbcbradford. This may be in reference Bradford’s diverse demographics, in which 26% of the population is Asian, with the city possessing one of the highest percentages of South Asians in the country, and in regards to religion, 24.7% being adherents of Islam. On the other hand, it was reported in the UK on the 27th of February 2019, by BBC News and various online news sources that in Bradford, nine men, as part of a grooming gang, were convicted of raping and sexually abusing young, vulnerable girls, who were initially in the care system. The men were all of Asian – predominantly Pakistani - origin. It could be the case that such usage of these particular hashtags and accompanying tweets as well as re-tweets are in reference to this incident, which although it was reported on in the media 2 weeks

previously, may still be still high in the UK public’s mind and in the online fora. Figure 33 demonstrates that NGO operators, academics and researchers can observe a nuanced analysis of hashtags, in respect to hashtags that are not outwardly Islamophobic and/or are benign but are affixed to tweets that are awash with Islamophobic statements.

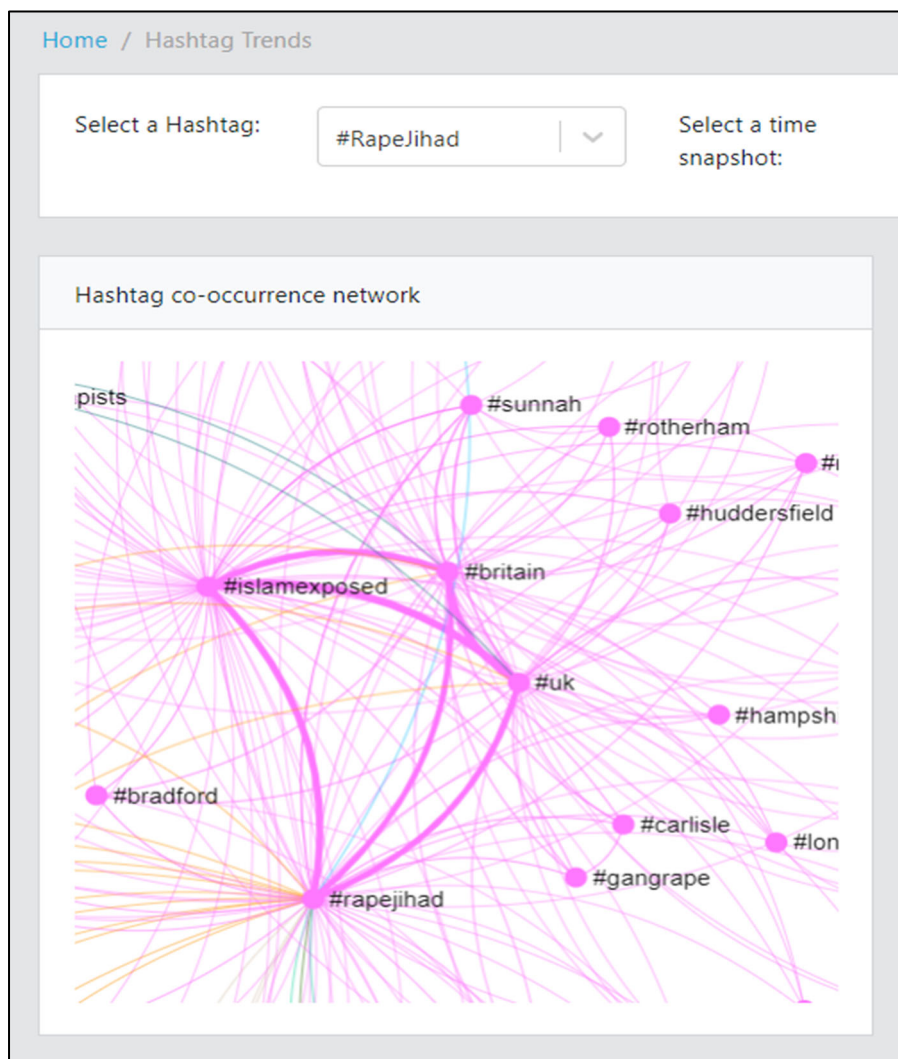
Figure 34 – “Hashtag Trends” Functionality, utilising “#RapeJihad”, April 2019



SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

On the 16th of April 2019, a high peak was recorded concerning the “#RapeJihad” hashtag. Figure 34 presents a visualisation of this hashtag, a day later on the 17th of April 2019. As is evident, the *Overall Statistics* and *Day-by-day Statistics* boxes demonstrate that of 13 tweets containing the “#RapeJihad” hashtag, there have been 1758 re-tweets, indicating active and robust dissemination activity. As Figure 35 below shows, in the “*Hashtag co-occurrence network*”, “#RapeJihad” is within a very broad network, comprised of interspersed and co-occurring Islamophobic hashtags that refer to grooming, gang rape, Muslims, and localities including Rotherham, Huddersfield, and Bradford. Some of the hashtags included in the “*Hashtag co-occurrence network*” included “#pedobritain”, “#pedogate” and “#pedos”, strongly suggesting that “#RapeJihad” has a high degree of crossover and interchange with this specific Islamophobic sentiment. It can also be observed that the #IslamExposed” hashtag has a strong link and by extension, network with “#RapeJihad”.

Figure 35 – “Hashtag Trends” Functionality, utilising “#RapeJihad”



SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

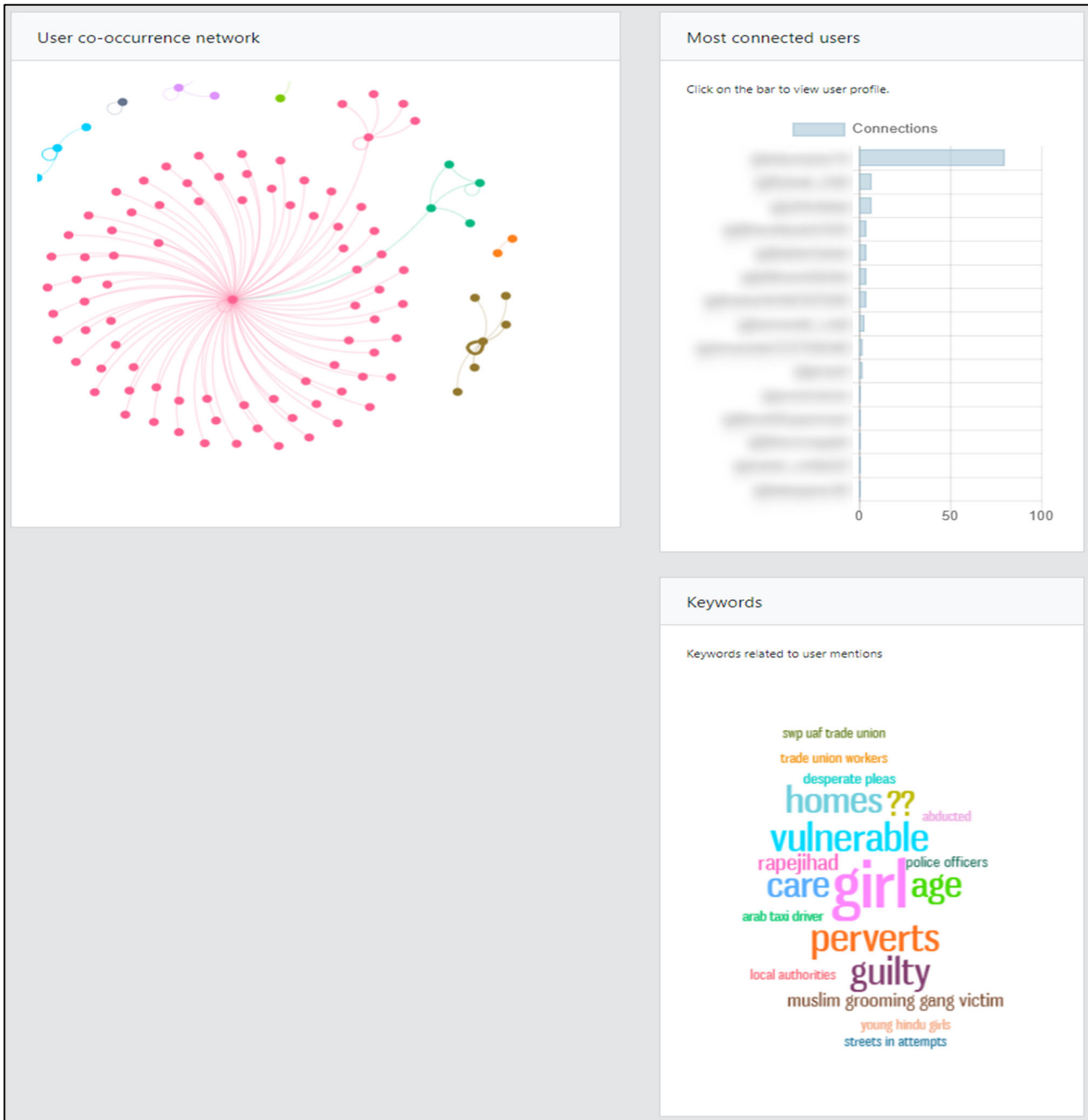
According to Carter (2017), the very overt lexicon of the #Rapejihad hashtag serves as an explicit call about the dangerous Muslim men who prey on Western women and girls, resulting in a striking example of what has been described as ‘digital Islamophobia’ (see Horsti, 2017). In this sense, the use and dissemination of #rapejihad in the online fora, is executed to supposedly reveal the ‘truth’ about Muslim

masculinity as dangerous through discussion of the paedophile rings in Rochdale and Rotherham, the auctioning of female sex slaves in Syria, and honour killings, etc., as well as positioning Muslim men as threatening (Carter, 2017). Such neologisms tap into and reinforce existing discourses that decry Muslim men's sexuality as uncontrollable and exaggerated (Evolvi, 2018), in which Muslim males, as an ethnic group, are deemed to be hyper-masculine, violently raping non-Muslim, unveiled women, resulting in a major issue that is believed to be unaddressed in the West due to feminism (Horsti, 2017). Kelsey (2017) points out that following the child abuse scandals in England, far-right groups, particularly the English Defence League (EDL) commonly used the term "rape jihad" to describe the behaviour of Muslim perpetrators towards children, young girls and women. Moreover, Kelsey (2017) argues that "rape jihad" functions as an intertextual term that dialogically functions through other inter-discursive connections, projecting the child abuse issue from within UK society to a foreign evil. In doing so, it internationalises via Islam, rather than humanity as a whole. Using "jihad", there are connotations of Islam and the war on terror as well as more specific signifiers such as rape being used as a weapon of war. Not only are stories of rape as a weapon familiar in reports of atrocities in foreign conflicts, but it also domesticates the acts of child abuse as an act of war against British victims.

It is striking to also note that in the "*Hashtag co-occurrence network*", the "**#RapeJihad**" hashtag possessed an international reach, where other countries, such as Pakistan, Sweden, Israel, Denmark, India and Finland were mentioned, with the latter country connected to a hashtag entitled "**#rapefugees**", a portmanteau of "rape" and "refugee" often used to connect Islam, as well as perceived migrant Muslim men with sexual violence, mainly by right-wing extremists (Würschinger *et al.*, 2016). It is salient to point out that in the *Most retweeted messages* box, the top tweet is referring to a criminal case of gang rape in India of a non-Muslim woman, by Muslim males. This highlights that the "**#RapeJihad**" hashtag is an Islamophobic narrative, that, whilst very relevant and potent in the UK, has also transcended international borders, and is ostensibly an issue of major concern in other nations. Indeed, it has been observed in India that false claims by the Hindu right of a "Love Jihad" organisation, which is forcing Hindu women to convert to Islam through false expressions of love, have gained much traction and incited moral panics (Gupta, 2009; Strohl, 2019). As such, **NGO operators, researchers and academics** are able to acquire a **fine-grained analysis of particular hashtags**, via being able to observe and discern which Islamophobic expressions are transnational in scope.

In the "**Hate Speakers**" function, the Hatemeter Platform can identify the most popular, active and connected anti-Muslim users, also known as influencers, which are displayed as a central node or nodes. After selecting a hashtag and choosing a time snapshot, the Hatemeter Platform exhibits the network of users that posted messages containing the given hashtag in the "**User Co-occurrence Network**" box, in which colours are automatically assigned by the network analysis algorithm to identify communities of users. As Figure 36 below demonstrates, the "**#RapeJihad**" hashtag is contained within a relatively large cluster of users in the "*User co-occurrence network*". Networks that are the same colour, usually communicate and interact with each other. Within the "**#RapeJihad**" hashtag, there is a high degree of connection between one main influencer and various users, as evidenced by the "*Most Connected Users*" frame, which showcases a ranked list of users, in descending order. Here, the top influencer has 80 connections, with others possessing much less, of around 4 to 7 connections, which signifies that this specific influencer, due to their high number of connections, is most likely to be posting and giving visibility to Islamophobic messages. The "**Keywords**" box provides a visualisation of the user mentions that contain the "**#RapeJihad**" hashtag, within the various networks, enabling NGO operators, academics and researchers to gain an idea of what Islamophobic narratives are linked. As is evident in Figure 36, numerous keywords refer to Muslims, grooming gangs and young girls, demonstrating the particular alleged Islamophobic sentiment of Muslims as rapists and sexual exploiters, across the networks.

Figure 36 - “Hate Speakers” Functionality, utilising “#RapeJihad”

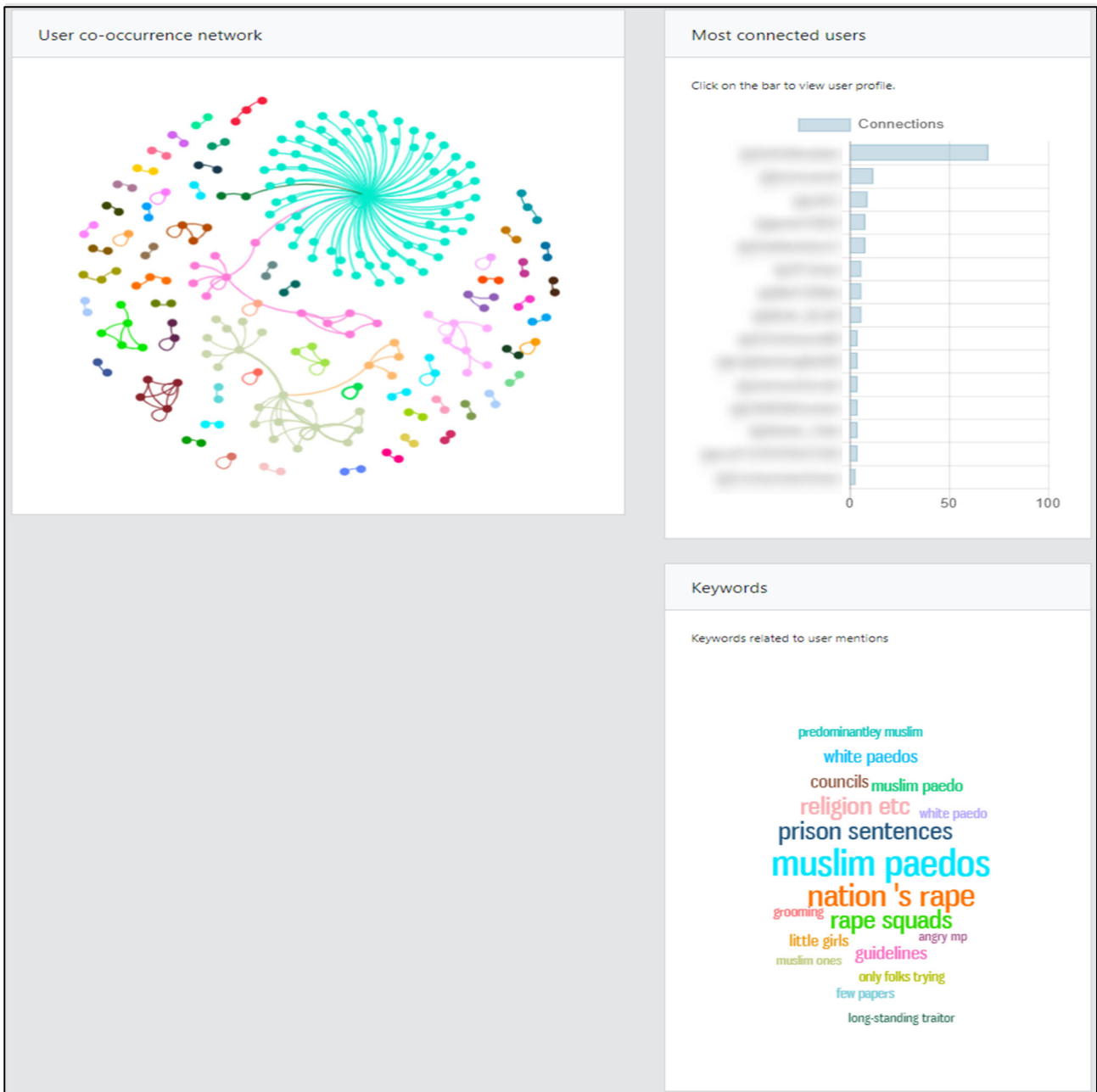


SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

In Figure 37, it can be observed that within the “Muslim Paedos” hashtag (October 2018), the blue cluster represents a very large extent of connection, again with one main influencer who possesses 70 connections and various other users, as evidenced by the “Most Connected Users” box. The thickness of the lines in this specific cluster exhibits that there is a large degree of traffic and interactions between the influencer and their connections within the “User co-occurrence network”, indicating that many tweets, re-tweets and replies have been posted and disseminated. Shifting away from the blue cluster, the visual analysis presents an interesting view of the various other networks and by extension, communities in which the “Muslim Paedos” hashtag is enclosed in. It is evident from the numerous colours in the “User co-occurrence network” that some networks have a link to other networks, and some

networks are smaller in size with less connections. The visual analysis suggests that there has been some discussion and dissemination of the “Muslim Paedos” hashtag within these various networks, albeit to a lesser extent than the blue cluster, which contains the majority of Islamophobic activity. The “Keywords” frame reveals that within these networks, the “Muslim Paedos” hashtag is heavily concomitant to purported Islamophobic notions of Muslims as paedophiles and rapists, committing offences against very young girls, and with mention of grooming.

Figure 37 - “Hate Speakers” Functionality, utilising “Muslim Paedos”

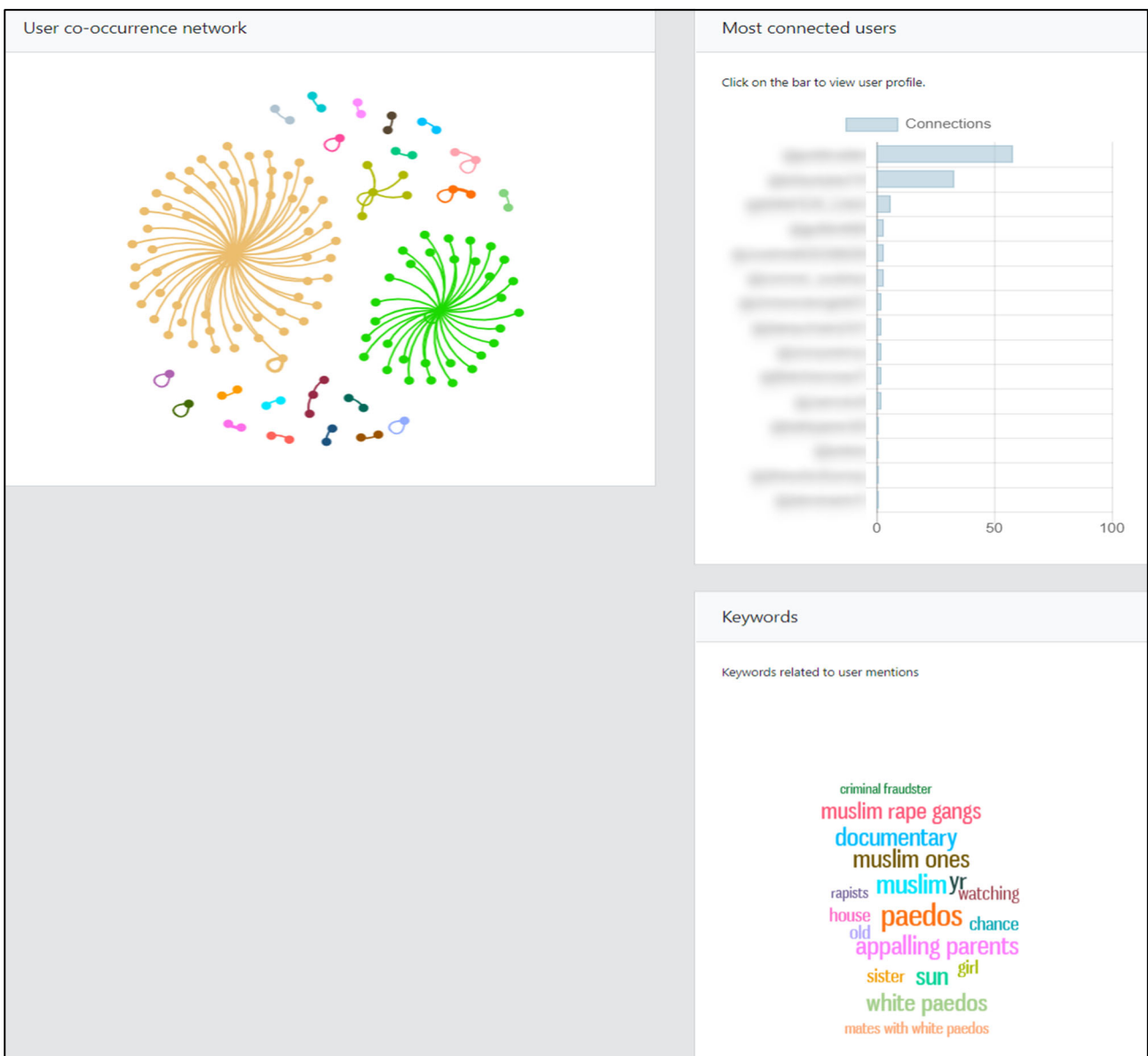


SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

Figure 38 shows the “Muslim Paedos” hashtag with a different time snapshot (May 2019), using the method of compare and contrast in order to analyse the usage of the hashtag in a different and more recent time period. The “User co-occurrence network” indicates that there are two main influencers, with

the **“Most Connected Users”** box exhibiting 58 and 33 respective connections, and two distinct clusters, as evidenced by the different colours. Within both networks, the thickness of the lines demonstrates that there is a high extent of traffic and interactions in the respective clusters, suggesting a strong concentration of the **“Muslim Paedos”** hashtag, well as active posting and diffusion of accompanying tweets, re-tweets and replies. The visual analysis reveals that beyond these two networks, the **“Muslim Paedos”** hashtag has less visibility, interactions and awareness, which can be observed by the networks that are much smaller in size in respect to user connections. The **“Keywords”** frame spotlights that within the various networks, the **“Muslim Paedos”** hashtag is still strongly linked to Islamophobic attitudes of supposed Muslim rape gangs and Muslims as paedophiles. As this time snapshot is relatively recent and is 7 months after the previous **“Muslim Paedos”** hashtag in Figure 37, the visual analysis suggests that this is an **enduring and contemporary Islamophobic narrative**, and an **Islamophobic discourse that has a wide reach** across time and space. As such, it is possible to evaluate and gain a nuanced analysis of hashtags, specifically focused on their trends over time.

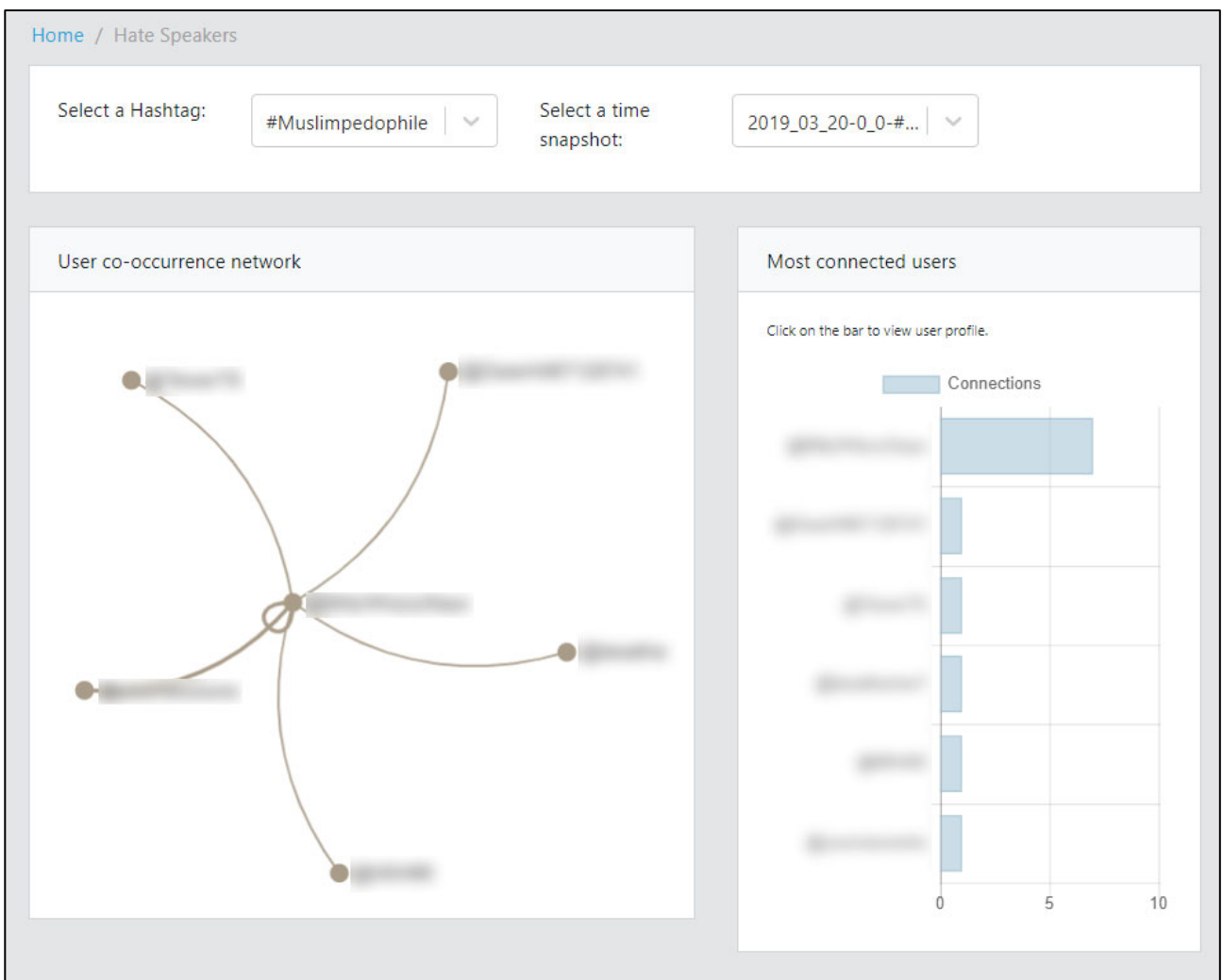
Figure 38 - “Hate Speakers” Functionality, utilising “Muslim Paedos”



SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

The **“Hate Speakers”** feature enables the observation of the nuanced structures of the **“User co-occurrence networks”**, and the types of communities of the networks. In Figure 39 below, it is strongly indicated that **the “User co-occurrence network” is an ‘echo chamber’**, where users predominantly interact with each other, in which beliefs are reinforced and amplified via communication and repetition within a closed system. This is suggested by the low number of just 6 users who are connected by the **“#Muslimpedophile”** hashtag. This suggests that beyond the 6 users, this hashtag does not have a wider reach nor is part of a network with larger links and connections. By doing so, the **“Hate Speakers”** function can provide fine-grained insights into online behaviour within virtual communities pertaining to Islamophobia.

Figure 39 - “Hate Speakers” Functionality, utilising “#Muslimpedophile”



SOURCE: Teesside University elaboration – Screenshot from the Hatemeter Platform

4. Suggestions and insights on the use of the Platform and of the Hatometer methodology for academics and researchers

While the **Hatometer Platform** has been implemented with specific objectives in mind, its **functionalities** are general-purpose and leave **ample room for future development and extensions** that can be implemented on demand.

Firstly, the main goal of the Platform is to support NGO's operators in tackling anti-Muslim hate speech online, by automatically monitoring and analysing Internet and social media data on the phenomenon. Nevertheless, the analyses presented above shed light on some interesting ideas for **possible employments** of the tool in **academia** and **in the areas of research**. Indeed, the analytical tools for a deeper exploration of Islamophobia, providing **descriptive statistics** concerning the amount and frequency of the hashtags identified as keywords indicating hate speech. Within the "**Alerts**" function of the Hatometer Platform, the ability to select and view data pertaining to particular hashtags, in the form of a date in which the Islamophobic hashtag has been highly present on Twitter, as well as the number of hashtags within this peak, can act as an important method of analysis, in various ways. NGOs, academics and researchers can engage in a **comparative investigation**, and observe within the peaks the trends of particular Islamophobic narratives – in the form of hashtags – and whether such accounts have increased or decreased over time (i.e. from a year ago, or six months ago, or two months ago to the most recent peak). In terms of low peaks and/or high peaks, NGOs, academics and researchers are able to acquire a sense of the public, online "strength of feeling" within certain Islamophobic hashtags. Interestingly, it is possible to investigate **pieces of public debate** (only on Twitter¹³ and YouTube comments) developing around a specific event: what are the most common users' reactions? Which are the most frequent hashtags? Do these hashtags indicate positive or negative attitudes toward the event?

Through the Platform, it is possible to have an idea of the **context** in which hate speech can develop (i.e. by looking at the entire conversation where a tweet, reply and/or retweet are included) and to understand meaningfully **narratives** and **patterns** around Islamophobia online. This possibility will support research and academic activities studying the phenomenon, providing empirical evidences on its development and main characteristics.

It is also possible to perform some **preliminary content and network analyses**, by exploring the words and expressions most frequently associated with specific hashtags. It can occur that the same hashtag is employed by both hate and non-hate groups, and it can be stimulating to understand how the meaning of a word or expression can change from time to time. The co-occurrence analysis of hashtags can also provide some very interesting indications on the political or ideological affiliation of the users, reflecting the connection with hot topics and voting attitudes.

The use of the "**Hate Speakers**" and "**Hashtag Trends**" features enables users to discern *why* there may be a peak in Islamophobic activity concerning a specific hashtag. Here, NGOs, academics and researchers can examine and cross-reference exact dates, which culminated in a high peak of the Islamophobic hashtag, as well as the dates a few days beforehand, leading up to the peak, which may be linked to events involving Muslims. In this way, NGO operators, academics and researchers can also view the content and trends of such related tweets. Such tweets may contain certain connected topics or refer to news stories, as well as *who* is disseminating such tweets, their connections to other users, and important statistics regarding the number of tweets, retweets and replies created, disseminated and discussed containing the Islamophobic hashtag.

¹³ The Hatometer database also comprehends comments from YouTube channels, but the manual is focussed on the use of Twitter.

Secondly, the focus of the Hatemeter Platform was directed to the **monitoring, analysing and tackling anti-Muslim hatred online**, but the ICT tool could also be directed to the investigation of **other hate phenomena** (e.g. homophobia, transphobia, online bullying, anti-Semitism and other racist speech). This new employment of the Platform could be easily realised by modifying the list of keywords and hashtag to be crawled on social media Platforms.

Moreover, the focus was on **three countries**, i.e. Italy, France and the UK, but the methodology could be replicated easily to investigate Islamophobia and/or other phenomena in **new languages**. Indeed, advanced functionalities are available depending on the NLP tools available for the given language.

Accordingly, to the expansion of the Platform to other phenomena and new languages, **the counter-narratives suggestions can also be enhanced**. From a **technical point of view**, new suggestions should be collected to create a specific pool of answers to be used by the tf-idf response retrieval model or for more advanced natural language generation algorithms (see 2.2.2). Secondly, this tool could be adapted to and implemented in different domains, such as for **education- or investigation-related purposes**. This could support also **media professionals** in their daily work to combat hate speech relating and related to published news on their websites or directly via **social media companies**.

The counter-narratives suggestions can also simply be directed to **activate awareness campaigns**. For example, when a minimum number of hashtags (e.g. two or three) indicating hate speech is employed in the same tweet or conversation (i.e. tweets and replies), information sensitising on the topic can be automatically published. Nevertheless, too much suppression on the social media Platforms could simply shift the problem to private messaging, where no intervention is possible.

Thirdly, and specifically under a technical point of view, **the ICT tool itself can be improved**. Future technical projects can **enlarge the dataset** of tweets and hashtags the Platform relies on, or realise a **lengthier exploration** of Islamophobia online and of other phenomena (e.g. over a period of years). It would also be possible to employ the Platform on **different types of social media** to verify if there is preferred channel to express hate speech and, if yes, why that specific version is better than the others.

Moreover, **more functionalities** can be added to the Platform. The addition of specific analytics related to the **users network structure** (e.g. centrality, in-degree and out-degree of the nodes) can help in understanding social network analysis, which explores how many users publish hate speech and if there are connections. This analysis could be useful also for **public security reasons**: if hateful communities proliferate on the Internet, could they also become more aggressive in the real world and commit concrete 'real-world' reprisals? Similarly, functionalities to **monitor specific hostile accounts** can be implemented. Obviously, for both these two possible developments, stakeholders should strictly comply with General Data Protection Regulation (GDPR) provisions.

References

- Aguilera-Carnerero, C. and Azeez, A.H. (2016). 'Islamonausea, not Islamophobia': The many faces of cyber hate speech. *Journal of Arab & Muslim Media Research*, 9(1), pp. 21-40.
- Alietti, A. and Padovan, D. (2018), 'Islamophobia in Italy: National Report 2017', in: E Bayrakli, F Hafez (eds.), *European Islamophobia Report*, Istanbul: SETA.
- Awan, I. (2014). 'Islamophobia and Twitter: A typology of online hate against Muslims on social media'. *Policy & Internet*, 6(2), pp.133-150.
- Baker, P., Gabrielatos, C. and McEnery, T. (2012). 'Sketching Muslims: A corpus driven analysis of representations around the word 'Muslim' in the British press 1998–2009'. *Applied linguistics*, 34(3), pp.255-278.
- Bayrakli, E. and Hafez, F. (2019a) *European Islamophobia Report 2018*, Istanbul, SETA.
- Bayrakli, E. and Hafez, F. (2019b) The State of Islamophobia in Europe, in Bayraklı E. and Hafez E., *European Islamophobia Report 2018*, Istanbul, SETA, 2019, pp. 7-58.
- BBC News (2018) 'Huddersfield grooming: Twenty guilty of campaign of rape and abuse'. [Online]. Available at: <https://www.bbc.co.uk/news/uk-england-45918845> [Accessed 18 February 2019].
- BBC News (2019) 'Bradford grooming: Nine jailed for abusing girls' [Online]. Available at: <https://www.bbc.co.uk/news/uk-england-leeds-47388060> [Accessed 10 October 2019].
- Bortone, R. and Cerquozzi, F. (2017), 'L'hate speech al tempo di Internet', *Aggiornamenti Sociali*, dicembre, pp. 818-27.
- Cediey, E. and Foroni, C. (ISM-CORUM) (2006) *Les Discriminations à raison de « l'origine » dans les embauches en France Une enquête nationale par tests de discrimination selon la méthode du BIT*. Genève, Bureau international du Travail. <http://ses.ens-lyon.fr/ses/manifichiers/testing2discriminationembauche.pdf>
- CCIF (2019), 'Our Manifesto'. <http://www.islamophobie.net/en/manifesto/>
- CNCDH (2017) *Rapport 2017 sur la lutte contre le racisme, l'antisémitisme et la xénophobie*. <http://www.cncdh.fr/fr/publications/rapport-2017-sur-la-lutte-contre-le-racisme-lantisemitisme-et-la-xenophobie>
- CNCDH (2018) *Rapport 2018 sur la lutte contre le racisme, l'antisémitisme et la xénophobie*. <https://www.cncdh.fr/fr/publications/rapport-2018-sur-la-lutte-contre-le-racisme-lantisemitisme-et-la-xenophobie>
- Ekman, M. (2015). 'Online Islamophobia and the politics of fear: manufacturing the green scare'. *Ethnic and Racial Studies*, 38(11), pp. 1986-2002.
- Evolvi, G. (2018). 'Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses'. *Religions*, 9(10), pp. 307-321.

- Faytre, L. (2019). 'Islamophobia in France: National Report 2018', in Bayraklı E. and Hafez E., *European Islamophobia Report 2018*, Istanbul, SETA, 2019, pp. 319-368.
- Froio, C. (2018). 'Race, Religion, or Culture? Framing Islam between Racism and Neo-Racism in the Online Network of the French Far Right'. *Perspectives on Politics*, 16 (3), pp 696–709.
DOI:10.1017/S1537592718001573.
- Giancalone, C. (2017), 'Islamophobia in Italy: National Report 2016', in: Bayraklı, Enes and Farid Hafez (eds), *European Islamophobia Report 2016*, Istanbul, SETA: 297–319.
- Giovannetti, M. and Minicucci, M. (2015), 'L'hate speech nei new social media: percezioni, esperienze, approcci, reazioni e risposte dei giovani utilizzatori e dei professionisti', Relazione al convegno Hate speech e libertà di espressione.
- Gouvernement.fr (2019), 'Bilan 2018 des actes racistes, antisémites, antimusulmans et antichrétiens', <https://www.gouvernement.fr/bilan-2018-des-actes-racistes-antisemites-antimusulmans-et-antichretiens>
- Gupta, C. (2009). 'Hindu women, Muslim men: Love Jihad and conversions'. *Economic and Political Weekly*, pp.13-15.
- Horsti, K. (2017). 'Digital Islamophobia: The Swedish woman as a figure of pure and dangerous whiteness'. *New Media and Society*, 19(9). Pp. 1440-1457.
- Kelsey, D. (2017). *Media and Affective Mythologies: Discourse, Archetypes and Ideology in Contemporary Politics*. Springer.
- Larsson, G. (2007). 'Cyber-islamophobia? The case of Wikislam'. *Contemporary Islam*, 1(1), pp. 53-67.
- Longhi, J. (2017). 'Tweets politiques: corrélation entre forme linguistique et information véhiculée'. #Info. Partager et commenter l'Info sur Twitter et Facebook, 2017. halshs-01841132.
- Malchiodi, M. (2016), *L'islam nei social media*, Pavia: Osservatorio di Pavia.
- Moore, K, Mason, P. and Lewis, J. (2008). *Media Images of Islam in the UK*. Cardiff: Cardiff School of Journalism, Media and Cultural Studies.
- Moretti, G., Sprugnoli R. and Tonelli S. (2015). Digging in the Dirt: Extracting Keyphrases from Texts with KD. In *Proceedings of the Second Italian Conference on Computational Linguistics*, Trento, Italy.
- Oboler, A. (2016). 'The normalisation of islamophobia through social media: Facebook'. In: Awan, I. (Ed.), *Islamophobia in Cyberspace: Hate Crimes Go Viral*. Routledge, New York, pp. 41-62.
- Osservatorio sulle Discriminazioni (2010). *Rapporto 2010*, Mantova: Osservatorio sulle Discriminazioni.
- Poole, E. (2006). 'The effects of September 11 and the war in Iraq on British newspaper coverage'. *Muslims and the news media*, pp.89-102.
- Richardson, J.E. (2004). *(Mis) representing Islam: The racism and rhetoric of British broadsheet newspapers* (Vol. 9). Amsterdam: John Benjamins Publishing.

- Runnymede Trust, (1997). *Islamophobia: A Challenge for Us All*, London: Runnymede Trust.
- Sian, K. (2018). 'Stupid Paki Loving Bitch: The Politics of Online Islamophobia and Misogyny'. In: Bhatia, M., Poynting, S. and Tufail, W. (Eds), (2018). *Media, Crime and Racism*, Palgrave Macmillan, pp. 117-138.
- Spärck Jones, K. (1972). "A Statistical Interpretation of Term Specificity and Its Application in Retrieval". *Journal of Documentation*. 28: 11–21.
- Strohl, D.J. (2019). 'Love jihad in India's moral imaginaries: religion, kinship, and citizenship in late liberalism'. *Contemporary South Asia*, 27(1), pp.27-39.
- Tahata, Y. (2018). 'Définir « les Français », une question de race? Analyse des formes de racisation dans les débats parlementaires sur la nationalité et l'immigration (1981-2012)'. *Mots, Les langages du politique*, 116. DOI : 10.4000/mots.23050.
- Tufail, W. (2015). 'Rotherham, Rochdale, and the racialised threat of the 'Muslim Grooming Gang''. *International Journal for Crime, Justice and Social Democracy*, 4(3), pp.30-43.
- Vox, Osservatorio Italiano sui Diritti (2019), *La mappa dell'intolleranza*, accessed online at: <http://www.voxdiritti.it/la-nuova-mappa-dellintolleranza-4/>.
- Würschinger, Q., Elahi, M.F., Zhekova, D. and Schmid, H.J. (2016). 'Using the Web and Social Media as Corpora for Monitoring the Spread of Neologisms. The case of 'rapefugee', 'rapeugee', and 'rapugee''. In *Proceedings of the 10th Web as Corpus Workshop*, pp. 35-43.

November 2019

HATE