Andrea Di Nicola
Daniela Andreatta
Elisa Martini

With:
Georgios A. Antonopoulos
Gabriele Baratto
Stefano Bonino
Serena Bressan
Shani Burke
Francesca Cesarotti
Parisa Diba
Jérôme Ferret
Marco Guerini
Bill Howe
Isis Koral
Mario Laurent
Elisa Martini
Giulia Pirozzi
Jen Schradie
Sara Tonelli

eCrime
ICT, law & criminology

06

eCRIME RESEARCH REPORTS

HATEMETER

HATE SPEECH TOOL FOR MONITORING, ANALYSING AND TACKLING ANTI-MUSLIM HATRED ONLINE

Beneficiaries

eCrime – Department 'Faculty' of Law, Department of Sociology and Social Research, University of Trento (Coordinator)
Fondazione Bruno Kessler (Co-beneficiary)
Universite Toulose 1 Capitole (Co-beneficiary)
Teesside University (Co-beneficiary)
Amnesty International Italy (Co-beneficiary)
Collectif contre l'Islamophobie en France (Co-beneficiary)
Stop Hate UK (Co-beneficiary)

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

eCrime Research Reports

No. 06

HATEMETER
Hate speech tool for monitoring, analysing and tackling Anti-Muslim hatred online

Andrea Di Nicola
Daniela Andreatta
Elisa Martini

With:
Georgios A. Antonopoulos
Gabriele Baratto
Stefano Bonino
Serena Bressan
Shani Burke
Francesca Cesarotti
Parisa Diba
Jérôme Ferret
Marco Guerini
Bill Howe
Isis Koral
Mario Laurent
Elisa Martini
Giulia Pirozzi
Jen Schradie
Sara Tonelli

Graphic design and layout by Damiano Salvetti

ISSN 2284-3302
ISBN 978-88-8443-886-7

eCrime – ICT, Law & Criminology
Department 'Faculty of Law'
University of Trento
Via G. Verdi 53, 38122 Trento, Italy
+39 0461 282336
www.ecrime.unitn.it

Trento, January 2020

Table of contents

Acknowledgements

Authors

Under the supervision of Andrea Di Nicola, Gabriele Baratto, and Elisa Martini, authorship is as follows:

Andreatta Daniela (2, 2.1.2, 4, 4.1, 5, 6), Georgios A. Antonopoulos (2.3, 4.3), Gabriele Baratto (4.1), Stefano Bonino (2.1.1), Serena Bressan (3), Shani Burke (2.3, 4.3), Francesca Cesarotti (4.1), Parisa Diba (2.3, 4.3), Andrea Di Nicola (1), Jérôme Ferret (2.2, 4.2), Marco Guerini (3.2, 5.2.1), Bill Howe (3.3), Isis Koral (3.2), Mario Laurent (2.2, 4.2), Elisa Martini (2.1.3, 4.1, 5.1, 5.2, 5.2.2), Giulia Pirozzi (4.1), Jen Schradie (2.2), Sara Tonelli (3.1, 5.2.3).

This report presents the results of project “Hatemeter - Hate speech tool for monitoring, analysing and tackling anti-Muslim hatred online” (hereinafter also referred to as “Hatemeter”, project reference: 764583). The project has been coordinated by the University of Trento (Italy) and especially by eCrime, the research group on ICT, law and criminology of the Department ‘Faculty of Law’, with the cooperation of the Department of Sociology and Social Research, in partnership with Fondazione Bruno Kessler (Italy), University Toulouse 1 Capitole (France), Teesside University (United Kingdom), Amnesty International – Sezione Italiana (Italy), StophHate UK (United Kingdom), and Collectif Contre l’Islamophobie en France (France). The project was funded by the European Commission - Directorate-General Justice and Consumers under the Rights, Equality and Citizenship Programme (2014-2020) and lasted 24 months: from February 2018 to January 2020.

The research project was coordinated by Andrea Di Nicola, associate professor of criminology at the Department ‘Faculty of Law’ of the University of Trento and scientific coordinator of eCrime, Department ‘Faculty of Law’, University of Trento. The eCrime team included Daniela Andreatta, Gabriele Baratto, and Stefano Bonino. The Department of Sociology and Social Research (University of Trento) team included Mario Diani, head of the Department, and Elisa Martini. University of Trento would like to especially thank the Video Director Matteo Scotton for the realisation of video of the Hatemeter awareness raising campaign, and the students Giulia Zantonello and Mohamad Baalbaki for the translations of the subtitles of the video.

The Fondazione Bruno Kessler (FBK) team included Sara Tonelli, Stefano Menini, Alessio Palmero Aprosio, Rachele Sprugnoli and Elisa Leonardelli from the Digital Humanities research group; Marco Guerini and Serra Sinem Tekiroglu from the Natural Language Processing research group; and Serena Bressan from the Project Management Group. Sara Tonelli was the technical coordinator of Hatemeter. The FBK team would like to express its gratitude to Giovanni Moretti and the students Mohamad Baalbaki and Y-Ling Chung, who provided unevaluable support throughout the Platform development.

University Toulouse 1 Capitole (UT1-C) team included Jérôme Ferret, professor of sociology, Mario Laurent, and for the first part of the project, Jen Schradie (currently professor assistant at the Observatoire du Changement of Sciences Po, Paris).

Teesside University (TU) team included Georgios A. Antonopoulos, professor of criminology, Parisa Diba and, during the final stage of the project, Shani Burke. Teesside University team would like to thank Nigel Copsey, Shahda Khan and David Temple as well as the participants to the study for their time and valuable insights.

Amnesty International – Sezione Italiana team included Francesca Cesarotti, Mariangela Pantaleo, and Giulia Pirozzi, as well as all the volunteers of the Italian Hate Speech Task Force.

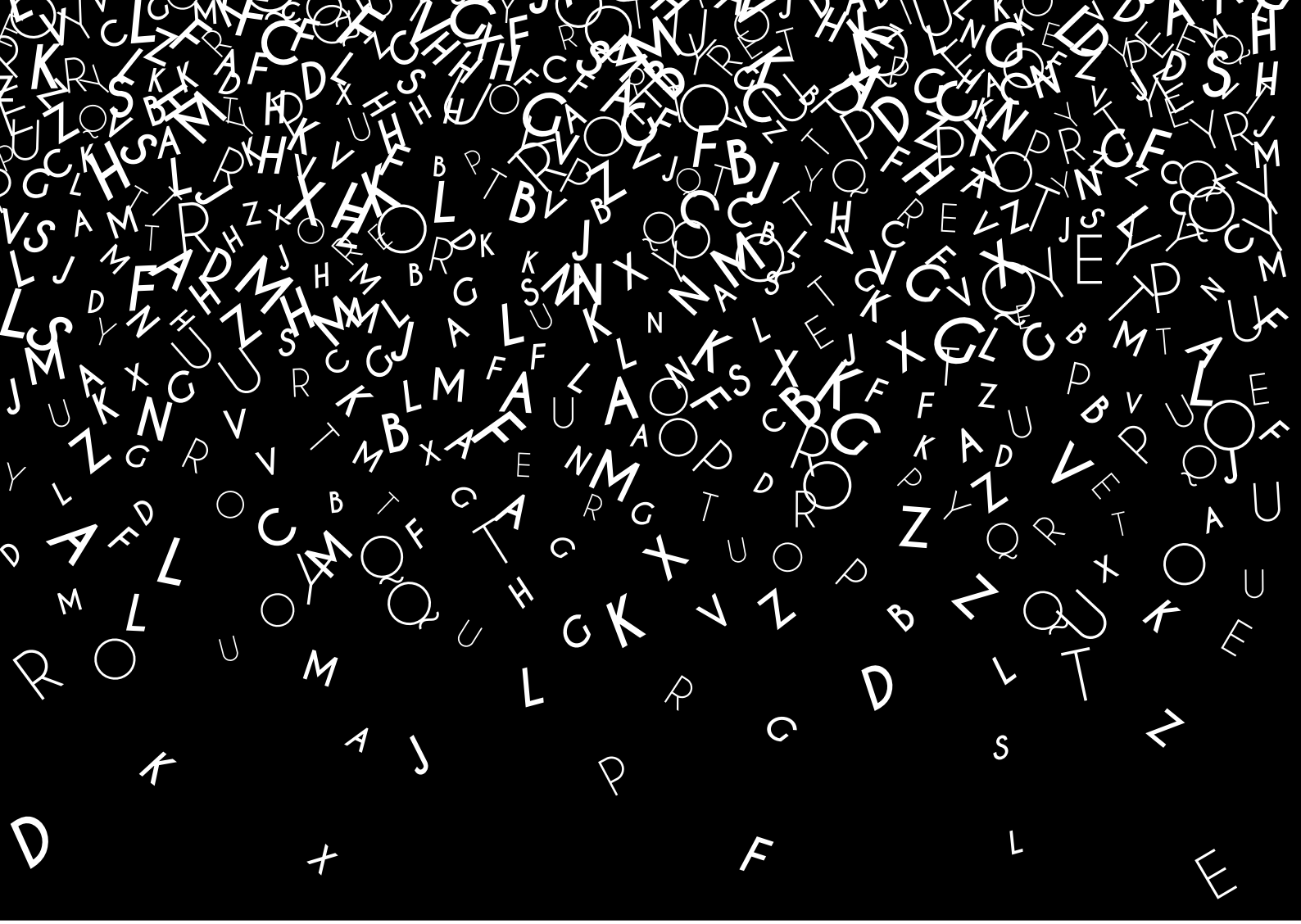
StophHate UK team included Rose Simpkins, Kylie Read, Bill Howe, Joseph Sankar, Rebecca Parker, Erika Samenaite, and Catherine El Ashkar.

Collectif Contre l’Islamophobie en France (CCIF) team included Isis Koral, Chafika Attalai, and all the voluteers.

We would like to acknowledge the 22nd edition of the festival “Voci per la Libertà – Una canzone per Amnesty” held in Rosolina Mare (Rovigo, Italy) from the 18th to the 21st of July 2019, which host Hatemeter project as partner of the event. Mostly, sincere thanks go to Roy Paci, Hatemeter Testimonial and Amnesty International Italy Award 2019, who as a musician and artist is committed in preventing and combating any adverse human rights impacts.

We would like to address a special thanks to all the academics, researchers and professionals members of the Hatemeter Lab, who participated in the seminars held in Toulouse on the 17th and 18th of December 2019.

This research work would not have been feasible without the help and assistance of a number of persons. Especially, we would like to express our gratitude to the administrative staff of the University of Trento and of all the Hatemeter partners, together with all the many other persons impossible to list here without whom this research would not have been accomplished. Last but not least, we would like to thank the Directorate-General Justice and Consumers of the European Commission and the Hatemeter Project Officers which, under the Rights, Equality and Citizenship Programme, have supported both the project and the realisation of this publication.



01

The project

Project “Hatemeter - Hate speech tool for monitoring, analysing and tackling anti-Muslim hatred online” aims at systematising, augmenting and sharing knowledge of 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, and 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 effective 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 by three NGOs in 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 Hatemeter Project 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 the project results on future research streams and operational strategies, the project favoured capacity building and training and the sustainability and transferability of the Hatemeter Platform among other target stakeholder groups (e.g., law enforcement agencies\agents, 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).

This document is the Final Report of Hatemeter Project and presents the results, standards and guidelines to favour the transferability of the Hatemeter Platform among other non-governmentalNGO/civil society organisations (NGO/CSOs) and stakeholders across the EU, and for other forms of hate speech and intolerance, as to augment the medium/long-term impact and sustainability of the project. The Hatemeter Partnership considers sustainability and transferability of its results as crucial elements of its activities towards a successful project implementation.

The first chapter gives an outline of the project and its aims.

The second chapter presents the background of the Hatemeter Project according to a subdivision in the three countries involved in the research, namely Italy, France and United Kingdom. For each country, the first subsection contains a brief overview of the current state of the art on Islamophobia online. The second subsection summarises the existing legislation relating to hate speech and hate crimes. The third subsection, reports a summary of the main points arising from interviews with stakeholders, to explore their current background experiences, needs and aspirations in the context of combatting Islamophobia online.

The third chapter is a technical description of the development of the Hatemeter Platform. It is divided in two subsections, the first one explaining method and data that have been necessary for the realization of the Platform, and the second one illustrating in detail the Platform and its various functions.

The fourth chapter presents evidence of online Islamophobia in Italy, France and the United Kingdom. Here, for each country, is an analysis from the three research teams involved in the project, and secondly the results of use of the Platform by operators from the three NGOs.

The fifth chapter describes the main results from the evaluation of the Hatemeter Platform. Firstly, it illustrates the validation strategy and the evaluation methodology underlying the assessment, and secondly demonstrates the results of some of the KPIs aimed at measuring the efficacy and efficiency of the Platform and an analysis of the scale of the User Experience Questionnaire (UEQ scales).

The last chapter concludes the report, providing some insights into potential use of the Platform by academics, researchers and NGOs outside the project.

02

Background of Hatemeter project

This chapter presents the background of the Hatemeter Project according to a subdivision in the three countries involved in the research, namely Italy, France and United Kingdom. The subsection for each country contains a brief overview of the current state of Islamophobia online and of the legal framework concerning hate speech and hate crimes. This is followed by a summary of the main points that emerged during interviews with stakeholders is reported, to explore their current background experiences, needs and aspirations combatting Islamophobia online.

2.1. Italy

2.1.1. What is known about Islamophobia online

The idea of public discourse has changed with the birth of the Internet and social media. Thanks to research conducted in the United Kingdom (Awan, 2014), it is well-known that cyber-hate is perpetrated by a variety of offenders, such as: the *trawler*, a person who goes through people's Twitter accounts to specifically target Muslims; the *apprentice*, who is new to Twitter but targets Muslims with the help of experienced online abusers; the *disseminator*, who tweets messages and pictures of online hate that target Muslims; the *impersonator*, who uses a fake account, profile and image to target Muslims; the *accessory*, who joins others' conversations on Twitter to target vulnerable Muslims; the *reactive*, who initiates an online hate campaign after a major incident (for example, a terrorist attack) or immigration-related issues and target that particular group and individuals (in this case, Muslims); the *mover*, who changes his/her Twitter account to target Muslims from a different profile; the *professional*, who has many followers on Twitter and launches major online hate campaigns against Muslims.

Italy is a "new immigration" country, which experienced the arrival of a larger Muslim population later than many countries in Western Europe. Yet, "*due to the spreading of information about controversial issues in other countries, the controversies surrounding the position of Muslims in society took shorter time to emerge in Italy than in European countries with 'older' Muslim populations*' (Serdar, 2015: 283). Importantly, *'Italy is also a destination and transit country for migrants, and its role in the so called refugees' crisis as well as the challenges that the country is facing in managing this situation appear on the public debate almost every day"* (eMore, 2017: 27). The debate concerning religious difference was absent from the foreground of public discourse until recently, when Islam became both a key theme of public debates on immigration and a negative issue in the media and political worlds (Sciortino, 2002). Nowadays, hate speech is on the rise and, with it, xenophobia, Islamophobia and anti-Semitic and racist messages, particularly since 2016 and as a result of both the humanitarian crisis that has led immigrants to European shores and terrorist attacks (Bortone and Cerquozzi, 2017).

Research conducted by the Commission on British Muslims and Islamophobia (2004) demonstrates that Islamophobia, at least in the sense of discrimination against Muslims *qua* Muslims, leads to the social exclusion of Muslim communities. On a scale ranging from 'very high' to 'low', Italy scores 'low' on governmental restrictions on religion but 'high' on social hostilities involving religion,[1] while most European countries score either 'moderate' or 'low.' Although it might be assumed that this is an issue related to a post-9/11 world, pre-9/11 studies on prejudice demonstrate that some European countries already suffered from discrimination against their Muslim and immigrant populations in the last century (Strabac and Listhaug, 2008), as part of a global Islamophobia trend that takes a transnational character (Morgan and Poynting, 2012). In comparative terms, people living in Spain (52 %), Germany (69 %), the United Kingdom (72 %) and France (76 %) hold more positive views of Muslims than people living in Italy (31 %) (Pew Global Attitudes

[1] The other European countries, which score 'high' on social hostilities towards religion, are Bulgaria, France, Germany, Greece, Romania, Sweden and the United Kingdom. See: Henne (2015).

Project, 2015). Indeed, in a study exploring prejudice against Muslims in sixteen Western European countries, Italy scores as the sixth most prejudiced country. The findings of this study suggest that people who live in countries with an official religion or a liberal citizenship regime, as well as intergroup contact among their population, are more likely to tolerate Muslims (Serdar, 2015).

The political and cultural debate in Italy is also increasingly coloured by xenophobic and racist contents due to the economic crisis and immigration-related issues. For the past twenty years, there has been a circularity and reciprocal influence among the political/institutional, media and social spheres. The target groups for online hatred in Italy tend to be migrants in general: hatred is associated with ideas connected to the economic and social unsustainability of immigration policies. Islamophobia becomes particularly noticeable insofar as migrants are popularly linked with Muslims and Muslims are then linked with terrorists.

As in other countries, in Italy there is an increasing connection among 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: social networks have a better capacity to convey messages, while open Platforms make individuals responsible for the content of their messages. Right-wing groups tend to be numerous, small and very fragmented: social networks help such small organisations to rise and to gain influence and visibility well beyond their limited resources, including national and international networking (Giovannetti and Minicucci, 2015). Indeed, social networks facilitate transversal and 'from the bottom' communication, make individual actions and behaviours immediately public, promote a sense of anonymity and the possibility to remain unpunished for one's hateful content and help messages, which would not normally appear in real life, be conveyed online (Giovannetti and Minicucci, 2015; Scaramella, 2016).

Compared to other European countries, particularly the United Kingdom and France, where the sustained migration of Muslims dates far back in the past, Italy is a relatively understudied locus for the life experiences of Muslims. Similarly, the issue of Islamophobia and anti-Muslim hatred, including their online mutations,[2] started to receive attention from the Italian academic community and civil society only in the last few years (see for example Pugiotto, 2012; Giovannetti and Minicucci, 2015; Gometz, 2017; Alietti and Padovan 2010, 2018). Nevertheless, there are several Italian projects that seek to monitor these phenomena and/or to develop educational tools in the fight against hate speech. Among them: the Intolerance Map,[3] Hate Barometer,[4] project 'REACT: Respect and Equality: Acting and Communicating Together,[5] "*Contro l'Odio*",[6] project 'PRISM: Preventing, Inhibiting and Redressing Hate Speech in New Media',[7] project 'BRICkS: Building Respect on the Internet by Combating Hate Speech',[8] project 'Media against Hate',[9] project 'Words are Stones',[10] and project 'Silence Hate'.[11]

According to Ljujic *et al.* (2016), Italy is the most 'anti-Muslim' European country after Hungary and intolerance of Muslims is higher in Northern areas of Italy and in and around Naples. 61% of Italians consider migration from Islamic countries to be a menace to the West; 70% of (right-wing) Italians have a negative perception of Muslims; 21% would not want to have Muslims as neighbours and 43% would not want to have a Muslim as a family member. On social media, particularly on Twitter, Italians communicate anti-Islam messages (72.3 %) more frequently than 'pro-Islam' ones (4.1 %); however only a very tiny minority (1.4 %) can be classed as hate speech or dangerous speech. 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). Italians tend to associate Muslims with the following words online: '*terrorista*', '*jihadista*', '*beduino*', '*abdullah*', '*tagliagole*', '*vu cumprà*' and '*marocchino*' (Vox, 2018).[12]

[2] The problems of drawing a line between online and offline Islamophobia is also an issue, as it is – for example – when making distinction between offline versus online violent radicalisation. See: Szmania and Fincher (2017).

[3] See http://www.voxdiritti.it/la-nuova-mappa-dellintolleranza-4/

[4] See https://www.amnesty.it/barometro-odio/ and https://www.amnesty.it/cosa-facciamo/elezioni-europee/

[5] See http://www.reactnohate.eu/the-rise-of-online-hate-speech-as-a-new-alarming-global-social-phenomenon-reasons-and-remedies/

[6] See https://controlodio.it/

[7] See https://www.rota.org.uk/content/prism

[8] See https://www.bricks-project.eu/

[9] See https://europeanjournalists.org/mediaagainsthate/

[10] See https://adice.asso.fr/en/projects/words-are-stones/

[11] See http://www.silencehate.eu/

[12] English translation: 'terrorist', 'jihadist', 'Bedouin', 'abdullah', 'cut-throats', 'street pedlar', and 'Moroccan'.

A recent study on Islamophobia in Italy of Alietti and Padovan (2018) recorded several discriminatory articles in newspapers and an important increase of instances of hate speech against Islam by Internet-based neo-fascist and Catholic fundamentalist groups. Muslims are the fourth most targeted group on Twitter (Alietti and Padovan, 2010), targeted on social media platforms together with Jews, migrants, homosexuals, women and disabled people (Vox, 2018).

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). Political statements that are discriminatory and racist towards Muslims, such as unwarranted generalisations, hate and incitement to violence and limitations of religious freedoms, are telling examples. The notion that adherence to Islam presents a threat to national security has also entered public and political discourse, particularly regarding notions that Muslims, or symbols of Muslimness (for example, the *burqa*), are associated with potential security issues. The mass media have an important role in shaping representations of reality. In particular, the Internet and social networks amplify this role, providing the media with quicker and easier ways to maintain the lead on mainstream information. The Internet offers immediacy, pervasiveness, amplification, replicability, social validation and persistence of certain messages. Social networks offer a polycentric proliferation of hate speeches and promote the diffusion of demagogic and propagandistic 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).

There is also the risk that current representations of Islam and Muslims do not grasp the complexities of the Muslim population in Italy and the dynamism of the Muslim world. These issues mean that the priorities to be dealt with when opening spaces for dialogue with Muslim communities have become areas of *cultural urgency* across right-wing and left-wing governments. At the same time, there is no agreement between the Italian State and the Muslim communities living in the country. There are two main reasons. First, Islam is pluralistic and therefore lacks a unifying voice due to being a religion shaped by dynamic processes among places, contexts, collective actors and state dimensions. Muslim organisations that the Italian State accepts as interlocutors do not represent the whole Muslim community; on the contrary, they simply represent the most visible and powerful organisations that have managed to negotiate their position in society with the Italian State. Second, the Italian State has recognised Muslim communities mostly in terms of religious difference. Cultural differences are conceived as more problematic. In addition, polls have shown that over half of the Italian population considers the migration and integration of Muslims in Italy to be more troublesome than the migration and integration of people from non-Muslim countries. Muslims are seen as carriers of social, cultural and religious differences that are more visible than others due to their 'distance' from the rest of the population. This distance is perceived to be mostly emerging from Muslim communities than from governmental and socio-political stances (Russospena, 2009).

2.1.2. Legislative Framework

A legal obligation to combat hate crime and hate speech makes it mandatory for governments worldwide to have in place and develop appropriate legislation (eMore, 2017). Hate speech in Italy is criminalised on the grounds of race, ethnicity, nationality or religion (Article 19, 2018).

The first legislative source to mention to hinder antidiscrimination is the Italian Constitution. Article 2 specifies that the Republic recognises and guarantees the inviolable rights of man, as an individual and in the social groups where he expresses his personality, and demands the fulfilment of the mandatory duties of political, economic and social solidarity. Article 3 then sets forth the principle of non-discrimination and prohibits distinctions related to sex, race, language, religion, political opinions, and personal and social conditions.

OSCAD is the Observatory for Security Against Acts of Discrimination (*Osservatorio per la Sicurezza contro gli Atti Discriminatori*) and provides a list of the fundamental pieces of legislation regarding antidiscrimination (OSCAD, 2020). Among them, the most relevant laws potentially pertinent with discrimination against Islam are:[13]

- Law 645/1952– Implementing rules of the XII transitional and final provision (first comma) of the Constitution against fascist activities (known as "*Legge Scelba*");
- Law 654/1975 – Known as "*Legge Reale*", it implements the International Convention on the Elimi-

[13] Other pieces of legislation regarding antidiscrimination, are: Law 962/1967 (Prevention and repression of crime of genocide); Law 104/1992 (Framework law for assistance, social integration and rights of disable persons).

nation of All Forms of Racial Discrimination. It has been modified by Law 115/2016, which adds one comma to article 3 regarding counter-activities and repression of the crime of genocide, crimes against humanity and war crimes, as defined by articles 6, 7 and 8 of the statue of the International Penal Court;

- Legislative Decree 122/1993, converted with Law 205/1993 – Known as "*Legge Mancino*", it contains urgent measures against racial, ethnic and religious discrimination;
- Law 482/1999, modified by Law 38/2001 – It extends the dispositions of the *Legge Reale* (article 3) and of the *Legge Mancino* to prevent and repress intolerance and violence phenomena to individuals that are part of linguistic minorities.

More specifically, *Legge Reale*, modified by Law 205/1993 (*Legge Mancino*) and Law no. 85/2006 (Law on thought crimes), "punishes those who propagandize ideas founded on racial or ethnic superiority or hate, or instigate someone to commit, or themselves commit, acts of discrimination for reasons of race, ethnicity, nationality or religion; those who, in every way, instigate someone to commit, or themselves commit, violence or acts which induce to violence for reasons of race, ethnicity, nationality or religion; those who take part or support organizations, associations, movements or groups which are aimed at subverting the socio-economic order" (eMore, 2017: 41).

The Italian Penal Code, then, has provisions regarding offences against a religious creed through contempt for people (article 403) and through vandalism (article 404), and desecration of tombs (article 408). Legislative Decree 21/2018 introduces in the Penal Code articles 604bis "Propaganda and incitement for reasons of racial discrimination, ethnical and religious", and 604ter "Aggravating circumstances".

From an international perspective, in Italy several pieces of regulation are in force (OSCAD, 2020):

- Universal Declaration of Human Rights, approved and proclaimed by the General Assembly of United Nations (1948);
- European Convention for the Protection of Human Rights and Fundamental Freedoms (1950);
- Additional Protocol no 12 to the Convention mentioned above (2000);
- New York Convention on the Elimination of All Forms of Racial Discrimination (1966);
- Charter of Fundamental Rights of the European Union (2000) that, according to the Lisboan Treaty (2007) has the same juridical value of treaties and it is binding for the institutions of the Member States.
- With Legislative Decrees no 215 and 216 (2003), Italy also ratified the EU Directive 2000/43/EC on Racial Equality and the so-called "occupational" Directive 2000/78/EC (eMore, 2017).

In Italian jurisprudence, there have been a few cases of ethnic and racial discrimination in which courts have recognised online associations as "criminal conspiracy aimed at incitament to hatred, racist propaganda and violence against ethnic or religious minorities" (eMore, 2017: 41), even though there were no physical interactions. Injuries aggravated by discriminatory reasons, incitement to violence and aggravated defamation both online and in person have also been recorded. At the same time, even when law enforcement agencies are inactive in responding to hatred and discrimination, independent authorities and civil society have sometimes taken the lead.

Constitutionalist Andrea Pugiotto (2012) suggests that prevention should be utilised to fight homophobic, xenophobic, anti-Semitic, Islamophobic, etc. hate speech. For example, he proposes the introduction of positive actions of formation and information rather than penal repression, and the employment of diversionary penalties such as community work rather than stronger sanctions. He also suggests that new crimes of expression would clash with Article 33, first 'comma' of the Italian constitution, which gives special protection to freedoms of historical and scientific research.[14] However, on a balance of rights, the right to freedom of expression can be circumscribed if and when other rights, goods and interests of a constitutional nature need to be preserved. In fact, there are intrinsic limitations to the right to freedom of expression in terms of an ideal public order, human dignity, principle of equality, international obligations and those instances when ideas become actions (Pugiotto, 2012). Lastly, messages and actions that are of a racist, xenophobic, homophobic and sexist nature tend to be stigmatised but there is no agreement, both on why they should be stigmatised and why they should be combated through criminal law (Gometz, 2017).

[14] By way of example, this protection has allowed: politician and academic Rocco Buttiglione to hold a conference on the 'immoral and unnatural dimension of homosexuality; various historians to give characterised as 'holocaust denial'; and the publication of a sociological study that theorises the superiority of heterosexuality (Pugiotto 2012).

2.1.3. Main results of in-depth interviews

In order to scope the online problem of Islamophobia and assess background experiences, needs and aspirations of NGO/CSOs and other target stakeholders groups, in-depth interviews have been organised. In the Italian context, six persons participated in the interviews, experts in different fields related to anti-Muslim discrimination and hate speech. Their identity is anonymous for confidentiality reasons, but their affiliations are as follows: University of Reading; Municipality of Turin; Associazione Carta di Roma; the Observatory for Security Against Acts of Discrimination; Unione delle Comunità e Organizzazioni Islamiche (UCOII) in Italy; and COSPE Onlus. Below, it is reported a summary of the main points emerged during the interviews.

- In Italy some issues emerge when trying to catalogue data on hate crimes and hate speech incidents. Official data collected by law enforcement agencies (LEAs) are held within a database, the National System of Criminal Data Collection (SDI, Sistema Di Indagine). It is a national database and contains data based upon violations of criminal law. However, it os not possible to disaggregate this data according to religious faith and for this reason it is very difficult to retrieve statistics on Islamophobic hate crimes and speech. Unlike the UK, where police forces have a memorandum with an NGO called Tell MAMA that measures anti-Muslim attacks, in Italy there is no close cooperation between the police and NGOs and this situation limits the possibility of an accurate representation of the phenomenon.
- Anti-Muslim sentiments intersect with xenophobic sentiments, such as anti-migrant ones and this is also evidenced by the latest report from Associazione Carta di Roma. Moreover, there tends to be an association between migrant boats and terrorism. Other associations are between migrants and crime, and migrants and diseases. This speaks of the multi-factorial nature of discrimination.
- Nowadays, Roma people, refugees and Muslims are the most hated groups.
- The religious element tends to be the most targeted in anti-Muslim hate speech incidents. For instance, mosques and spaces for prayer evidence a higher negative association with Islam.
- The perception is that there is hatred springing from certain political parties and the mass media. On social media, anti-Muslim hatred is perceived to be very high, to be increasing and to be particularly linked with terrorist attacks. However, anti-Muslim discrimination seems to be lower than in other European countries that experienced Islamist terrorist attacks.
- Facebook is one of the most problematic social media networks in terms of anti-Muslim hate speech but it occupies a difficult position. Censorship would be an extreme measure, while stigmatising hate speech a more moderate position.
- Islamic organisations (UCOII) blame the media and social networks for double standards when dealing with anti-Muslim sentiments. By way of example, the President of UCOII had his Facebook page blocked for posting pictures of nuns on a beach in response to the French Prime Minister's hardening stances against the veil in public.
- Hatred is more malicious online than offline because the online world offers anonymity and 'protection' to hateful content.
- Muslims do not report hatred to avoid giving visibility to haters, because they do not know that they have a right to do so and because of a lack of a 'reporting culture'.
- Reactions to hatred are different and range from expressing grievances to reacting in a violent fashion.
- A study of online hate speech after the homicide case in Fermo, when a far-right man assassinated a Nigerian person, found that there were just a few significant negative tweets but these increase when mainstream media acts as an echo-chamber, thus giving resonance to the hashtag '*io sto con Amedeo*'.[15]
- Words such as '*musulmano*', '*islamico*', '*islam*', '*terrorista*', '*imam*' and associated words such as '*barconi*' should be monitored.[16]
- Exposing hate speech and engaging with haters to divert their ideas towards more moderate stances could be used as strategies to deal with Islamophobic people. Similarly, ignoring online hate speech may work to avoid incitement to hatred.
- Providers and social media Platforms should moderate online content.
- Counter-narratives can and should include irony as a weapon to challenge Islamophobia online.
- Online groups are very mobile and easily transition from one website or one Facebook webpage to another.

[15] English translation: I support Amedeo (i.e. the murderer).

[16] English translation: 'Muslim', 'Islamic', 'Islam', 'terrorist', 'imam', 'boats'

- Italy lacks 'Muslim neighbourhoods' (although it has ghettos based on a shared ethnicity) as in Great Britain and France, therefore there is not a feeling of being 'invaded' by Muslims.
- There are conspiracy theories about Muslims that are reminiscent of those that were directed against Jews in the 1920s.
- Recommendations for counter-narratives include having psychological expertise.
- There are several monitoring campaigns such as the European Council's 'no hate speech'.
- The Catholic Church under the leadership of Pope Francis has managed to take forward inter-religious dialogue between Christianity and other faiths (including Islam).

2.2. France

2.2.1. What is known about Islamophobia online

In order to prevent online hate speech and to understand the specificities of hate directed against Muslims in all its complexity, it is necessary to analyse the historical and legislative context. Several parameters participate in the construction of these discourses, sometimes independently, sometimes jointly.

There is both the influence of the law, on what forms of expression are permissible or prohibited, and the influence of society, conveyed by past and recent historical context but also by mass media, which define current hot topics and the ways these are discussed.

It is also important to always consider the communication tool used to develop these discourses, since its specific rules will shape both the form of the discourse and the target of the discourse according to the objectives of the users. The objectives of the sender of the speech are specific to the exchange framework. In our research, it is specific to the social networks and, more particularly, Twitter.

The form of hate speech implied by Twitter limitation must be considered. The maximum size of the text in a tweet prevents the development of well-constructed speech and forces writers to be more concise and often more divisive in their statements. Furthermore, the system of followers push users to make prominent remarks to attract attention and increase their audience. Thus, the exchanges between support and anti-Muslims groups are radicalized. 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. its number of sharing, seems to be linked to the controversial nature of its content.

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

The influence of each of the parameters mentioned above on hate speech and on the feeling it gives to transmitters that they have permission to espose such views must be taken into account, and they must not be underestimated.

The opinions that are acceptable to express in public, whether it is online or on mass media can be placed at the centre of a freedom spectrum of speech, and the extremist ideas at the edges. The centre is sometimes referred as 'Overton Window' (for a theoretical definition and discussion, see: Szałek, 2013). Dematteo (2018), illustrated how far-right activists in France and Italy are both working on the public perception of opinions to shift this so-called window in order to make their anti-migrants speech more tolerable and popular along mass media and social networks.

Regarding the historical context, as explained in previous reports for the Hatemeter Project, French colonial history with regions of Muslim culture in Africa, particularly Algeria, has created a form of tension and resentment. As for most colonial empires, France built its colonies on both economic exploitation and racial oppression, based on a belief of racial superiority. Today in France there remains frequent discrimination against anyone that displays any characteristic that might be related to North African population, whether that beclothing, names (Cediey and Foroni, 2006), cultural habits or religion (CNCDH, 2017).[17]

In the 20th century, the growing acceptance of Islamophobia, or hatred of Muslims, has its roots in the growth of the far-right in France. Simultaneously with an economic decline in the manufacturing sector, there has been an upsurge in far-right discourse that associates many societal problems with the presence of immigrant populations from Northern Africa. Although

[17] See https://sos-racisme.org/discrimination-au-logement-un-rapport-edifiant/

this discoursed had grown in significance since the mid-1990s, a milestone was reached in 2002 when the party of the National Front (FN) became an important political force by reaching the second round of the presidential election. The FN then saw its anti-immigrant speech broadcast by the media and thus legitimised (Wieviorka, 2013). This encouraged other politicians to follow suit and stereotype immigrants as the cause of many societal problems. The recent terrorist attacks, which were claimed by groups that espouse Islam, have only accelerated this stereotypical association. Despite the many calls not to associate murderous extremists with the global Muslim population, the media and politicians frequently make these shortcuts.

In recent times, the role of Muslims in France has become a major topic of political debates, specifically taking place on the most accessible of communication tools – the Internet. Following the terrorist attacks in Toulouse in 2012 and then in Paris in 2015, these online public discussions have only become more heated. Panic over Islamic fundamentalists has led some columnists, intellectuals and politicians to directly oppose Islam.[18] Researchers are also describing and analysing this issue (Geisser, 2003; Marwane 2014). Debates have played out in a variety of media formats, but there is a growing concern on the role of online media that can more easily cross the line into hate speech.

2.2.2. Legislative Framework

It would seem that French laws would limit this type of hate speech, particularly online. First, article 24 of the Law of 1881 (J.O., 29 July 1881, modified in 2012), on *liberté de la presse* (media freedom), states that discrimination, hate and violence based on origin, ethnicity, nation, race or religion is forbidden.[19] Another part of the same law specifies the different types of hate speech that are punishable. It includes speaking in public, writing in public venues but also writing anything exposed to the public by electronic or digital means.[20] Yet for those who might claim that some digital forums are private, a decree of the Penal Code (article R624-398) states that discrimination even in a non-public space due to origin, ethnicity, nation, race or religion is prohibited (J.O, 30 march 2005).[21]

In conjunction, either direct hate speech (insults, threats, etc.) or indirect hate speech (incitement to hate speech) directed against any person because of their religion or origin is illegal in France, both publicly and privately. All technical means of expression are covered by the law, including digital content.

On the surface, French law seems to favour the fight against anti-Muslims hate speech online. However, another law-related concept (perhaps more embedded in French society) complicates this assumption, that is the so-called *"laïcité",* which is defined directly in the French constitution text (see this references for a critical discussion of laicity principle and its impact on religious citizen: Baubérot, 2012; Fabre, 2017; Amiraux, 2015). This law, upon which French secularism is founded, is a cornerstone of the country's public policy and is an essential part of citizenship tests. Yet its interpretation is not without controversy. First adopted in 1905, in order to separate the state and the church and make French state institutions independent of any religious rules or representatives, at the same time it insures the right of French citizen to practice any religion.

Nevertheless, some interpretations of *laïcité* relegate religious beliefs to the private sphere, limiting the expression of religious practices in the public sphere, far beyond the law which simply prohibits conspicuous signs within public institutions, such as in schools or government offices. It is also culturally frowned upon to carry signs that may be associated with religion or to declare one's religious practice in any public space, such as a conference, a festival or any event that might receive media coverage, although the law does not prohibit it. This tension is even more pronounced when it comes to a person showing or declaring a religious affiliation during a televised debate or an interview. As a result, any type of religious sign, even if it is not accompanied by religious speech, can be interpreted

[18] See the tribune published in Le Figaro on the 19 March 2018: L'appel des 100 intellectuels contre le séparatisme islamiste. http://www.lefigaro.fr/vox/societe/2018/03/19/31003-0180319ART-FIG00299-l-appel-des-100-intellectuels-contre-le-separatisme-islamiste.php

[19] Original text: *Ceux qui, par l'un des moyens énoncés à l'article 23, auront provoqué à la discrimination, à la haine ou à la violence à l'égard d'une personne ou d'un groupe de personnes à raison de leur origine ou de leur appartenance ou de leur non-appartenance à une ethnie, une nation, une race ou une religion déterminée, seront punis d'un an d'emprisonnement et de 45 000 euros d'amende ou de l'une de ces deux peines seulement*.

[20] Original text: [...] s*oit par des discours, cris ou menaces proférés dans des lieux ou réunions publics, soit par des écrits, imprimés, dessins, gravures, peintures, emblèmes, images ou tout autre support de l'écrit, de la parole ou de l'image vendus ou distribués, mis en vente ou exposés dans des lieux ou réunions publics, soit par des placards ou des affiches exposés au regard du public, soit par tout moyen de communication au public par voie électronique* [...].

[21] Original text: *La diffamation non publique commise envers une personne ou un groupe de personnes à raison de leur origine ou de leur appartenance ou de leur non-appartenance, vraie ou supposée, à une ethnie, une nation, une race ou une religion déterminée est punie de l'amende prévue pour les contraventions de la 4e classe*.

as proselytising. It is even difficult for people adorned with religious symbols to simply walk on the street freely (Baehr and Gordon, 2013). This phenomenon is amplified for Muslims, especially for women wearing the *hijab* who are the most visually identifiable faith population.

Debates on the veil in French society have been frequent in 30 years. The first issues arriving in 1989 when two students wearing veils were refused access to the college by the school principal.[22] In December 1989, the Minister of National Education issued a circular, stating that teachers were responsible for accepting or refusing students wearing the veil in the classroom, on a case-by-case basis, leaving doubt as to whether religious symbols could be worn in the school. This circular remained valid until 2003.

Another period of tension occurred in 2004, when a specific law on the veil in public school was finally debated and adopted. The law of 15 March 2004 is regulating, following the principle of *laïcité*, the wearing of signs or clothing showing religious affiliation in public schools, middle schools and high-schools. It prohibits the wearing of any "ostensible" religious sign, including the veil.

Recently, for several months we have seen the return of these debates to the forefront, with a media peak in October 2019. Another terrorist attack occurred at the Paris police prefecture[23] was followed by government statements on the willingness to detect potential terrorists through the presence of "weak signals" of radicalization,[24] these weak signals being sometimes simple practices linked to Islam, some of which are not necessarily signs of religious extremism (such as the way of behaving and clothing adopted in public). This further reinforces the connection between religious practice and support for terrorism, which sustains a climate of mistrust and can then encourage online hate speech.

This event gave way to a media sequence focusing on the place of Islam in French society, during which a speech by the columnist Eric Zemmour, who had previously been convicted for incitement to hatred against a group of people because of their religion,[25] was broadcast live by the LCI channel. In this speech, he refers to Muslims as a threat for French society.[26]

A few days later, an elected representative of the *Rassemblement National* verbally questioned the president of the *Conseil Regional de Bourgogne* about a mother accompanying a class on a school trip wearing the veil. He urged the President of that Council to demand that this mother remove her veil, and, on the top of this, he staged himself doing it by filming his request during this regional council meeting and publishing the video on social networks.[27] The continuous news channels made it the main subject of their talk show for a week. Some observers pointed out the absence of veiled women during all this TV debates[28] and the media harassment against Muslims in a 'tribune' signed by 90 public personalities.[29] On the other hand, another tribune signed by 100 French Muslim citizens argues that the wearing of the veil is not a religious sign but shows the connection to particular movements of some Muslims with a particular reading of religious texts.[30]

All these events led the government to take the floor on this topic on several occasions and the subject remains in the news and against a background of the tension between critique of French laws and the sense of laïcité itself. All of this raises questions about the boundary between the rejection of multiculturalism and Islam in French society and a real hatred of Muslims.

At the same time, researchers analysing anti-Muslim discrimination, whether offline or online, have legal constraints as another law, "*Informatique et libertés*", passed in 1978, prohibits the collection of personal information on religious beliefs. According to this law, one cannot collect or process any personal data

[22] See https://fr.wikipedia.org/wiki/Voile_islamique_dans_les_écoles_en_France#Premiers_cas_(juin_1989)

[23] See https://www.lefigaro.fr/actualite-france/attaque-a-la-prefecture-de-police-de-paris-le-deroule-de-la-tuerie-minute-par-minute-20191005

[24] See https://www.bfmtv.com/police-justice/quels-sont-les-signaux-qui-permettent-d-identifier-un-cas-de-radicalisation-islamiste-1783573.html

[25] See https://www.lemonde.fr/societe/article/2019/09/20/eric-zemmour-definitivement-condamne-pour-provocation-a-la-haine-raciale_6012389_3224.html

[26] See https://www.liberation.fr/checknews/2019/10/28/discours-de-zemmour-sur-lci-le-csa-a-t-il-deja-effectue-d-autres-signalements-a-la-justice_1760147

[27] See https://www.europe1.fr/politique/video-virale-dun-elu-rn-prenant-a-partie-une-femme-voilee-au-conseil-regional-de-bourgogne-franche-comte-3925024

[28] See https://www.liberation.fr/checknews/2019/10/17/une-semaine-sur-les-chaines-d-info-85-debats-sur-le-voile-286-invitations-et-0-femme-voilee_1758162; https://www.liberation.fr/debats/2019/10/20/la-femme-voilee-ne-parle-pas_1758721

[29] See https://www.lemonde.fr/idees/article/2019/10/15/jusqu-ou-laisserons-nous-passer-la-haine-des-musulmans_6015557_3232.html

[30] See https://www.marianne.net/debattons/tribunes/le-voile-est-sexiste-et-obscurantiste-l-appel-de-100-musulmanes-de-france

that show, directly or indirectly, one's racial origins or ethnicity, as well as political, philosophical or religious opinions. Failure to respect this prohibition is punishable by Article 226-19 of the Penal Code, leading to five years' imprisonment and a 300,000 euro fine (J.O 12 December 2018). According to a decision from the *Conseil Constitutionnel*, which modified this law in 2007, researchers are able to collect data on ethnicity and religious opinions, but only if they obtain the authorization of the *Commission Nationale de l'Informatique et des Libertés (CNIL)* (J.O, 15 November 2007).

This challenge, however, does not prevent people from making assumptions about the number of Muslims in France, which is often over-estimated. According to a survey conducted in 2016, those polled estimated that 31% of the population is Muslim but the true figure is only 7.5% (Duffy, 2016). By the end of 2017, Muslims in France are projected to comprise about 5.7 million or 8.8% of the total population, yet some far-right websites claim that the number is as high as 20 million (Banet and Fauchet, 2018). The last official data aggregated by the French government on the number of practitioners of different religions date back to the census of 1872, during the Third Republic, before the secularism law (Dargent, 2009).

Yet, a scholarly understanding of Islamophobia remains challenging. The use of this kind of data on ethnicity and religion divides researchers, anti-racist associations and politicians. Some argue that this so called "*statistiques ethniques*" can be used in a discriminatory way to validate racist theories, whereas others argue it is needed to verify the facts and help in the fight against racism. For instance, online debates flourished after the mayor of Bézier was sued for gathering data on students' religions in the different schools of the city. An identical debate, on which position should be adopted and whether it may be racist or anti-racist, applies for the use of the word "*race*" and categories such as "*racisés*" (racialised) and "*blanc*" (white). These terms are discussed equally by the media, the political sphere or the academics (Devriendt, Monte and Sandré, 2018; Simon, 2008). In the summer of 2017, the word "*race*" was even removed from the Article 1 of the French constitution, as it was viewed a racist tool. The constitution previously mentioned that France ensures equality before the law of all citizens without distinction of origin, race or religion. It had been the only place where "*race*" was mentioned in the constitution, yet it has now been erased, with the word "*sexe*" being added instead.[31]

Although the question is divisive across academia and the general public, the vote among politicians was unanimous.[32]

In July 2019 a new law was passed at l'Assemblée Nationale specifically regarding online hate speech. It prohibits a variety of content on social network, including abusive messages inciting hate, discrimination or violence against people, on the basis of their sex, sexual orientation, gender identity, disability, ethnicity, nationality, "race" or religion . The social networks specified are: Facebook, Instagram, Twitter and Youtube. Any French citizen will have the right to report illicit contents. If a message or content is clearly illegal, it have to be deleted in less than 24 hours and replace by a message indicating its deletion. Any social network or search engine that does not cooperate shall be punished by one year's imprisonment and a fine of 250,000 euros.

During the debates preceding the adoption of the text by the Assembly, Deputy Avia, who proposed the law, agreed to amend her text after complaints made on Twitter by secularist claimants. In the list of prohibited hate ideologies propagated online, the term "Islamophobe" was initially proposed but secular activists have a negative view of the term "Islamophobia", which they fear will be used later to prohibit any criticism of the religion itself rather than to defend the rights of Muslim citizens. So, "Islamophobe" had been replace by the term "anti-Muslim". This term switch in the law once again highlights the particular tension around the place of Muslims in society and the epistemological battles over terms qualifying hatred of Muslims.

Finally, we can also note that a shift occurred in the recent years on the way racism is discussed and understood by politicians. This change is accompanied by a new division between political parties. Tahata (2018) studied this question by analysing 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 needed to be defined, by virtue of its majority status, and was opposed to the "other". Then in the after 2000 period, Tahata shows that the French is regularly define in the discourse and associated to strong moral values typical to French society and French

[31] See Europe 1 (2018) Constitution française : l'Assemblée supprime le mot «race», interdit la «distinction de sexe» http://www.europe1.fr/politique/constitution-francaise-lassemblee-supprime-le-mot-race-interdit-la-distinction-de-sexe-3708722

[32] See Libération (2018) Supprimer le mot «race» de la Constitution: oui, mais… http://www.liberation.fr/debats/2018/07/10/supprimer-le-mot-race-de-la-constitution-oui-mais_1665506

republic. A foreigner is then not just seen as other but as a person with different culture and different values that are not compatible with French society. This way of viewing foreigners is shared by a part of all political groups, particularly on the basis of the willingness of *laïque* (secular) and atheistic society. This new opposition between French culture and foreign culture, seen as invasive, is sometimes described as racism without "race" or neo-racism. According to Froio (2018), it is now possible to distinguish two racist positions, which are called "racism" and "neo-racism", the first based on biological arguments, and the second supported by cultural and religious arguments and claiming incompatibility between two civilizations. According to the principles of this second position, the values of republic and secularism are used as justification to exclude Muslims.

2.2.3. Main results of in-depth interviews

In France, four people have been in-depth interviewed, again in order to scope the online problem of Islamophobia with the support of stakeholders groups. Interviews have been conducted with activists on the field from three different NGOs, who have expertise in the subject and are direct witness of anti-Muslim hate speech. The four activists are Muslim women, working respectively for an association fighting against online harassment, an association which is producing positive content on the life of French Muslims women, and an association fighting for the right of Muslims in society (one interviewed works for both the last two associations mentioned). Islamophobia is everywhere in French society: on the Internet, on the street, in the schools (law against the veil in classroom), in health care (some are treated differently for stereotypical consideration based on origins - Arab and black women are seen as better able to bear pain), when looking for work, etc.

- Islamophobia takes multiple forms.
- Racism is part of a coherent narrative which is followed by some individuals who are sometimes among the political or scientific elite.
- The main goal of Islamophobia is to silence Muslims.
- Bringing the harassment of Muslim people to justice cost a lot of money and is time-consuming.
- Many Muslims do not know about the organisations fighting Islamophobia and do not know their rights, so they are not reporting the attacks or the hate speech. This leads to a significant dark figure (i.e. underestimated number) of Islamophobia.
- Most of the time, Islamophobic groups just share information about Muslim activists without any comment to prevent accusations of racism. Or they use subtle language.
- Those who have the letter "Nun" (ن) in their Twitter name are usually people supporting the Christians of the Middle-East, and many of them are Islamophobic because they are against ISIS and think a lot of French Muslims are pro-ISIS.
- Harassment is often taking the form of raids, following a more or less organised talk in the Islamophobic groups. Islamophobia is basically triggered by anyone who shows themselves in public and speaking about any subject, even non-political, while being identified as a Muslim. For example: a woman making YouTube videos about beauty/fashion, or just people expressing themselves on a political, not religious related, topic but while wearing a veil (e.g. Maryam Pougetoux).
- Muslim famous person in France are often object of online control by Islamophobic groups, who seek information on their private life to spread (e.g. Mennel Ibtessim or Tarik Rammadan).
- Muslim women targeted by Islamophobic groups online can easily be victims of misogyny, revenge porn, disclosure of private photos, private information or home address. They also receive several private messages, death threats or pictures of beheaded women (accusations of supporting ISIS).
- On the Internet, people feel invincible because they are anonymous, and they can even be viewed as heroes or martyrs to the eyes of their groups or networks if they are confronted and if their account is deleted.
- News articles containing fake news accusing a Muslim association to be an extremist group exist online. Once an association is suspected to be related to extremists, their rights to employ someone or to benefit from all the usual state help is jeopardised.
- Every time there is a news media controversy including a Muslim women, other Muslim women or associations of Muslims can be harassed again as they are all considered the same.
- Sometimes government leaders or journalists say things that validate an online harassment campaign.

- On social media, Islamophobia is concentrated around the fight between the secularist group and the so called "Islam-politique" group (note that "Islam-politique" is a right-wing label, not the definition they give to themselves).
- Far-right activists are not necessarily the source of Islamophobia, as they just use the fact that Islamophobia is tolerated and arises from political discourse including that of the French government. Asking Facebook and Twitter to fight Islamophobia themselves and accusing isolated and extremist people of being the perpetrators of hate speech, is actually hindering the real fight of the deconstruction of societal Islamophobia.
- Cyber harassment can result in victims suffering from burn-out, anxiety, depression, suicide attempts, and the same post-traumatic effects as sexual assault. Symptoms exhibited by victims include crying, feeling trapped, being hyperactive, vomiting, going into a trance, etc.
- Shame often prevents harassed people to ask for help, including their family. Victims often shut down their social media accounts and become isolated.
- An important consequence of online harassment is that being only offline nowadays cannot be considered as a good solution since everyone needs to use social media for their personal, professional and associative activities.
- An association tried to work with Facebook and Twitter, but they said they already do their best. These Platform companies need to track IP addresses and ban them instead of just deleting the accounts.
- To report someone, you need to identify with your official ids, and it can be difficult, then if you succeed, they just delete the tweet and it doesn't help much because the person continues to have an influence on social media. Sometimes, the account is deleted but the person can create a new one, and people who have been deleted by Twitter are considered as heroes or martyrs by the other people of the same community, so that is not effective in the end.
- There was some success in tackling online hate speech in the past, particularly when the victim was a famous person if there is a racist or sexist speech involved, but never when there is Islamophobic speech.

2.3. United Kingdom

2.3.1. What is known about Islamophobia online

The Internet has afforded opportunities for like-minded individuals to locate each other and bypass traditional mass media gatekeepers, resulting in a huge number of online communities appearing into exist below and across national communities with varied social, political, religious or commercial raisons d'être (Kohl, 2018). This is particularly acute when observing social media Platforms with multifarious groups, subgroups and networks (Kohl, 2018). Furthermore, online spaces have been noted to continually invert and blur the boundaries between the private and public, in which content which may have been typically restricted to privatised spaces, now encroaches public, online spaces. As such, online spaces create environments of "social disruption" where what is deemed acceptable and unacceptable, private and public, legitimate and illegitimate become progressively obscured distinctions (Allen, 2014: 2). Consequently, the Internet and social media Platforms such as Facebook and Twitter have become a popular sphere for online hate, partly due to their accessibility and the anonymity they offer for perpetrators who use it to intimidate, harass, and bully others (Christopherson, 2007).

The Internet has been utilised by far-right groups such as the English Defence League (EDL) and Britain First, who have used the web to create and establish a public presence, being successful in using social media Platforms, including Facebook and Twitter, to disseminate further online hate and intolerance toward people of Muslim faith (Barlow and Awan, 2016). Indeed, it has been noted that the EDL, described as an "Islamophobic new social movement" (Copsey, 2010: 5), was one of the first far-right movements to make extensive use of social media, being proactive in exploiting the virtual environment and using worldwide events to incite hatred towards Islam and Muslims (Awan and Zempi, 2015). Analysing data from the Islamophobia monitoring project *Tell MAMA* (Measuring Anti-Muslim Attacks), Copsey *et al.,* (2013) found that the majority of incidents of anti-Muslim hate crime reported were online incidents and "300 – 69 percent – of these online cases reported a link to the far-right" (p. 21), specifically the EDL and British National Party (BNP). Of the online incidents that reported a link to the far right however, it was the EDL, rather than the BNP that was named in 49 percent of such cases, signifying that the EDL is the far-right organisation most active on the Internet in terms of circulating anti-Muslim sentiments (Copsey *et al.,* 2013). In addition, Copsey *et al.,* (2013) also discovered that most of the online hate reported

to *Tell MAMA* was committed by males, comprising of 321 incidents, whereas women were responsible for 78 incidents, constituting 18 percent of the online incidents reported, with 18 of these cases reporting a link with the EDL.

In a study by Brindle (2016) which examined two corpora to analyse discourses produced by the EDL's group leaders on their official website, and on the official EDL Facebook page by group supporters produced similar findings. Brindle (2016) observed that in the former corpus, EDL group leaders focused on the opposition to extremism within Islam, whilst in the latter corpus, EDL supporters constructed Islam and Muslims as a threat to their position in society and their way of life, with no effort made to differentiate between radical and non-radical forms of Islam, to which the religion was viewed as being in fundamental conflict with 'Englishness'. As such, Brindle's (2016) findings suggest that the EDL is an organisation that is opposed all forms of Islam and immigration of Muslims to the UK, which stands in contrast with the organisation's mission statement which declares that the aim of the EDL is to oppose forms of radical Islam in the UK. Likewise, an investigation of active users on an online EDL message board undertaken by Cleland *et al.,* (2017) revealed that the existence of several threads operating within a broader theme of Islamophobia in which posters discussed Muslims as socially and culturally problematic and Islam as the opposite to British values and identity, and a key cause of social decline in the UK. Moreover, many posts across the message board were found to be replete with openly racist language directed towards Muslims and Islam, concerning perceived cultural differences that deemed Muslims to be cultural outsiders and a threat to British culture Cleland *et al.,* (2017). A significant finding of the research was that at no point was racist language challenged on the message board; rather it was mutually supported by other users, with such uncontested behaviour providing important evidence to dispute the EDL's claim of being an anti-racist organisation.

In a recent study, Awan (2016) conducted research focusing on Islamophobia on Facebook. Awan (2016) created a typology of five offender behaviour characteristics based on the themes that emerged from examining 100 different Facebook pages, comments and posts, in which 494 instances of specific anti-Muslim hate speech was discovered. These five types which encompass perpetrators who have been engaged with Facebook as a means to target Muslim communities with online hate are: the *Opportunistic* (someone using Facebook to create a posts and comments of hate directed against Muslim communities after a particular incident. In particular, these individuals are using Facebook to post threats and promote violence); the *Deceptive* (someone creating fear through the use of posts which are specifically related to false events in order to intensify the Islamophobic hate comments online. For example, a number of people were attempting to capitalise on false stories with links to incidents such as Peppa Pig and Halal meat); *Fantasists* (someone using Facebook webpages to fantasise over Muslim deaths and consequences with respect to Muslim events. In particular, these individuals have blurred the lines between reality and fiction and are making direct threats against Muslim communities); *Producers* (people who use and promote racist images and videos which are used as a means to create a climate of fear, anti-Muslim hate and hostility. These individuals are closely linked to the distributors); and finally, *Distributors* (people who use social media and Facebook in order to distribute messages of online hate through posts, likes, images, videos and comments on Facebook). Awan (2016) found that the majority of people involved in these acts were males (805) and females (20%). A number of the individuals were predominantly based in the UK (43%), however there were also a number of online users who were identified as being from the United States (37%) and Australia (20%). Additionally, Awan (2016) identified that a number of comments and posts revealed individuals with direct links to organisations such as Britain First and the EDL.

Monitoring organisations such as *Tell MAMA* have suggested that levels of Islamophobic discourse online are ongoing and increasing, especially through Twitter (Allen, 2014). In 2016, *Tell MAMA* documented 340 anti-Muslim crimes or incidents, of which 311 were verified that were classified as 'online', occurring on social media Platforms such as Facebook, Twitter, YouTube and Instagram, or other Internet-based websites, such as forums and online newspapers. Most of the online incidents reported to *Tell MAMA* (2016) fell under the category 'Abusive Behaviour' at 84% (n=261), with some falling under the rubric of 'Threats' being 9% (n=29) and Anti-Muslim literature comprising of 7% (n=21) – a term which broadly includes racialised memes in the online sphere. As Awan and Zempi (2015) highlight, online anti-Muslim abuse occurring on social networking sites, such as Twitter, can be categorized as being "cyber harassment", "cyber bullying", "cyber abuse", "cyber incitement/threats", and "cyber hate" (p. 12). According to Awan (2014), many of the comments that are posted online through social networking Platforms possess an extremist and incendiary undertone, and also transpire on blogging sites (see Ekman, 2015), online chat rooms and other virtual Platforms which have been used to propagate online anti-Muslim hatred, often in the form of racist jokes and stereotypical 'banter'. As such, the Internet and

social media sites are popular arenas for online hate to flourish, partly due to their accessibility and the anonymity they offer for offenders who use it to intimidate, harass, and bully others (Awan, 2014).

According to Awan and Zempi (2015), the prevalence and severity of virtual and physical world anti-Muslim hate crimes are influenced by 'trigger' events of local, national and international significance. Terrorist attacks carried out by individuals who identify themselves as being Muslim or acting in the name of Islam, such as the Woolwich attack, the atrocities committed by ISIS and attacks around the world such as Sydney, the Charlie Hebdo attack in Paris and attacks in Copenhagen and Tunisia have induced a significant increase in participants' virtual and physical world anti-Muslim hate crime experiences. Additionally, national scandals such as the grooming of young girls in Rotherham by groups of Pakistani men, and the alleged 'Trojan Horse' scandal in Birmingham framed as a 'jihadist plot' to take over schools, were also highlighted by participants as 'trigger' events, which increased their vulnerability to anti-Muslim hostility. Such assertions are supported by Feldman and Littler's (2014) research, who discovered that in the wake the brutal murder of the soldier Lee Rigby in Woolwich, reported incidents to *Tell MAMA* skyrocketed – there were nearly four times more online and offline reports (373%) in the week after 22 May 2013 than in the week beforehand.

Törnberg and Törnberg (2016) investigated patterns of representation around the words 'Muslim' and 'Islam' in a 105 million-word corpus of a large Swedish Internet forum from the years 2000 to 2013. The findings illustrated that Muslims are portrayed as a homogeneous outgroup implicated in conflict, violence, with such negative characteristics described as emanating from Islam. Islam was depicted as a monolithic group as opposed to a religion that contains a diverse multitude of outlooks among its adherents. Törnberg and Törnberg (2016) reported that these patterns are often more extreme versions of those previously found in analysis of traditional media forms.

Oboler (2016) reported that Facebook is being used to normalise Islamophobia and associated discourses. After a qualitative analysis of 349 posts on Facebook, Oboler found several themes that depict Muslims as manipulative, dishonest and a threat to security and to Western way of life. In an earlier study concerning online Islamophobia conducted by Oboler (2013), culminating in a report for The Online Hate Prevention Centre, over 50 different Facebook pages were analysed, and which illustrated a clear correlation of hate speeches online that were targeted toward Muslims. Overall, 349 separate instances of online hate speeches directed against Muslims were discovered, including a number of Facebook pages created in order to specifically target Muslim communities. For example, the 'Boycott all Halal products in Australia!' page, which has over 520 likes, the page "The truth about Islam" which has over 150,000 likes, the 'Islam is Evil' page (418 likes), the 'Prophet Muhammad Still Burns in Hell' page which has 470 likes (Oboler, 2013). The Internet, therefore, was found to reinforce existing discourses in traditional media. Allen's (2014) study found similar strong links of Facebook users and growing public opposition about mosques. Allen (2014) found that members were engaged actively in online discourse which was opposed to the mosque. Some of the themes that emerged from this, included, issues regarding social identity, otherness and the Islamification of Britain. Another problem that emerged within the discourse of Islamophobia was the issue of Muslims being a threat to national security.

In his exploration of anti-Muslim hate crime on Twitter, Awan (2014) discovered that were a number of terms that were used to describe Muslims in a negative manner; these included the words "Muslim pigs" (9%), "Muzrats" (14%), "Muslim Paedos" (30%), ''Muslim terrorists'' (22%), "Muslim scum" (15%), and "Pisslam" (10%). Likewise, Awan's (2016) examination of Islamophobia on Facebook revealed, via a word frequency count of comments and posts to ascertain words and patterns directly related to anti-Muslim hate, the presence of six key words that depicted Muslims in an overtly prejudicial way, including the words; "Muzrats" (18); "Paedo" (22); "Rapists" (24); "Paki" (25), "Scum" (28) and "Terrorists" (22). These words were accompanied by images and texts that were posted following high-profile incidents, including spate of beheadings by ISIS and the Rotherham abuse scandal in the UK. Similarly, the Islamophobia monitoring project *Tell MAMA* undertook an investigation into the use of words to label Muslims from the time period of January 2013 to December 2013 of incidents received, collating high-frequency words that were directly related to anti-Muslim hate and prejudice. *Tell MAMA* also found that the terms "Muzrats", "Ninja" and "Paedo" were being used against Muslims (Tell MAMA, 2014). As a report by *Tell* MAMA (2016) affirms, the usage of the term 'Muzrat' demonstrates the unique vernacular of dehumanising language when discussing Muslims in online spaces and remains a rhetoric that does not always translate into street-based abuse (see also Oboler, 2013).

As part of wider efforts to understand the scale, scope and nature of uses of social media that are possibly socially problematic and damaging, research conducted by the Centre of the Analysis of Social Media

(CASM) at Demos measured the volume of messages on Twitter considered to be derogatory towards Muslims over the duration of a year, from March 2016 to March 2017, also yielded interesting results. Over the course of a year, researchers Miller and Smith detected 143,920 tweets sent from the UK that were considered to be derogatory and anti-Islamic, totalling around 393 tweets a day, with such tweets being sent from over 47,000 different Twitter users. These tweets fell into a number of different categories, from directed angry insults, to broader political statements, with a random sample of hateful tweets being manually classified into three broad categories: a) 'Insult', in which tweets were used as an anti-Islamic slut in a derogatory way, often directed at a specific individual; b) 'Muslims are terrorists', constituting around one fifth of tweets characterised by derogatory statements that generally associated Muslims and Islam with terrorism; c) 'Muslims are the enemy', comprising just under two fifths of tweets denoting statements that claimed that Muslims, generally are dedicated to the cultural and social destruction of the West (Miller and Smith, 2017). Demos' (2017) research study also identified six online tribes, to which the largest group was 'Core political anti-Islam', a politically active group engaged with international politics, composed of about 64,000 users that included recipients of tweets. Miller and Smith (2017) found that hashtags employed by this group suggested engagement in anti-Islam and right-wing political conversations, including: #maga (Make America Great Again), #tcot (Top Conservatives on Twitter), #auspol (Australian Politics), #banIslam, #stopIslam and #rapefugees (Miller and Smith, 2017).

A report published by *Tell MAMA* (2016) has highlighted the emergence of British Muslims as a racialised threat, an 'alien other' that possesses beliefs that contrast with mainstream society, which have become synonymous with 'deviance', 'un-Britishness' and terrorism. Whereas British Muslims are a heterogenous group that comprise many different ethnic and cultural backgrounds (see Abbas, 2010), as well as religious practices and various orientations of Islam, such as Sunni, Shia and Ahmadiyya, negative and salacious media coverage reduces such complexities into binaries of cultural difference. Due to this form of political rhetoric and sensational media reporting, Muslims, particularly Muslim men have been constructed as 'The New Folk Devils' - aggressive hotheads who are in danger of being brainwashed into terrorists (see Gill and Harrison, 2015; Tell MAMA, 2016). This has intensified in recent years with the onset of child sexual exploitation (CSE) scandals such as the aforementioned Rotherham incident, which have focused on the race, ethnicity and faith of organised criminal 'grooming gangs' targeting vulnerable young people across the UK. Such cases have brought into question the role of the potentially 'dangerous masculinity' of British Muslim men, which has resulted in a conflation between the Pakistani community and the constructed idea of the 'Muslim fundamentalist'. Consequently, the racial epithet 'Paki' has become interchangeable with British Muslims regardless of ethnic background, with others using this term to group sexual deviance with Islam or Muslim identity more broadly (Tell MAMA, 2016). Indeed, as a statement from a poster from the Faith Matters (2014) study highlights: "I noticed on the BBC News they said Pakistani Asian men, they wouldn't say MUSLIMS" (p. 9). Allen's (2014) research examining a pilot study which sought to investigate opposition to a proposed "super mosque" in the town of Dudley in the West Midlands region of the UK also revealed similar sentiments. Focusing on the Facebook group 'Stop Dudley Super Mosque and Islamic Village', members were engaged online to explore why they opposed the mosque, with disparaging responses stating that: "[the mosque would] ...mean more paki's will commute into the Dudley area, thus creating a curry infested atmosphere and I for one despise the cunts" (Allen, 2014: 8). However, Allen (2014) stresses that only a minority of responses were laden with such overtly discriminatory ad offensive expressions. In the minds of some respondents at least, utterances that were racist and racialised discourses were evident, where religious and racial markers – Muslim and 'Paki' – were interchangeable if not entirely the same (Allen, 2014).

2.3.2. Legislative Framework

In the United Kingdom, there is no single hate crime and hate speech legislation. In England, Wales and Scotland, the Public Order Act 1986 prohibits, by its Part 3, expressions of racial hatred toward someone on account of that person's colour, race, nationality (including citizenship), ethnic or national origin. Section 18 of the Act states:

"A person who uses threatening, abusive or insulting words or behaviour, or displays any written material which is threatening, abusive or insulting, is guilty of an offence if—

(a) he intends thereby to stir up racial hatred, or

(b) having regard to all the circumstances racial hatred is likely to be stirred up thereby".

Offences under Part 3 carry a maximum sentence of seven years imprisonment or a fine or both.

The Racial and Religious Hatred Act 2006 amended the Public Order Act 1986 by adding Part 3A. That Part declares, *"A person who uses threatening words*

or behaviour, or displays any written material which is threatening, is guilty of an offence if he intends thereby to stir up religious hatred". The Part protects freedom of expression by stating in Section 29J:

"Nothing in this Part shall be read or given effect in a way which prohibits or restricts discussion, criticism or expressions of antipathy, dislike, ridicule, insult or abuse of particular religions or the beliefs or practices of their adherents, or of any other belief system or the beliefs or practices of its adherents, or proselytising or urging adherents of a different religion or belief system to cease practising their religion or belief system".

The Criminal Justice and Immigration Act 2008 amended Part 3A of the Public Order Act 1986. The amended Part 3A adds, for England and Wales, the offence of inciting hatred on the ground of sexual orientation. All the offences in Part 3 attach to the following acts: the use of words or behaviour or display of written material, publishing or distributing written material, the public performance of a play, distributing, showing or playing a recording, broadcasting or including a programme in a programme service, and possession of inflammatory material. In the circumstances of hatred based on religious belief or on sexual orientation, the relevant act (namely, words, behaviour, written material, or recordings, or programme) must be threatening and not just abusive or insulting.

Concerning online hate speech and hate crime, it has been argued that current UK legislation is not fit for purpose in the digital age (Bakalis, 2018), resulting in an under-prosecution of online hate crime. As such, significant discussions have ensued about the unique challenges involved in effectively tackling hate crime that occurs online (SELMA, 2019). In 2015, a Minister for Internet Safety and Security was appointed, with tackling online hate included within their portfolio. In January 2018, a National Online Hate Crime Hub was launched, staffed by a team of specialists tasked with assessing and managing cases of online hate and referring to the appropriate authorities for handling (SELMA, 2019).

2.3.3. Main results of in-depth interviews

As in the other countries, also in the UK context experts have been in-depth interviewed to explore the online problem of Islamophobia and assess background experiences, needs and aspirations of NGO/CSOs and other target stakeholders groups. Interviewed experts were ten. One expert for Nottingham Trent University, University of Limerick, La Trobe University Law School, University of Leicester, Lancaster University, and Malmö University respectively, two experts from Teesside University, and one NGO representative. Below, it is reported a summary of the main points emerged during the interviews.

- All interviewees mentioned that there are difficulties in measuring the online space in terms of anti-Muslim hatred, how can it be quantified, detected and controlled.
- There must be better, effective methods in online reporting functions that enable hateful content to be followed up with a good, tangible outcome for the victim, especially from popular social media Platforms.
- Both the offline and online contexts are part of the same phenomena, and part of the same manifestations of anti-Muslim hatred and Islamophobia.
- Compellingly, all the interviewed experts stated that they had received abuse in the online sphere, comprised of verbal abuse in the form of threats as well as disparaging remarks about Islam and Muslims, and were called 'traitors' and 'quislings' due to their work on anti-Muslim hatred and Islamophobia.
- Although Anti-Muslim hatred has existed historically, Islamophobia in the current context has been shaped by important geopolitical factors such as the Iranian Revolution of 1979, the media publicity surrounding Ayatollah Khomeini, the fatwa issued against Salman Rushdie in 1989 to name a few have all contributed to the present state of Islamophobia today.
- Certain crimes such as grooming gangs have gone unchallenged, and are conflated with points of view that the Prophet Mohammed had a wife of 9 years of age, and is used make the argument that all Muslims are paedophiles as it has been sanctioned in the Qur'an, which has been advantageous for far-right groups to justify their views.
- Given the nature of law enforcement monitoring activities on social media websites, it is very likely to activities pertaining to online hate will move into the private, online sphere, such as WhatsApp group chats, or private Facebook pages.
- Accessing such pages raises human rights and legal issues.
- The anti-Muslim discourse of the far-right is much more aggressive than previously, in a very short period of the time in the UK. We as a society have failed to tackle it.
- More support and protection is needed for victims of online anti-Muslim hatred, especially psychological support, as many people develop mental health problems due them suffering Islamophobic victimisation.

- Research is needed into the impact of mental health and/or the health of Muslims as a result of experiencing anti-Muslim hatred; perhaps through working together with the NHS, the police, and victim support organisations to raise awareness and support.
- Article on Muslims can easily be shared on Facebook and/or Twitter.
- The source rather than the Platform is more important to monitor however, as the source will use multiple Platforms, for example far-right groups such as Generation Identity possess a dedicated website which also features links to their Twitter page, Instagram account, Facebook profile, demonstrating a huge convergence right across the Platforms.
- Certain markers can indicate whether an online group, page or sole profile is associated with the far-right and engaging in anti-Muslim activity, such as images of 'Pepe the Frog' and/or symbols such as Skull and Crossbones, use of the Fraktur font, triangles and the Arabic symbol for 'Christian' – ن.
- As regards to online counter-narratives, it would be prudent to avoid repeating the anti-Muslim narrative of the person who is engaging with anti-Muslim content online, and engage in a different way, from a divergent, more positive angle.
- An effective and important counter-narrative that could be disseminated in both the offline and online realms would be a campaign based around normalising Muslim people, in the sense of featuring average Muslim men and women that highlights the contributions that Muslims make to society, and demonstrates the normalcy Muslims as being both British and Muslim. In doing so, this would go a long way to allay societal fears and questions that being Muslim and British are mutually exclusive, incompatible elements.
- It is salient to recognise that hateful language used online is fluid – keywords can change, but what is imperative to examine are the hidden, coded messages, some of which are ambiguous and split up in smaller messages, but once joined together reveal the actual message.
- The use of humour can be deployed as a powerful tool via counter-narratives – a good online approach could be playing along with the negative generalisation and/or inaccurate facts, try to build on the poster's arguments, push the envelope further and further, so that the argument looks more and more ridiculous.
- Legislation does not change attitudes, it just curbs behaviours – there needs to be solid political leadership in the UK that challenges hostile views, especially those that are Islamophobic, which can help to prevent anti-Muslim ideas and views from growing and fermenting, in both online and offline contexts.
- Setting up fact-checking websites in regards to the religion of Islam and Muslims is an important counter-narrative, especially for people who may start to tentatively access the online far-right culture and explore associated narratives.
- Offline counter-narratives are also salient, in terms of myth-busting inaccurate myths, stereotypes and/or ignorant views.
- Doing so that can beneficially aid in efforts to break down views and misconceptions about Islam and Muslims and also for people who hold stronger, ingrained anti-Muslim views, look at the consequences of their involvement, how it will affect victims and the potential perpetrator or affiliate as well.
- An effective method could be to hold offline workshops such as those that contextualise issues such as verses and hadiths from the Qur'an being taken out of context, resulting in distorted meanings, and explain their meanings accurately.
- It is essential that social media companies develop more robust reporting mechanisms that constantly keep the victim or NGO or CSO in communications of the process, including what is going to be done, the progress of the report, and what the outcome is going to be. Failure to keep in contact with the victim or NGO or CSO after a report will most likely result in a lack of faith of the particular Platform' reporting process, elicit feelings of frustration for the victim, and worryingly, cause reluctance to report further instances of online hate experienced.
- Importantly, online counter-narratives, either disseminated by NGOs or CSOs or software tools, should not be following 'scripts' to form counter-messages, as this creates artificiality and consequently lessens the salient human element, which deters human interaction.
- Internet users are cognisant in noticing patterns with the use of scripts, believing them to be fake profiles or 'bots', and can alert other users of their suspicions or simply use the block function. This can be viewed as a failure, as it closes down an avenue for engaging in cogent dialogue with online actors.
- Effective counter-narratives that are going to be utilised and disseminated by NGOs, CSOs and software tools online to tackle anti-Muslim hatred in the fora should be comprised of three distinctive

elements – facts concerning Islam and Muslims to clear myths and negative misconceptions, the use of humour and/or witty banter to engage with users, and the implementation of a person-focused approach.

- This creates a dynamic process where lucid dialogue can take place online to effectively counter users that are disseminating and actively participating activities pertaining to anti-Muslim hatred.

As mentioned in the first chapter, Project Hatemeter seeks to strengthen cooperation between key actors, and to guarantee the broadest distribution and enduring impact of project results on future research streams and operational strategies. To do so effectively, the project encourages capacity building and training and the sustainability and transferability of the Hatemeter Platform among other target stakeholder groups, including victims of Islamophobia and anti-Muslim hatred, LEAs, journalists/media and civil servants. Within this context, a focus group with target stakeholder groups was conducted on June 27, 2018 at Middlesbrough Campus of Teesside University in the United Kingdom, with the aim of supporting the project's objectives of reinforced collaboration, wide circulation, as well as sustainability and transferability of the Hatemeter Platform. The focus group lasted approximately 2 hours (10:15-12:15 UK time). Eight participants attended the focus group. The participant were comprised of: Hate Crime Investigative Officers, British Police (2), local Council Strategic Cohesion and Migration Manager (1), victim of anti-Muslim hate (1), independent expert on hate crime (1), expert in Islamophobia, Centre for Fascist and Anti-Fascist Studies (1), Community Development Officer (Integration and Community Cohesion), Council/Police Independent Advisory Group -Hate Crime (1), Professor of Criminology, Teesside University (1), Research Associate, Teesside University (1).

- LEAs (Law Enforcement Agents) particularly noted that social media Platforms and constantly dynamic, fluid, and evolving in number, which increases the possibility for online hate to manifest and be perpetrated as well as consequently experienced on an array of online forms.
- All participants noted that the youth tend to use Instagram and Snapchat, whereas older people use Facebook and Twitter, with the use of certain Platforms and sites reflecting age demographics.
- In the UK, the participants were cognisant of several far-right groups that are active online, and were able to identify such groups, including: the Football Lad's Alliance (FLA), North-East Infidels (off-shoot of the English Defence League), Bishop Auckland Against Islam, Middlesbrough Against Mass Immigration, and also Military Veterans' groups e.g. Veterans Against Terrorism.
- These groups have used the web to create and establish a public presence, being successful in using social media Platforms, including Facebook and Twitter, to disseminate further online hate and intolerance toward people of Muslim faith. These groups never make a distinction between radical and moderate forms of Islam.
- A common representation in these groups narratives is that Muslim communities undermine British culture and traditions. The problem, however, might be the diffusion of these representation and the acceptance by individuals who are not affiliated with the aforementioned groups, especially the least known for their anti-Muslim sentiments.
- Several participants stated that they had noticed a disturbing trend in receiving reports of school-aged children aged between 8, 9, or 10 years of age engaging in hate (anti-Muslim among other) activities online. LEAs noted the average perpetrator of online (and offline) anti-Muslim hatred is White males between 40-60 years of age, with the average victims online being Muslim women, often those who wear traditional Muslim clothing.
- The experience of reporting online victimisation is exactly the same as reporting offline victimisation. There is the perception on the part of the victims that the authorities are not doing enough to deal with the issue. Participants identified that there are significant issues and barriers with the online reporting of online hate, particularly in reference to the reporting process. A major barrier is the lack of support from the major social media Platforms – Twitter and Facebook – to back LEAs.
- LEAs spoke of their negative experiences with online reporting functions, citing their struggle to remove an inflammatory, offensive Facebook post since March 2018, which they had reported numerous times, but still exists on the Platform and is causing offence to the targeted individual.
- LEAs agreed that anti-Muslim hatred is easier to deal with in the offline context. For example, if an individual is shouting obscenities in the street and/or using derogatory terms, it is easier to secure evidence and enable a prosecution through acts such as Public Order Offences. In the offline context, LEAs stated that there is a greater likelihood of attaining a positive outcome for the victim through an arrest and conviction, providing satisfaction to the victim.
- One LEA remarked that it is important to consider Facebook as just a Platform, with the likelihood

being that if suspected perpetrators are using one Platform to engage in online hate activity, they will be using other Platforms too.

- There are challenges in investigating, arresting and engaging in attempts to secure a conviction concerning online hate crime. LEAs mentioned that they receive many alerts of suspected online hate crimes occurring on Facebook, however there are certain criteria that must be met for such behaviour to be classed as a 'hate crime'. One LEA commented that on social media Platforms such as Facebook and/or Twitter, a user posting indirect statuses or Tweets is highly unlikely to be dealt with. If a user is, however, directing hate speech, such as religiously aggravated utterances towards a certain individual, this online behaviour then falls under 'malicious communications' and can be categorised as an offence.
- Participants were also aware of the existence of a fine line between legitimate criticism of Islam and anti-Muslim hatred, such as abusive, threatening speech.
- Participants highlighted that in regards to online hatred in general, it is important to address the technical issue of the offline-online distinction. Is online anti-Muslim hate crime more serious, are the effects on victims more consequential?
- An expert commented that it is salient to assess the causal relationship between the online-offline – will the tool be able to pick up evidence of the association? The online arena can be seen as a 'vent' for hatred, particularly anti-Muslim abhorrence, and can act as a 'safety valve' – it can displace activity, especially from far-right groups' offline.
- The participants posited an important question: how can the ICT tool differentiate, or is it able to differentiate, between a user posing a legitimate question or criticism about Islam or a question that possesses an Islamophobic undertone?
- Some participants, especially previous victims of anti-Muslim hatred were interested in whether the tool and software could pick up on soft-language that is often used in comment boards to pass online censors, comprised of sarcastic comments i.e. 'now I know the meaning of the religious of peace', derogatory references to perceived Islamic 'cultural practices', or action towards Muslim such as 'wrap them in bacon'.
- Much discussion was concerned on the issue of the counter-narratives: some participants were wary that reframing explicit Islamophobic comments to comments that are less Islamophobic legitimises Islamophobia in a different way through the use of subtle, less detectable speech.
- Another significant issue of discussion revolved around who is going to provide the counter-narratives – are they going to be predominantly Muslim organisations? Or government-funded think tanks e.g. Quilliam (which is perceived as highly controversial among Muslim groups? It was suggested that is absolutely essential that broad Muslim representation is included, especially from the different and often contesting sects within Islam, the two major branches being Sunni and Shia. As such, produced counter-narratives must be mindful of, and take into account the intra-religious dimension of Islam, and the different interpretations of Sunni and Shia Muslims, respectively.

03

Development of the Hatemeter platform

The present chapter provides the technical description for the development of the Hatemeter Platform. The first section explains the method followed and the data employed for the realisation of the Platform. The second section illustrates in detail the functioning of the Platform and its many functionalities.

3.1. Explanation of method and data necessary for the realization of the Platform

The Hatemeter database represents the "knowledge store" of the project, since it contains all information incrementally extracted from social media (Twitter and YouTube) and harmonised for the three project languages (English, French and Italian). It contains data that have been extracted on the basis of both the requirements of the NGOs involved in the Hatemeter Partnership and on the results of the quali-quantitative data collection and analysis.

The three universities and the three NGOs involved in the project have defined a first list of keywords and hashtags to be monitored on social media, which entail anti-Muslim content (for example #banislam, #NoMoschee, #IslamHorsDEurope). These were used to query Twitter APIs (Application Programming Interfaces) and retrieve content containing at least one of the hashtags/keywords of interest over a period of time starting from the end of September up to November 2019. After an initial stage in which the same list of keywords was used both for Twitter and for YouTube, we manually filtered and refined YouTube query terms as a consequence of a qualitative evaluation of the database, during which we noted that the first list was too generic. In Hatemeter, we focus on the processing of textual data, but we collect anyway also information on posted videos (not the video itself), based on the assumption that the accompanying description and the title are a proxy for their content.

We report in Table 1 the number of hashtags and keywords on anti-Muslim hate speech being monitored for each project language on Twitter and on YouTube. For Twitter, we manually divided the hashtags into primary and secondary: the former can be found alone and belong unambiguously to anti-Islam discourse (e.g. #muslimpedophile, #Londonistan), while the latter can appear in Islamophobic tweets but they can also belong to other kinds of messages (e.g. #terroristi, #Jihad). For YouTube, we started from the Twitter list but we refined some terms and removed some others, so to make the search through YouTube APIs more specific and targeted to islamophobia.

The messages containing one of the keywords or primary hashtags are then stored in the relational database

Table 1 - Hashtags and keywords on anti-Muslim hate speech monitored for each project language on Twitter and YouTube

	Number of primary hashtags (Twitter)	Number of secondary hashtags (Twitter)	Number of keywords (Twitter)	Number of query terms or expressions (YouTube)
French	13	22	5	11
English	36	-	7	6
Italian	20	122	10	7

Source: FBK elaboration – Hatemeter Project

Table 2 - Hatemeter database content extracted from Twitter and YouTube (as of November 10 2019)

	Tweets	Replies	Retweets	N. of unique users	YouTube videos	YouTube comments
French	159,822	272,847	1,222,710	261,039	782	101,982
English	132,402	111,477	1,023,743	520,507	497	37,180
Italian	33,736	28,362	179,834	47,816	611	32,655

Source: FBK elaboration – Hatemeter Project

of the project. A summary of the database content retrieved from Twitter and YouTube so far is reported in Table 2. For Twitter, we list the number of tweets containing Islamophobic content, the number of replies, the number of retweets (showing that the content is very likely to have large visibility on the social network), the number of times the Islamophobic hashtags and/or keywords appear in such messages, and the number of Islamophobic messages that contain also links to other media content. For YouTube, the APIs are not enabled to search for content directly in the comments, but only starting from videos. Therefore, the extraction process is different from the one used for Twitter: we first search for Islamophobic hashtags or keywords in the videos (they can be mentioned in the title, description or subtitles). This is done by setting the language as either English, Italian or French. Then, we collect all comments posted below the retrieved videos.

An analysis of Table 2 shows that French is the language for which we have more data, although English language is generally more represented on Twitter and YouTube. Since this trend is confirmed both on Twitter and on YouTube, we may argue that in French Islamophobic content is particularly widespread and present in the online debate. This may also depend on the fact that the removal of offensive content, especially on YouTube, occurs more frequently and is more effective on English data than on other languages. Also, Table 2 confirms that all the project languages are well represented in the database.

Therefore, we rely not only on the language declared in the Twitter user profile but also on some linguistic information to perform a language-based selection of the tweets. This feature affected in particular Italian users, which in many cases had set “Italian” as the profile language, but tweeted in another language.[33]

3.2. How does the Platform work

The Hatemeter Platform is a web-based tool designed to support NGO operators in analysing anti-Muslim speech online. Online content is continuously monitored using text processing tools that access content related to anti-Muslim hate speech and activities through keyword-based and hashtag-based search, focusing on English, French, and Italian.

The collected tweets are analysed using text processing tools to extract the most relevant information related to anti-Muslim hatred online, such as the metadata connected to the messages (i.e., user, date, frequency), the content popularity (number of replies and retweets) 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 is then 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 to make the tool language and country-independent.

As regards YouTube, the Platform displays videos that include comments with Islamophobic remarks, ranking them by date. For this purpose, YouTube APIs are used and then state-of-the-art sentiment analysis tools are run on all comments, so that only those with a negative

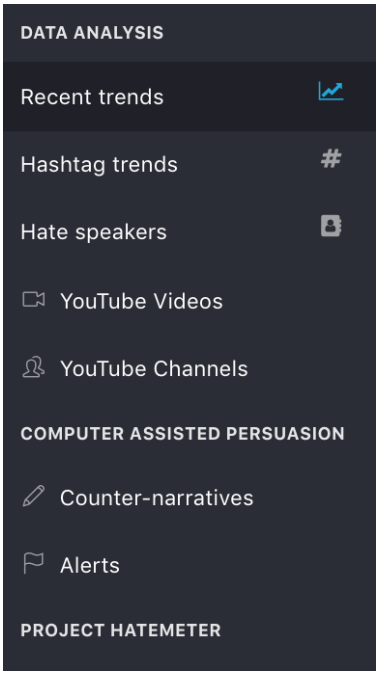
[33] Note that the data collected are intended for the internal use of the project partners and European Commission only and may not be distributed externally or reproduced for external distribution in any form without express written permission of the Scientific Coordinator of the project University of Trento and the Technical Coordinator Fondazione Bruno Kessler. During the validation of the Hatemeter Platform and national pilots, we did not process any personal data of the users.

sentiment are displayed. The Platform also displays a list of YouTube channels that have broadcasted videos related to anti-Muslim hatred, providing also the links to such content.

Each time anti-Muslim tweets or YouTube comments are detected, the corresponding posts are retrieved and stored in the project database. Then, the analytics provided in the Platform allow to better understand Islamophobic messages. As mentioned, the Platform is available in three different languages (Italian, English, French), according to which the use of the functionalities relies on different versions of the underlying database, containing material that has been leveraged following the specific NGO requirements. The analytic tools, instead, remain the same across the three versions, since they have been designed so to be language-independent.

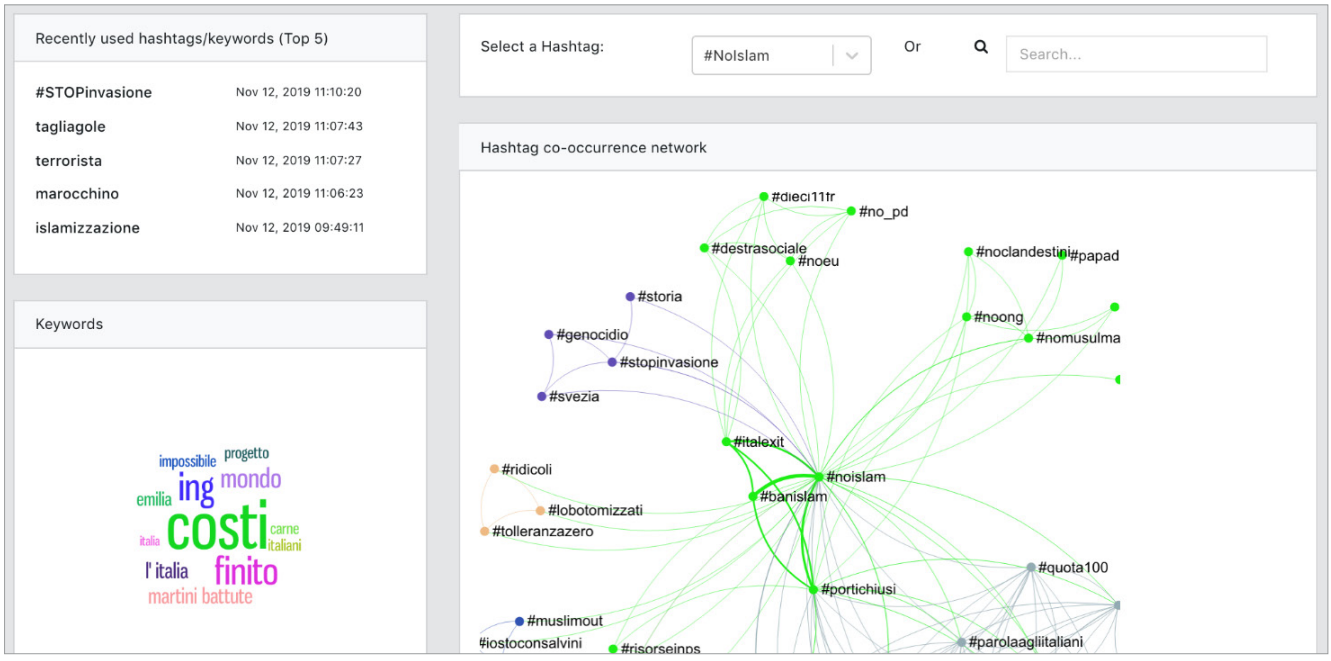
After accessing the Platform, the page displays a menu on the left (Figure 1) and the corresponding analysis outcome on the right (Figure 2). The analytics available in the Platform are those listed below under "DATA ANALYSIS", while the functionalities to support the creation of counter-narratives are displayed under the "COMPUTER ASSISTED PERSUASION" menu item. Both are described in more detail in the following sections, with a focus on YouTube videos and improved alerts.

Figure 1 – General view of the menu with all the Platform functionalities



Source: FBK and UNITRENTO elaboration – Hatemeter Project

Figure 2 – General view of the "Recent trends" tab



Source: FBK elaboration – Hatemeter Project

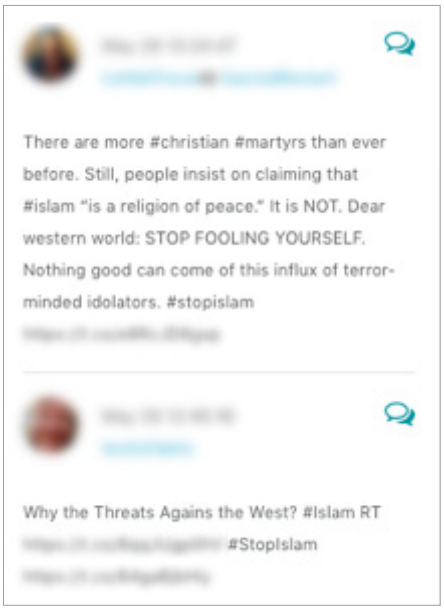
3.2.1. Data Analysis

The "DATA ANALYSIS" item presents three views: "*Recent trends*", "*Hashtag trends*", "*Hate speakers*", "*YouTube videos*" and "*YouTube channels*" (Figure 1).

The "*Recent trends*" view allows users to monitor current Twitter activities around predefined Islamophobic keywords and hashtags, or looking for search 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 (see Figure 2): 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. On the right it is possible to see the list of tweets, ranked by date, containing the hashtag or keyword selected by the user. This information was requested by NGO operators, who asked to have 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. As shown in Figure 3, 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 to obtain suggestion on possible responses. More details on this are reported in Section "Computer-Assisted Persuasion".

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 on 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 hashtag co-occurrence network and the day-by-day statistics showing the hashtag presence on Twitter over time are displayed, so that it is possible to view the message popularity day by day, and also to compare different snapshots (see Figure 4). An additional box shows the list of most retweeted messages among those collected through the query of interest.

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



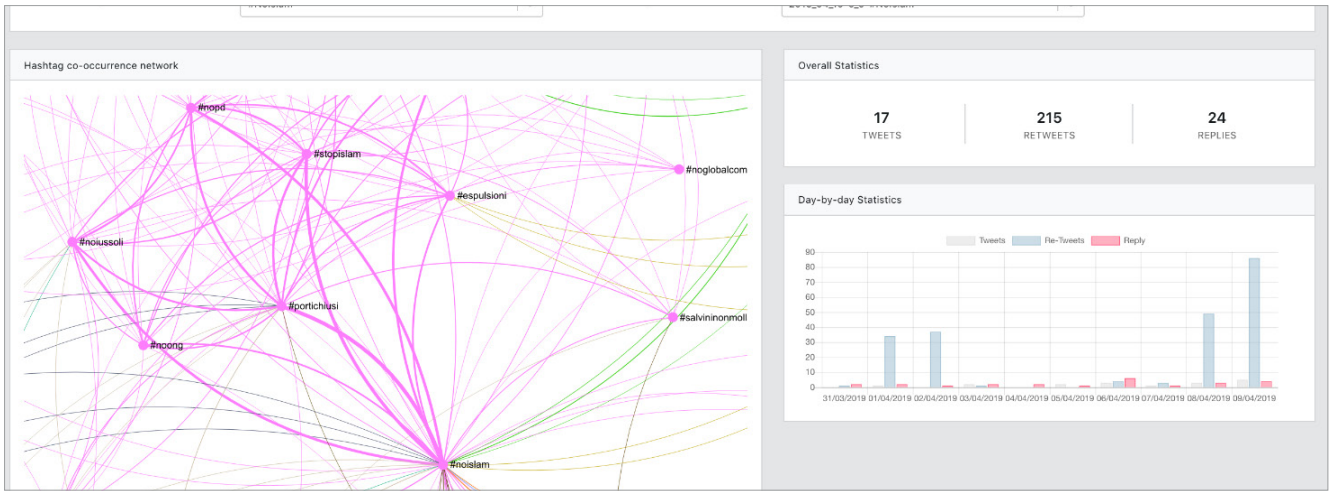
Source: FBK elaboration – Hatemeter Project

By clicking on the "*Hate speakers*" item under the "DATA ANALYSIS" menu, users are presented with the analyses related to the users that are most active in the community spreading Islamophobic messages online. After selecting a hashtag and a time snapshot, the Platform displays the network of users that posted messages containing the given hashtag (see the "User co-occurrence network", in which colours are automatically assigned by the network analysis algorithm to identify communities of users). 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.

The "YouTube Videos" tab in the Hatemeter Platform shows a list of the most relevant Islamophobic videos for each language, ranked by relevance. The videos were retrieved using as query term one of the keywords displayed in the menu on top of the window. These keywords were obtained starting from the list provided by the project domain experts and used also to collect tweets. However, it was manually modified and reduced after noticing that some keywords were too generic and were not effective in targeting Islamophobic discourse. Therefore, we revised the list and made it more specific, also through the use of multiword expressions.

For each keyword, YouTube API was queried to retrieve the videos that were most related to the keyword,

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



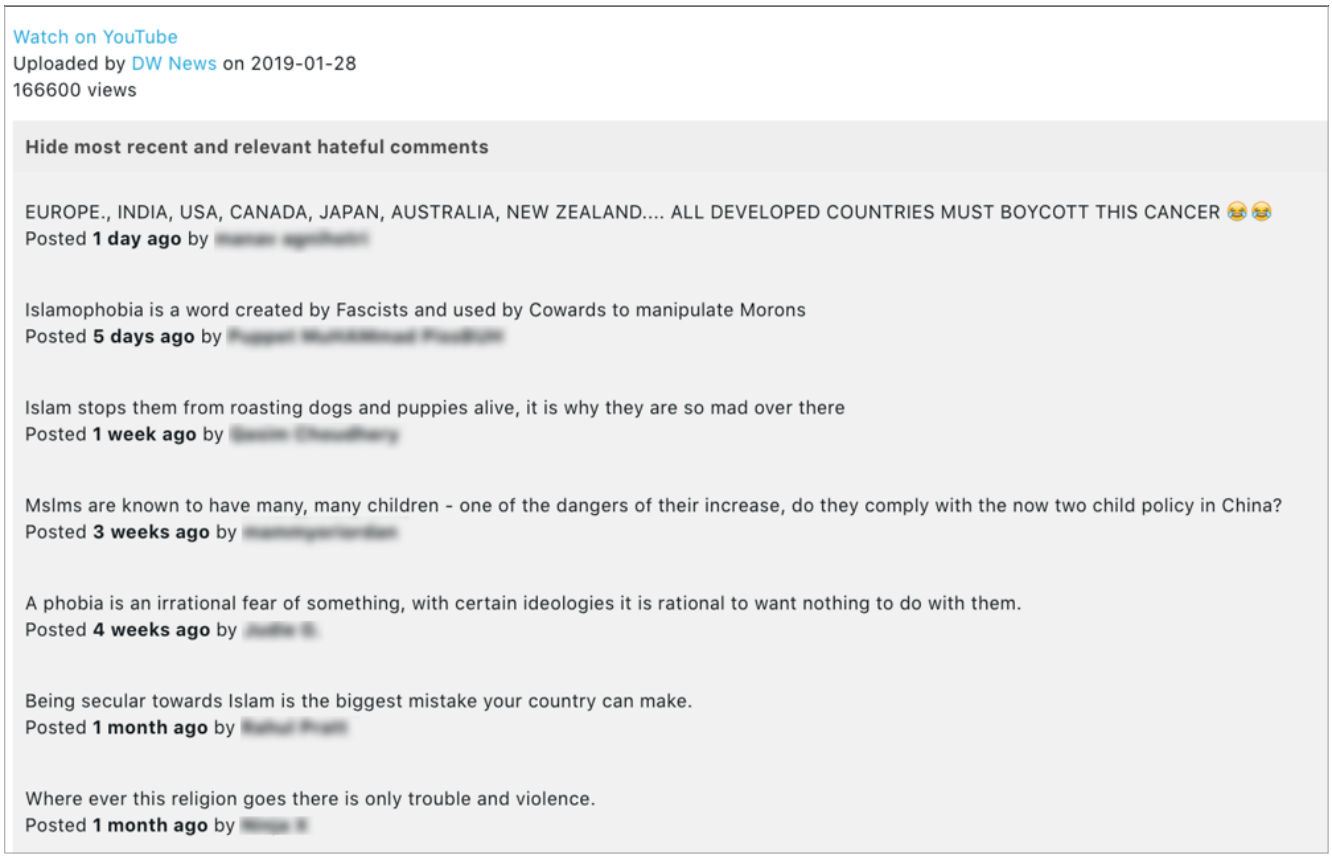
Source: FBK elaboration – Hatemeter Project

ranked by relevance. Although the algorithms used by YouTube to retrieve and rank the videos is not fully transparent, we observe that the ranked list of videos is indeed relevant to our domain. For each video, we then query the API again to collect all comments posted below it, and apply sentiment analysis to keep only the comments with a remarkably negative score. For each language, the sentiment analysis tool is different: for English we use StanfordCoreNLP (Manning *et al.*, 2014); for Italian, we build a dictionary-based sentiment analysis tool that compared the lemmas of the linked news with a list of affective terms extracted from WordNet Affect (Strapparava and Valitutti, 2004); and for French, the MeaningCloud API for sentiment analysis was used.

After running these analyses, the Platform allows users to watch the retrieved videos, which are displayed with some basic information, for example when it was posted by whom, and how many views it has. Below the video, it is possible to read the comments containing a negative sentiment, which in most cases show Islamophobic content. An example is reported in Figure 5 above. Note that, since the text analysis is completely automatic, some messages may be misclassified.

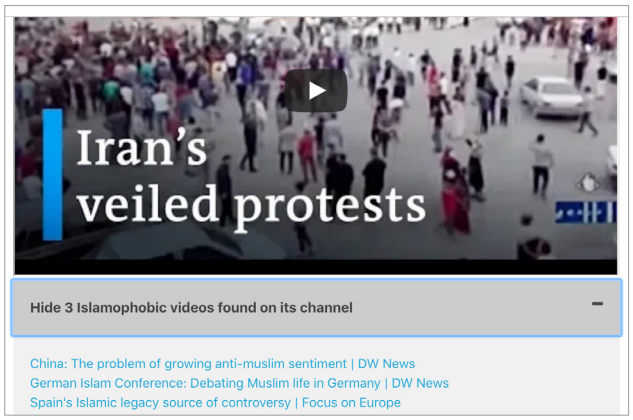
Running the queries detailed in the previous subsections led to the collection of several YouTube videos related to Islamophobic keywords. Since we observed that in some cases, several videos were broadcasted through the same channels, we added a tab to the Platform displaying a list of Platforms that have dealt with Islamophobic content. The channels are ranked according to the number of retrieved videos that are related to Islamophobic content, and that are listed when clicking on the bar "*Show N videos related to islamophobia found on this channel*". By selecting one of the videos on the list, it is possible to watch it on YouTube. This view may assist operators and researchers in discovering new sources of anti-Muslim hatred, identifying the most active channels, similar to the "*Hate speakers*" view provided for Twitter users. However, this list contains also channels of "official" news sources from traditional media that happened to show controversial videos or comments on Muslim-related events through the Platform. An example is reported in Figure 6.

Figure 5 – YouTube comments related to a video that are likely to contain Islamophobic content, ranked by date



Source: FBK elaboration – Hatemeter Project

Figure 6 – YouTube channel showing content related to islamophobia, with relevant videos listed below



Source: FBK elaboration – Hatemeter Project

3.2.2. Computer-Assisted Persuasion

The features available under the "COMPUTER-ASSISTED PERSUASION" (CAP) item are "*Counter-narratives*" and "*Alerts*".

The main tool for CAP relies on the "*Counter-narratives*" feature. This is a chatbot-like application that, given an Islamophobic short text in input, provides five suggestions that could be used to counter the hate speech or deescalate the argument. These suggestions are automatically retrieved from a DB of responses that has been collected during pilot days. A screenshot of the interface is displayed in Figure 7, where the message above is the input and those below are suggestions provided by the chatbot that could be used to build counter-narratives.

The input hate message, which triggers the generation of possible replies, can be selected by the operator according to different strategies: if in the "*Recent trends*" view, 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

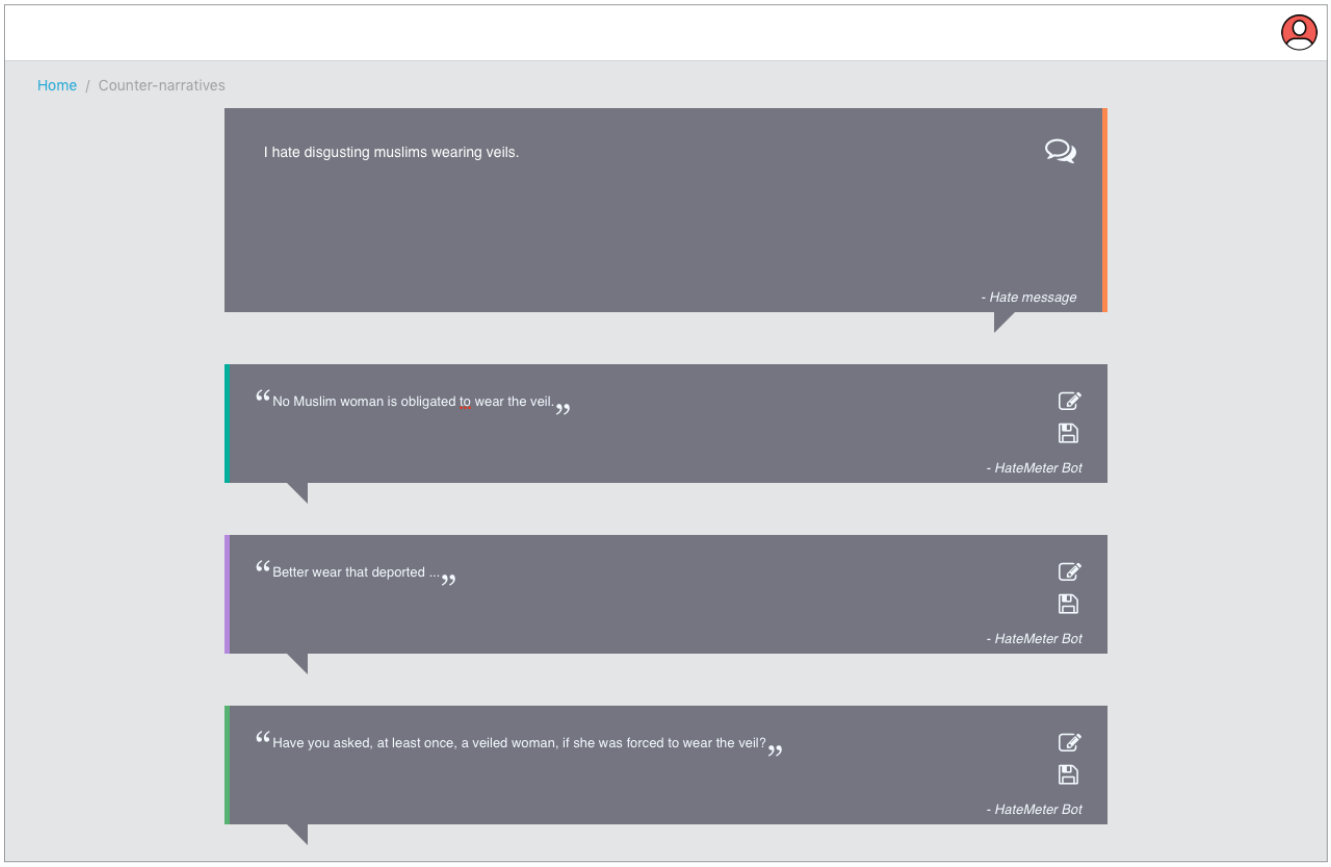
icon (see Figure 3 above). In this way, the content of the tweet is fed as input to the counter-narratives tool, which is automatically opened after clicking the speech bubble icon, and five possible answers are automatically retrieved by the chatbot-like application. As an alternative, the operator can also write an input text, and call the chatbot to provide possible replies, or copy and paste it from other sources.

After checking the answers, an operator can choose to use one of them to reply to an Islamophobic message. Therefore, each counter-narrative 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 protect operators' online identity, a different solution was implemented, requiring to copy and paste 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. We implemented a tf-idf 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 presenting in the interface the top five most relevant responses.

The "*Alerts*" view instead 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 "DATA ANALYSIS". In this view, users are asked to select one of the predefined hashtags or

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



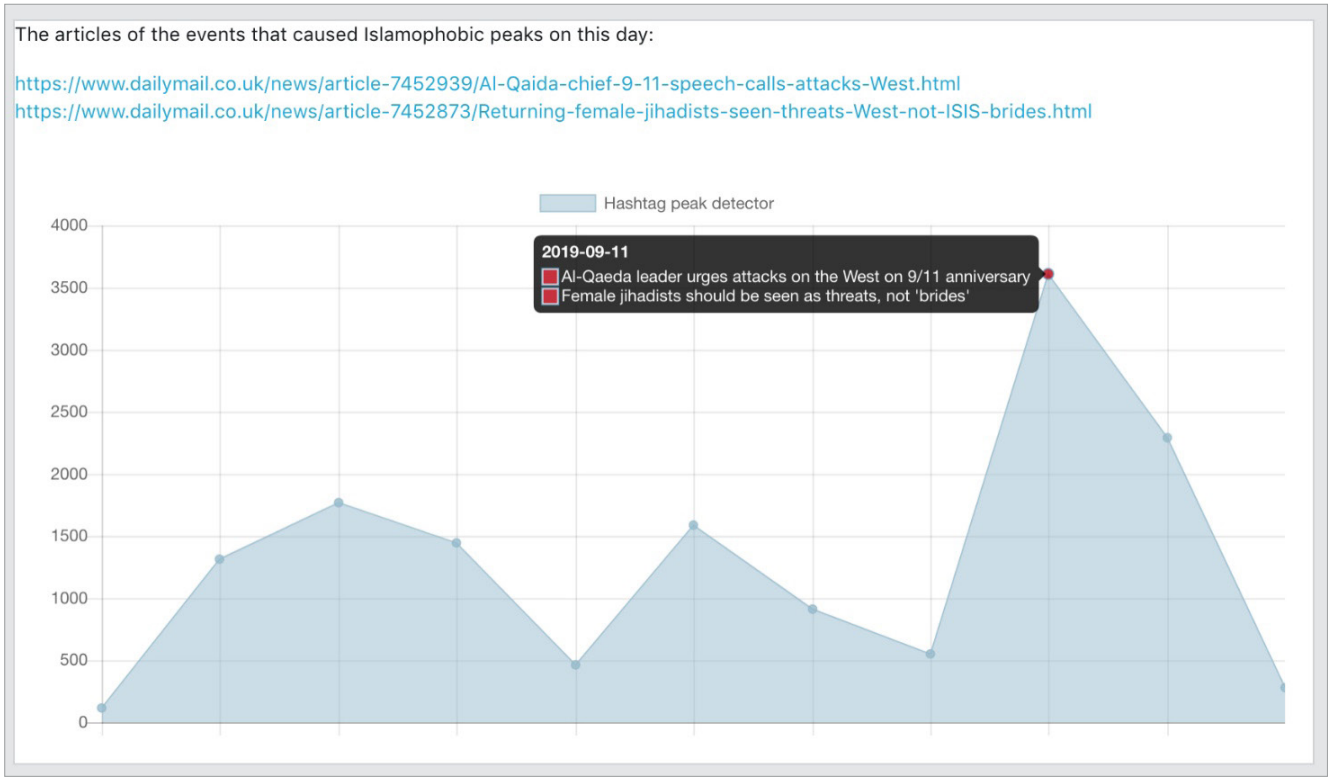
Source: FBK elaboration – Hatemeter Project

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, all aggregated. The visualisation signals peaks 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.

In order to understand why a peak was observed (i.e. finding evidence to explain the increased use of Islamophobic tweets) and to extend the range of information sources spreading Islamophobic messages (i.e. not only social media but also news from traditional media), we added to the displayed trends a list of possible news dealing with an event that may be connected with a peak in the use of an Islamophobic hashtag. We called the free version of the *Event Registry* APIs that, given a language, a date and one or more terms in input, returns a list of online news related to the query terms. We then need to select only those that deal with some negative events such as attacks, criminality, etc. that may trigger anti-Muslim prejudice. We therefore use sentiment analysis to select only the news sources with a negative score, assuming that news with a hateful note are likely to be classified as negative. While for English the sentiment analyser is already provided through the Event Registry API, for Italian and French we adopt the two solutions applied also to YouTube comments (see subsection “YouTube Videos”).

Since in some cases the Event Registry APIs do not retrieve any news sources related to the input keywords, or the corresponding sentiment score is below a certain threshold (i.e. negative enough), for some hashtag peaks the system does not display any link to external news. If instead some links are retrieved, they are presented at the top of the hashtag peak window, and can be clicked on to open the related news in a new tab. An example is displayed in Figure 8.

Figure 8 – Hashtag peak detector. Peaks in the use of Islamophobic hashtags on a given day are marked in red; if available, news related to the peak are displayed on top of the window



Source: FBK elaboration – Hatemeter Project

04

Using the Hatemeter Platform to increase knowledge of online Islamophobia and support effective counter-strategies

This chapter presents evidence of online Islamophobia in Italy, France and the United Kingdom, according to a subdivision in three subsections. For each country, are reported firstly the analyses realised by the three research teams involved in the project to increase the knowledge on online Islamophobia, and secondly the results of the use of the Platform by the operators of the three NGOs in regard to the counter-narrative tools.

4.1. Italy

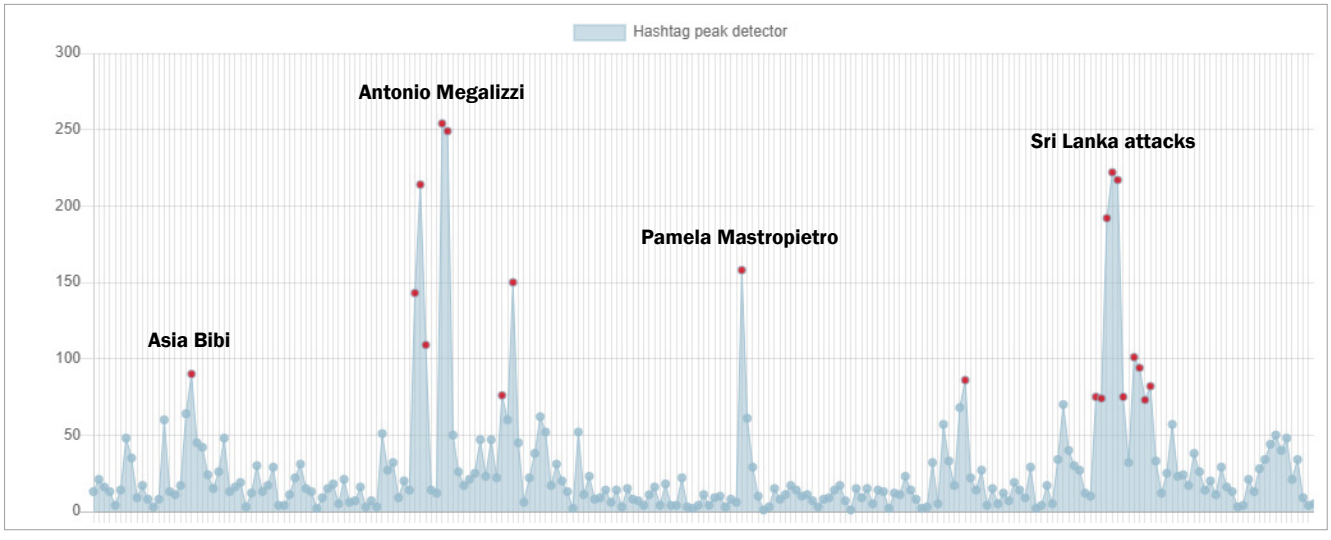
The preliminary results of the Hatemeter Project show that hate speech against Muslims on social media is increasing and closely linked to Italians' perception of migration and terrorism. Moreover, as it will be shown, the increase in the use of hashtags indicating hate speech is often connected to political debate and discourse.

Starting from a first list of keywords and hashtags entailing anti-Muslim content (for example, #StopIslam, #banislam, #NoMosques, #IslamHorsDEurope), the Hatemeter 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 9 shows four primary peaks identified using the hashtag "STOPislam", which correspond to events occurred in Italy and/or abroad:

- **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);
- **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);
- **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

Figure 9 - "Alerts" functionality, Hashtag peak detector #STOPislam in Italy



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

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);

- **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).

In order to gain a better understanding of anti-Muslim hatred online, preliminary data on Islamophobia was gathered and analysed via content analysis. As mentioned, the Hatemeter Platform has available several functionalities to create preliminary network and content analyses, primarily directed at supporting NGOs in their activities against Islamophobia and hate speech. Nevertheless, there are other analytical tools tailored for specific content and network analyses, which can be more effective for scientific research purposes. For this reason, some analyses of the Italian context have been conducted by downloading the Hatemeter database (see 2.1. Explanation of method and data necessary for the realization of the Platform) and importing it in T-lab software. T-Lab, is a linguistic and statistical software that segments and lemmatizes the imported texts, allowing researchers to conduct content analysis, co-occurrence analysis, thematic analysis and comparative analysis. The analyses were undertaken between January and April 2019.

Firstly, a content analysis was performed on a sample comprising of more than 3,000 tweets, referencing at least one of the following hashtags:

#Islamisation	#Muslims + shit	#afro-Muslims
#NoMosques	#STOPIslam	#Eurabia
#Bansharia	#BanMuslims	#StopMuslim
#NoIslamisation	#NoIslam	#Europastan
#BanIslam		

When choosing automatic settings, the list of keywords includes up to a maximum of 5,000 lexical units automatically selected by T-Lab, which belong to the category of content words: nouns, verbs, adjectives and adverbs (see Figure 10). The selection criterion uses the following algorithm: it selects the words with occurrence values higher than the minimum threshold; it computes the tf-idf or applies the chi-square test to all the crosses of each selected word for all the texts being analysed; it selects the words with the tf-idf or chi-square highest values, that is those words that, in the corpus (i.e. collection of one or more texts, selected for the analysis), make the difference. The hashtags used to extract the data from the Platform have been eliminated from the count of the keywords, because this would have resulted in them being the most numerous group. From this first descriptive analysis of the keywords, it seems that hate speech in Italy revolves around three main aspects:

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, defense, invasion*.

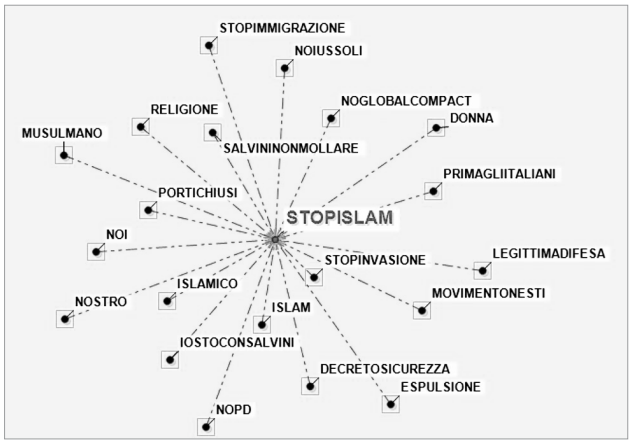
Figure 10 - Word cloud of keywords extracted by T-Lab software: results of the quantitative content analysis of tweets corpus

SOURCE: University of Trento elaboration

Secondly, the hashtag identified as most frequently used in the corpus is submitted to co-occurrence analysis. This presents a map linking a central keyword and other associated words. The most frequently used hashtag in our corpus is #Stopislam. In Figure 11, 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.

The association option shows the words in the corpus that share co-occurrence contexts with selected key-terms (#Stopislam). The selection is carried out by the computation of an Association Index (the Cosine Coefficient). The proximity to #Stopislam indicates the degree of association of each considered word. The words

Figure 11 - Co-occurence analysis for key-term #Stopislam: results of the quantitative content analysis of tweets corpus



SOURCE: University of Trento elaboration with T-Lab

considered as a whole are the context of relationships, which define the local meaning of "Stopislam" in the corpus. The results of this analysis suggest that the key-term "Stopislam" is associated to:

- Migration issues: *stop invasione*, *prima gli italiani*, *decreto sicurezza*, *no ius soli*, *stop immigrazione*, *porti chiusi*, *espulsione*, *no global compact*;[34]
- Some particular political parties or politicians of the centre-right coalition: *Salvini non mollare*, *io sto con Salvini*, *movimento onesti*.[35]

Finally, from the identified and analysed words composing the corpus, a thematic analysis is realised. Specifically, the corpus was explored, using a small number of significant thematic clusters. Each cluster consists of a set of elementary contexts (i.e. sentences and paragraphs or short texts like responses to open-ended questions) characterised by the same patterns of key words, and described in terms of their lexical units (i.e. words, lemmas or categories.

Figure 12 shows a factorial plane on which the clusters are projected, as the result of a process aimed at identifying topics that are internally homogeneous and heterogeneous among themselves. Thematic analysis identified three main themes in the analysed corpus:

- Terrorism (theme 01);
- Phobia (theme 02);
- Migration issue in political debate (theme 03).

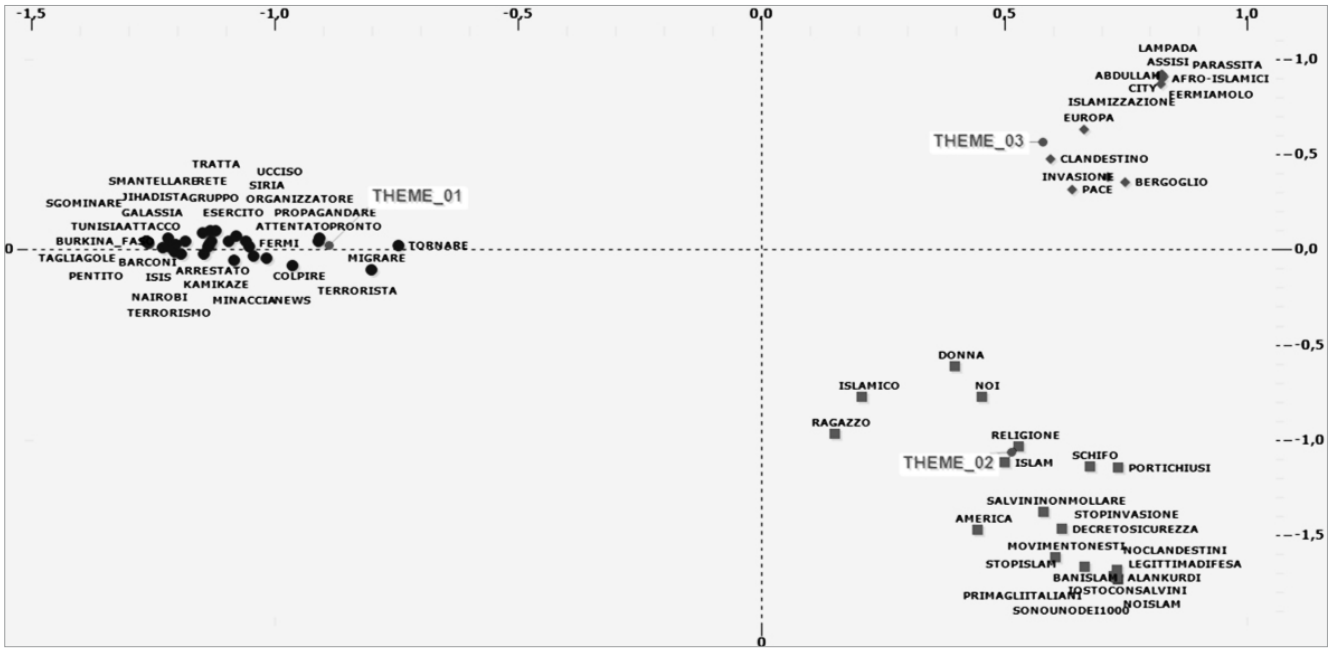
Nevertheless, the Hatemeter Platform incorporates excellent analytical tools that can similarly help in investigating the origin of hate speech. Starting from the events identified with the "*Alerts*" functionality, Hashtag peak detector the "*co-occurrence analysis*" also presents a map linking a central keyword and other associated words. By way of example, the second peak identified in Figure 9 above is here explored in detail, by using the functionalities of the Hatemeter Platform. The peak 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 13) shows that #Stopislam is associated with the hashtag "Stopinvasion", but together with #noglobalcompactformigration. Moreover, the main network has "large meshes" and it is composed of several different underlying clusters linked among them, as it is possible to notice from the different colours of the sub-networks that compose the whole network.[36] As previously underlined, these are connected respectively to immigration issues, national security and social threat. This peak refers to the period 11-21 December 2018, during which 47 tweets, 559 retweets and 49 replies were published.

[34] English translation: stopinvasion, Italians first, decree law on security, no ius soli, stop immigration, closed ports, expulsion, no global compact ("Global compact for migration" is an agreement, prepared under the auspices of the United Nations, which describes itself as covering "all dimensions of international migration in a holistic and comprehensive manner").

[35] English translation: Salvini (the Former Italian Ministry of Interior) do not give up, I support Salvini, movement of honest people.

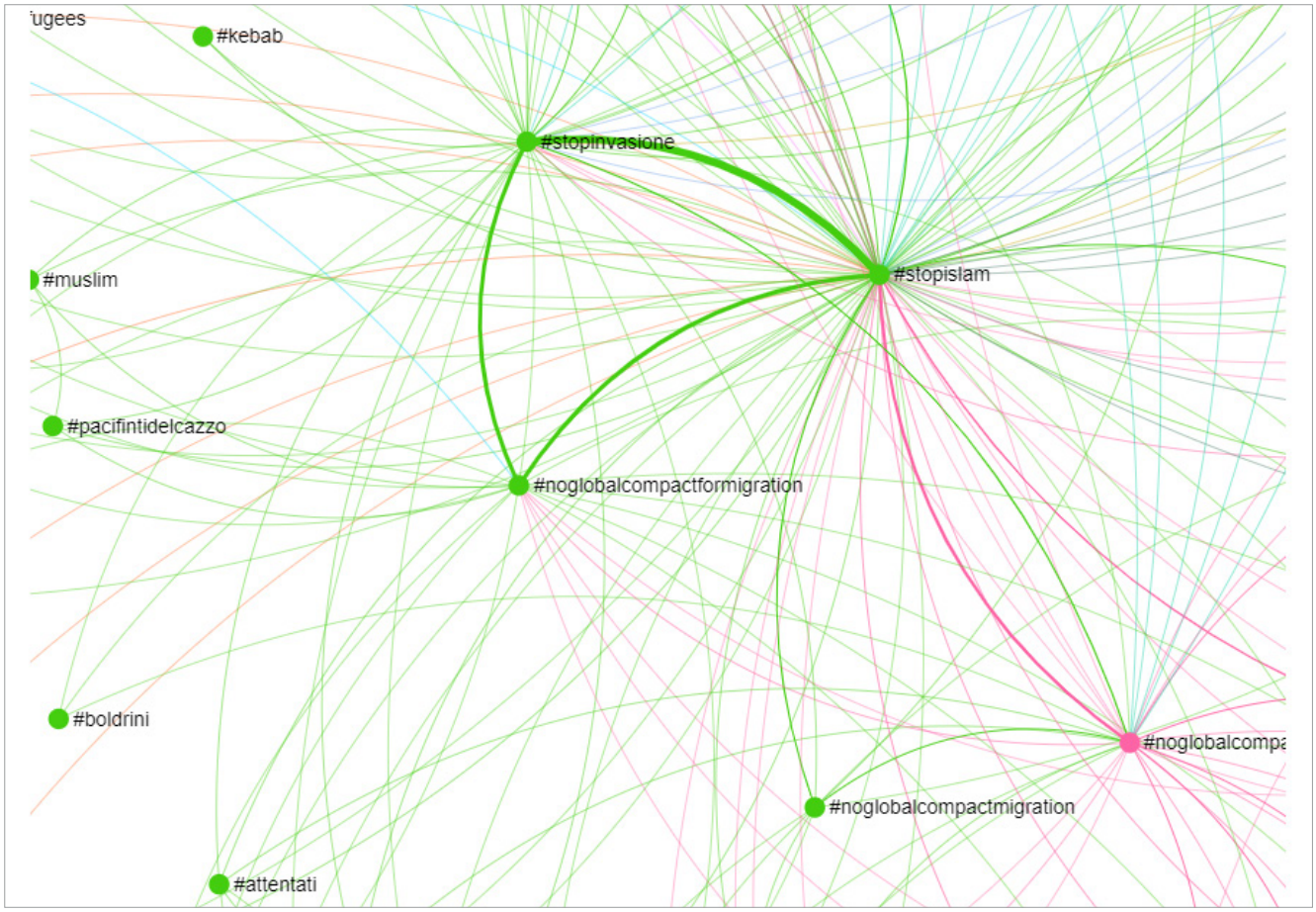
[36] The Platform relies on one of the possible algorithms usually employed to perform community detection. More specifically, the employed algorithm is "Louvain modularity", often used in social network analysis.

Figure 12 - Thematic analysis of the corpus: results of the quantitative content analysis of tweets corpus



Source: University of Trento elaboration with T-Lab

Figure 13 - #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, which have been retrieved through the Platform.

Figure 14 – First example of tweet regarding 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 15 - Second example of tweet regarding the death of Antonio Megalizzi

English translation
They also rape dead bodies. #noglobalcompactmigration #stopIslam #NOGLOBALCOMPACT

SOURCE: University of Trento elaboration – Screenshot from Twitter

Figure 16 - Third example of tweet regarding 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

In the Italian context, Amnesty International Italy also employed the Hatemeter Platform. In November 2018, the NGO started working jointly with the Hatemeter Project focusing on the fight against online Islamophobia and supported the creation of the Hatemeter Platform by providing examples of hate speech and Islamophobic discourse followed by counter-narrative suggestions.

During the Hatemeter piloting sessions from January to March 2019 and from June to September 2019, as well as in May 2019, Amnesty International Italy collected, through specific algorithms, more than 4 million pieces of content. Over 180 trained activists assessed 100,000 pieces of content, with the aim of detecting possible correlations between tone/spirit and the political rhetoric of politicians, and sentiments of social media users toward specific topics and groups of people. On this occasion, the Italian NGO utilised, among other tools, the Hatemeter Platform, both for the collection of hateful content and the rapid generation of alternative replies to messages posted on Twitter. The use of the Platform shed light upon the rise of hateful messages directed at religious minorities where these are connected to “terrorism” (Islam).

Starting from the piloting sessions of the Hatemeter Platform and simultaneous activities of Amnesty[37], it has emerged that the adjective "Islamic" is frequently associated with other terms, such as: "Islamic extremism", "Islamic terrorism", "Islamic fundamentalism", "problem", "immigration", and "danger" (Faloppa, 2019). Anti-Muslim sentiments find their roots not only in the connections between Islam and "invasion", "terrorism", and "barbarities", but also in the idea that Islam presents an obstacle to the advancement of the feminist and LGBTI movements. Moreover, these hateful messages are more likely to have recourse to acrimony and personal attacks.

Such monitoring activities confirm some trends already registered with an online monitoring realised by Amnesty International Italy in 2018 (i.e. during the political campaign). Even if politicians seem to maintain more moderate tones/spirit in comparison to their followers, candidates to the European elections wrote almost one in five tweets that were considered negative because of discriminatory content directed at minor religious groups in Europe (Vitullo, 2019). The Muslim community is confirmed as the religious group most targeted by political discourse in Italy (Vitullo, 2019).

The following two examples show the use of the "*Counter-narratives*" tool of the Hatemeter Platform in the Italian context. The concept of "invasion" is frequently cited and is also referenced within the political debate. As shown by the examples, the counter-narrative functionality can assist in responding to these tweets, by providing suggestions and insights to NGOs operators and thereby accelerate the response.

In both cases, the counter-narratives functionality suggests that users reflect upon the concept of "invasion", primarily by providing objective data on the presence of Muslims in Italy (i.e. the Platform affirms that Muslims represent between 4 and 5% of the total population). The aim is to present a basis for reflection, which may then curtail the spread of the hate speech. Nevertheless, it can happen that not all sentences are always suitable answers to the tweet and for this reason, NGO operators should read the counter-narratives carefully to choose the most efficient one.

Example 1_ITALY: The "counter-narratives" Functionality of the Hatemeter Platform (English translation)

Tweet: The African and Islamic invasion that the European Union is imposing upon us will be the end of the Western world.
First counter-narrative suggestion: Where do you get these conclusions about this alleged 'invasion' from?
Second counter-narrative suggestion: In Italy, Muslims are less than 5%, personally I would not define this an invasion.
Third counter-narrative suggestion: I do not think so. Muslim population in Italy represents only the 4% of the total population. I would not define this an invasion.

Example 2_ITALY: Hate tweet and answers provided by the "counter-narratives" Functionality of Hatemeter Platform (English translation)

Tweet: We must end this African and Islamic invasion… they must stay in Africa… we do not need them. They are dangerous for us and all Europe.
First counter-narrative suggestion: Where do you get these conclusions about this alleged 'invasion' from?
Second counter-narrative suggestion: In Italy, Muslims are less than 5%, personally I would not define this an invasion.
Third counter-narrative suggestion: I do not think so. Muslim population in Italy represents only the 4% of the total population. I would not define this an invasion.
Fourth counter-narrative suggestion: Hello, could you please explain more clearly what you mean by Islamic invasion? And can you provide me with data to support your statement?

[37] See Amnesty (2018).

4.2. France

The data collected for the French part of the Hatemeter 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. CCIF). However, during the development of the project, new hashtags and keywords were 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.

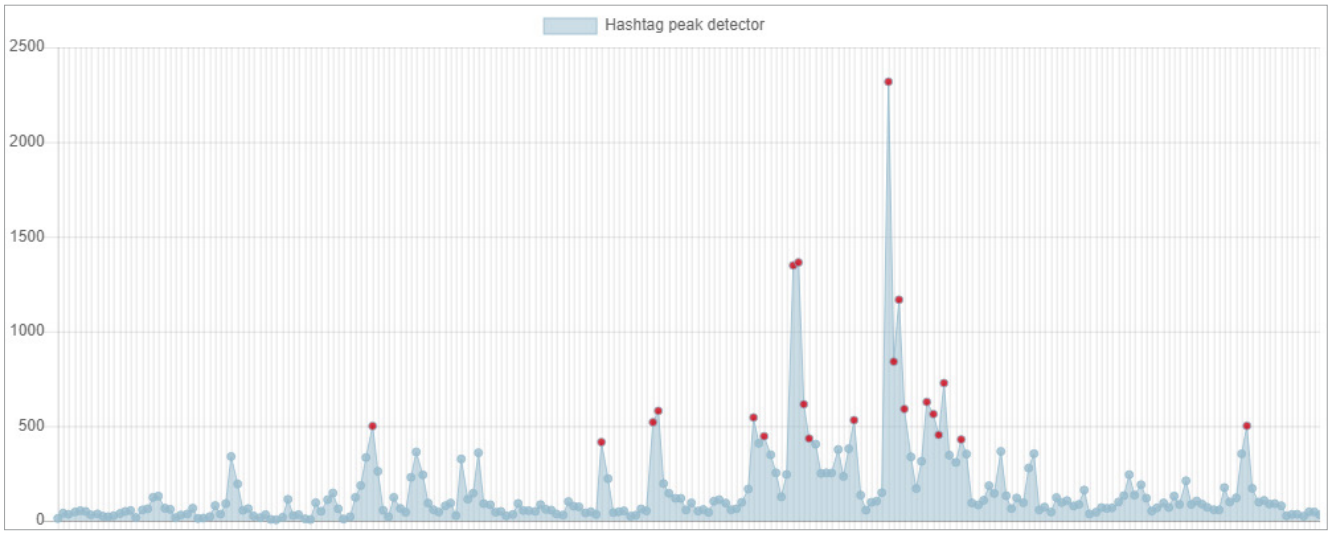
Some of these hashtags and keywords are those employed by hate speakers, whereas others are used by people talking about anti-Muslim hate itself or the place of Muslims in society. Both are useful as they help us to analyse the event in relation with the evolution of the online hate speech.

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?*

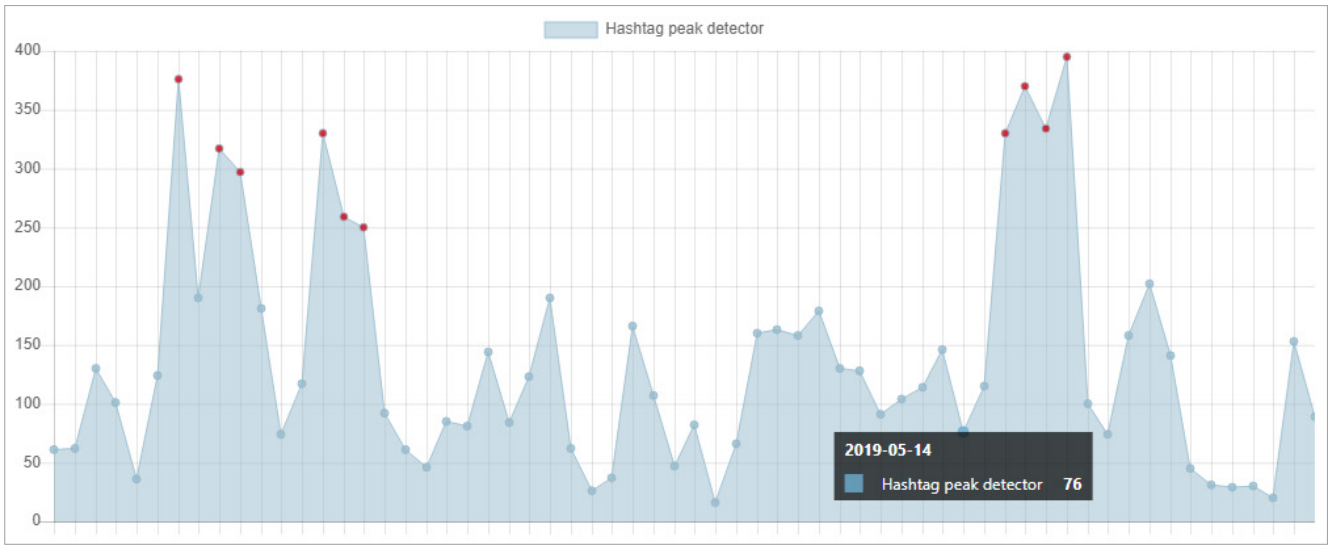
Figure 17 shows the example of a peak identified employing the "Alert" tool of the Platform and the hashtag "Islamophobie" (i.e. Islamophobia), over the period from October 2018 to March 2019. #Islamophobie is widely used, probably because not only extremist movements employ it, but also activists denouncing discrimination and people writing ironic tweets complaining about the victimization of Muslims. The 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 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

Figure 17 - "*Alerts*" functionality, Hashtag peak detector #Islamophobie in France



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

Figure 18 - "*Alerts*" functionality, Hashtag peak detector #GrandRemplacement in France



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

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.[38] 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.

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

Looking at the graph generated for #GrandRemplacement in Figure 18, it is possible to 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 19, 20 and 21 show the employment of #GrandRemplacement for other very different topics.

Figure 19 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 ironic 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.

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

Figure 19 - 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 #GrandReplacement (see Figure 20) 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 20 - 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 21), 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 21 - 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 Hatemeter Platform

The Hatemeter Platform allows the French NGOs Collectif contre l'islamophobie en France (CCIF) to gather evidence of online Islamophobia. For example, a Twitter user misrepresented the concept of French "*laïcité*" on his twitter feed to promote an interview he gave. In the following tweet displayed by Figure 22, gathered thanks to the Hatemeter tool through the *"Recent trends"* functionality, he completely negated the reality of Islamophobia.

Utilising the Hatemeter Platform, any NGO that wishes to prove the usage of #laïcité in an Islamophobic context, can search the hashtag "*laïcité*" and instantly browse through all the most recent tweets that used that key word, and see the links made with other topics, usually indicating Islamophobic discourse. For example, here we can see that *laïcité* was linked to "*musulmans*" (Muslims) and "voile" (headscarf).

The Platform allows analysis of the most recent trends and use of semantics by Islamophobic accounts and tweets. However, in our findings, we can often deduce other conclusions than the use of specific words of justification.

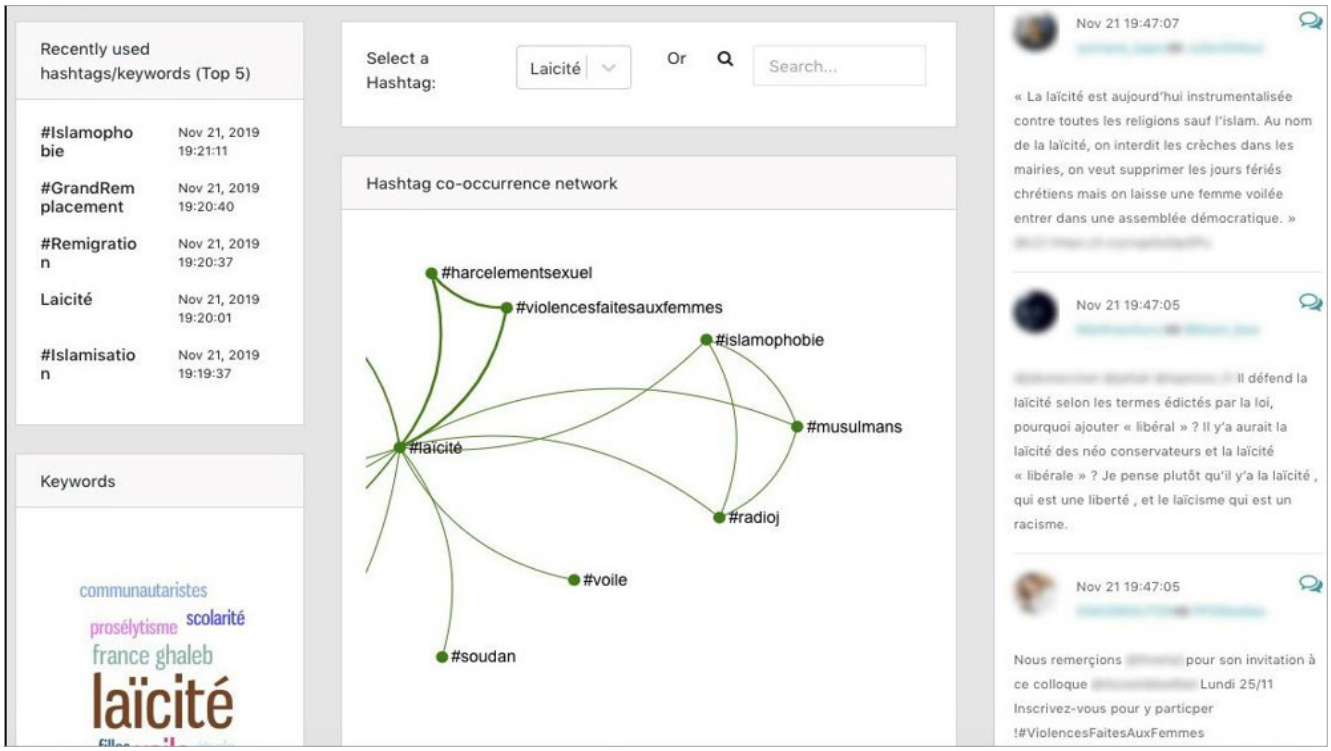
Moreover, the "*Hate speakers*" functionality facilitates delving deeper into a specific account to analyse its influence or any other specific reason for the data analysis. Here, it is possible to research a specific topic and see all the accounts that gained the most traction due to the use of a hashtag. For instance, by searching #GrandRemplacement in the most recent timeframe the Platform offers, one of the twitter accounts that mostly use this hashtag is highlighted (see Figure 23).

By clicking on the user's account, the NGO can realise the impact this Twitter user might have. Here, the user has 161 connections to #GrandRemplacement. Their tweets range from declaring a "cultural war" to protesting in the streets against Islam with extremist groups like Génération Identitaire.

The Platform offers the possibility of searching for keywords with the use of timeframes; this allows linking the rise of a specific point of view to dated events. In France, activists are aware that elections are often accompanied by a rise of Islamophobic sentiment. However, thanks to the data that can be collected by the Hatemeter Project, NGOs can now evidence the online effects of political statements. This will be crucial for the NGOs' work in fighting Islamophobia, because they can acquire greater legitimacy by providing facts through data analysis.

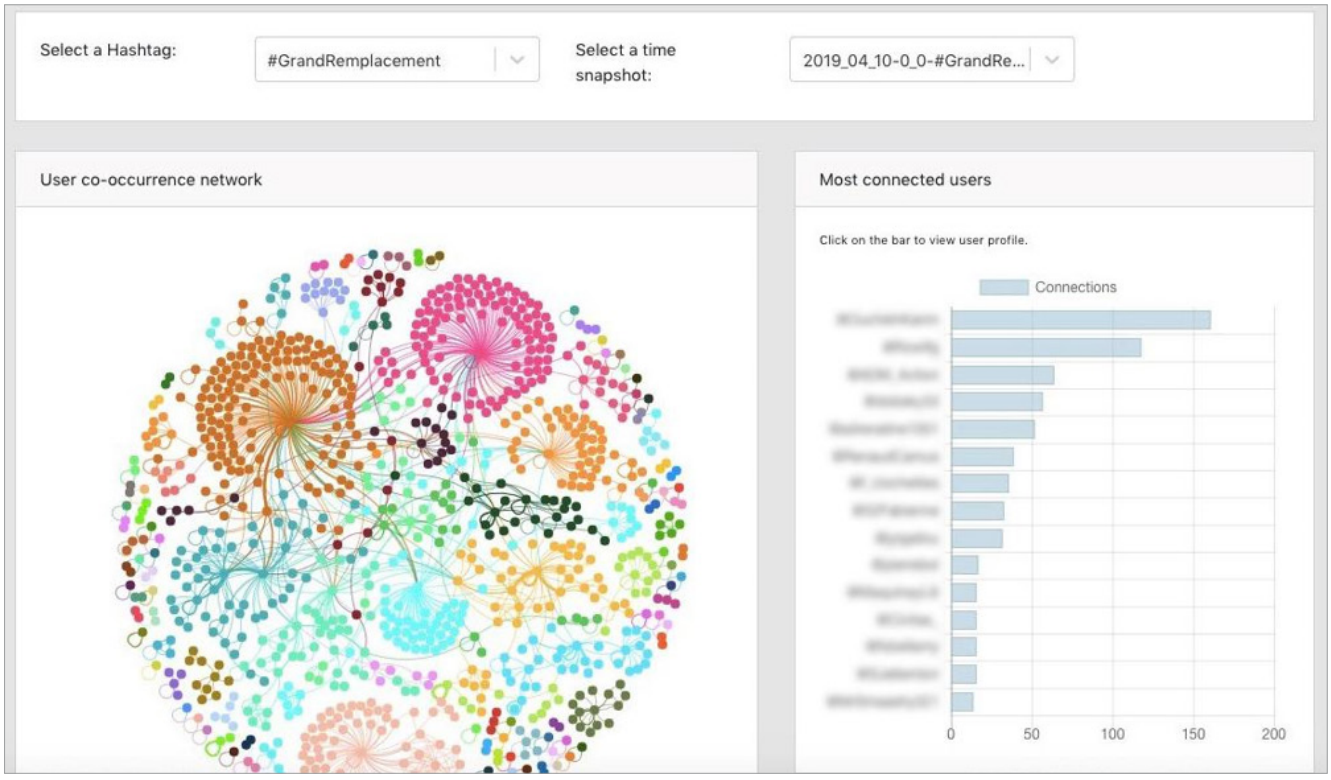
Furthermore, the cumulative use of these timeframes and keywords after televised events allow us to see the online response to political debates. For instance, by using the "*Hashtag Trends*" functionality with #Islamisation from 8th to 19th November 2019, we could analyse, compare, and draw conclusions from overall and day-by-day statistics (see Figure 24).

Figure 22 - "*Recent Trends*" functionality, utilising #Laïcité



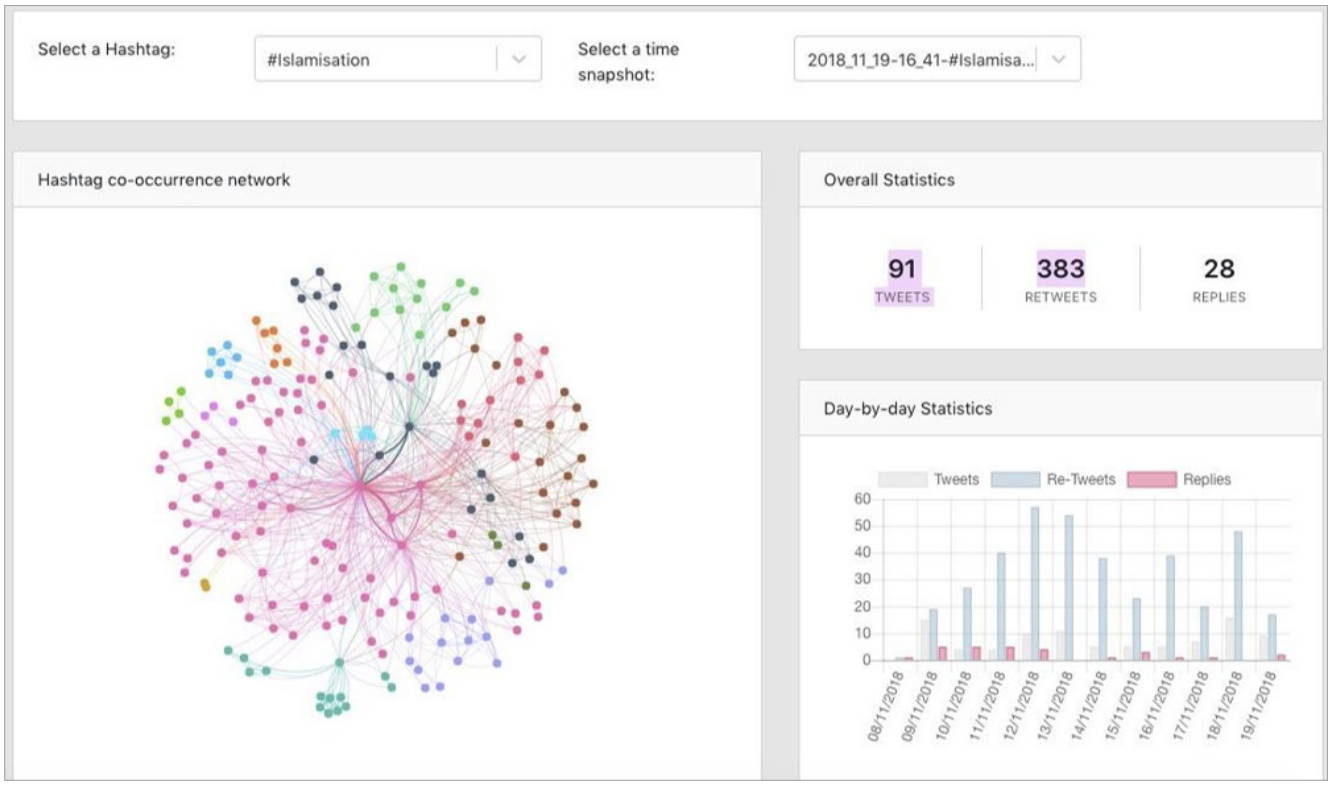
Source: CCIF elaboration – Screenshot from the Hatemeter Platform

Figure 23 - "*Hate speakers*" functionality, utilising #GrandRemplacement



Source: CCIF elaboration – Screenshot from the Hatemeter Platform

Figure 24 - "*Hashtag Trends*" functionality, utilising #Islamisation



Source: CCIF elaboration – Screenshot from the Hatemeter Platform

The analytical use of trends and time frames can also be used as a tool to see the tone of different discourses on social media after terrorist attacks. How does public opinion of Muslims change after a terrorist attack committed by someone in the "name of Islam"? How do people respond to online Islamophobia? What are the numbers on each side? Hopefully, these questions will be able to be answered thanks to the Platform. Furthermore, in addition to answering these questions, the Hatemeter Platform will provide NGOs with enough quantifiable data to hopefully prevent such hate speech: because the same key words and arguments appear repeatedly, thanks to the "Counter-narratives" functionality, activists will be able to provide ready-made responses to online hate speech. These counter-narratives will cover as many topics as possible, and are created to save time for activists, preventing them from having to repeat themselves. As Example 1 below shows, any operator of the Platform can answer multiple points raised by a form of hate speech with a click of a button, thanks to the counter-narrative tool. Platform users (e.g. NGO operators) can copy and paste a tweet into the Hatemeter tool for counter-narratives, and have available multiple potential answers from which to select a reply.

Example 1 provides an example of the types of answer the Platform can offer in response to a hate-tweet (reported in the first rectangle).

Example 1_FRANCE: "Counter-narratives" functionality of the Hatemeter Platform (English translation)

Tweet: The veil is a symbol of submission, political Islam invades France.
First counter-narrative suggestion: Islam is a religion not a political party. The veil is a fabric not a symbol.
Second counter-narrative suggestion: Islam is a religion. The veil is a sign of devotion and not a political flag.

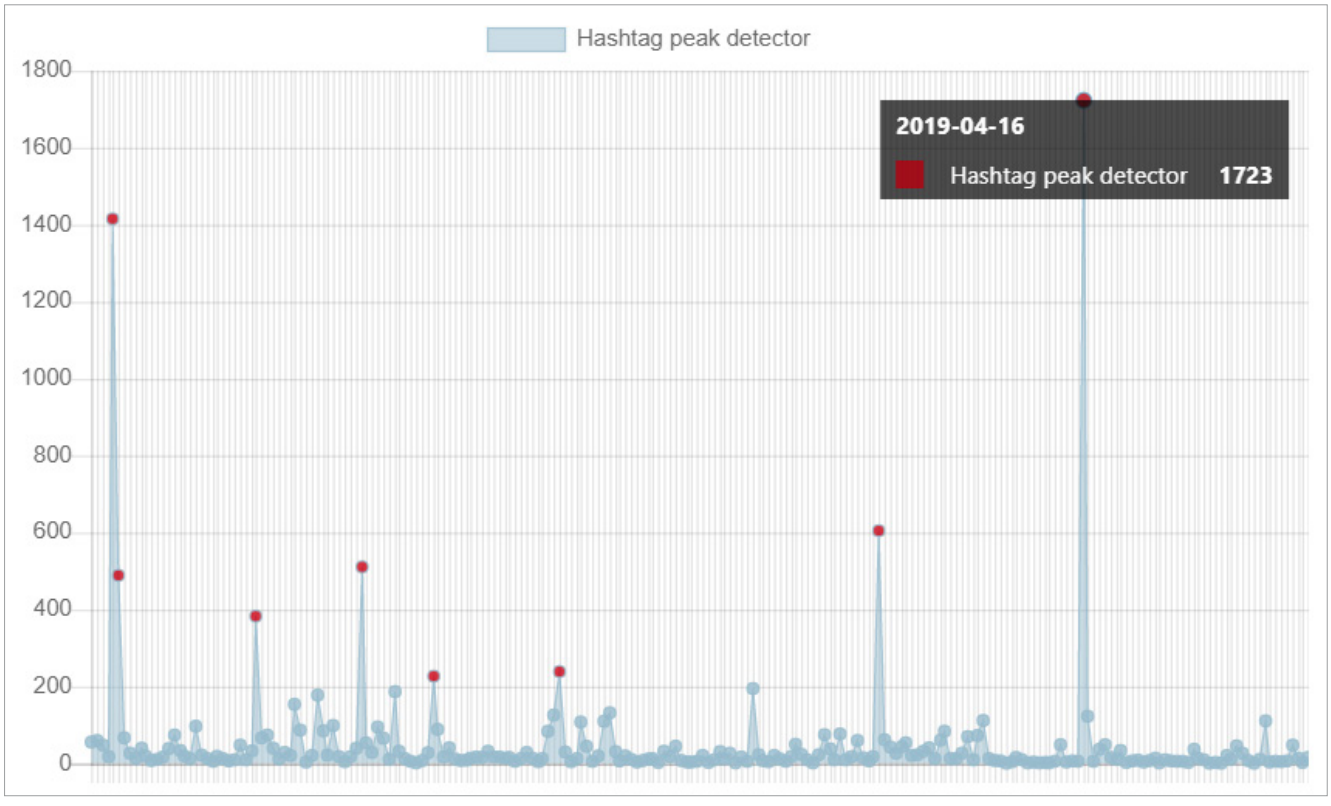
4.3. United Kingdom

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 (Sadique *et al.*, 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.

Figure 25 displays the signals peaks resulting from the selection of #RapeJihad, which appears 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.

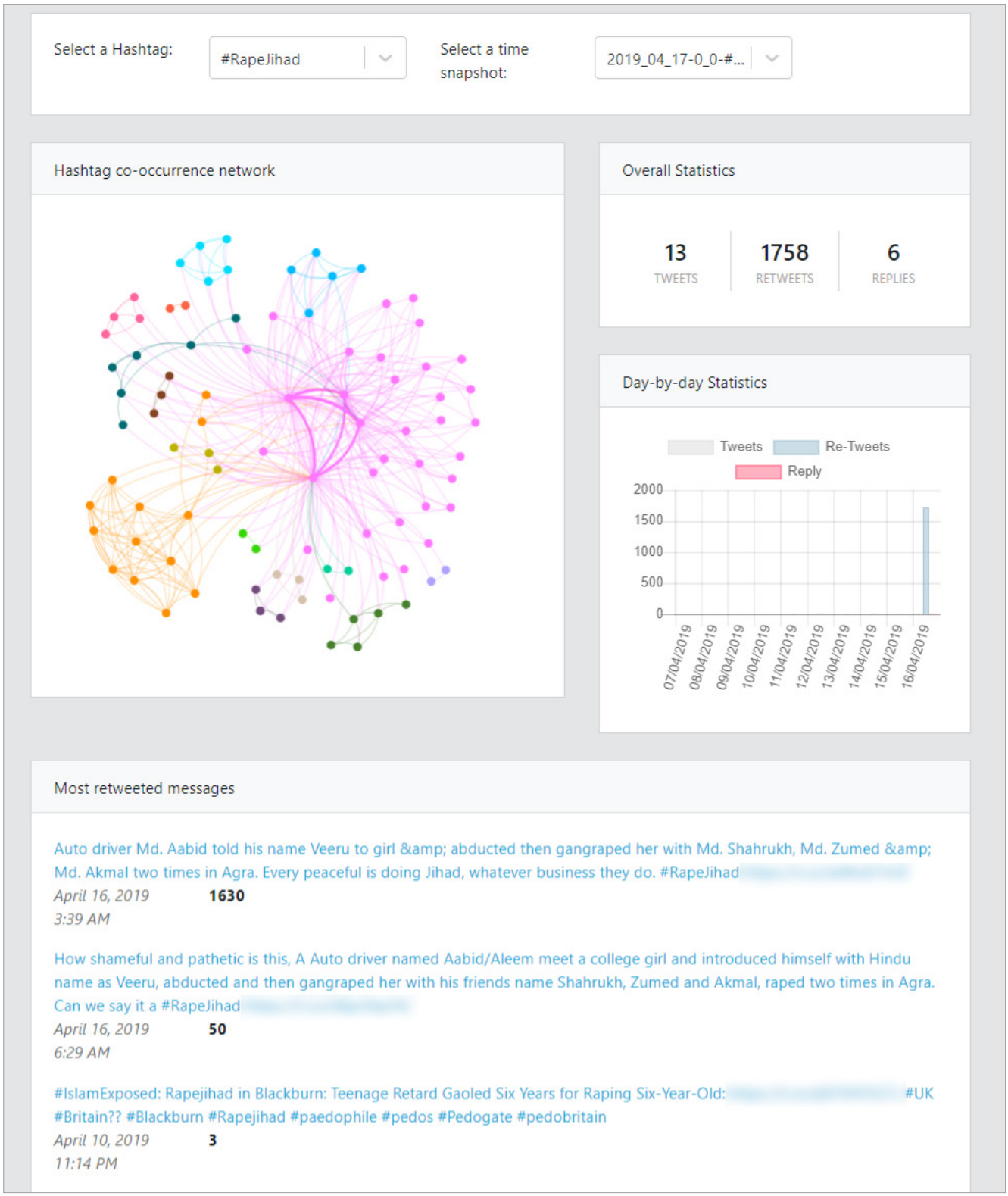
On the 16th of April 2019, a high peak was recorded concerning the #RapeJihad hashtag. Figure 34 presents a visualisation of this hashtag, a day later on the 17th of April 2019. As is evident, the *Overall Statistics* and *Day-by-day Statistics* boxes demonstrate that of 13 tweets containing the #RapeJihad hashtag, there have been 1758 re-tweets, indicating active and robust dissemination activity. As Figure 26 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 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 25 – "*Alerts*" functionality, Hashtag peak detector #Rapejihad in the UK



Source: Teesside University elaboration – Screenshot from the Hatemeter Platform

Figure 26 – "*Hashtag Trends*" functionality, utilising #RapeJihad



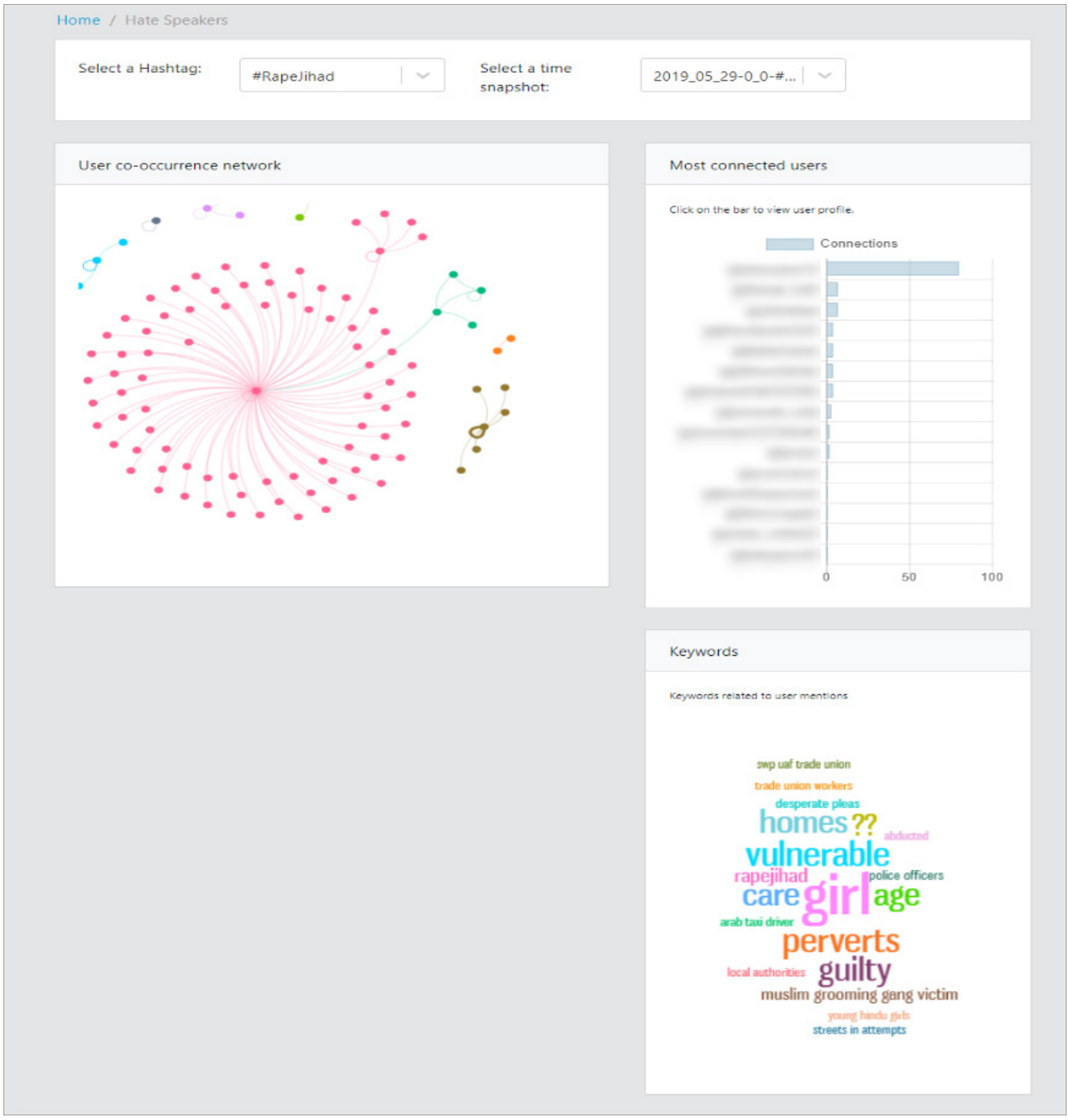
Source: Teesside University elaboration – Screenshot from the Hatemeter Platform

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"* functionality, 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 32 below demonstrates, the #RapeJihad 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, 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, 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 27, 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 27 - "*Hate speakers*" functionality, utilising #RapeJihad



Source: Teesside University elaboration – Screenshot from the Hatemeter Platform

In the UK context, Stop Hate UK's usage of the Hatemeter Platform has produced positive results in terms of assisting new volunteers/staff members to quickly and accurately identify sources of Islamophobic hate speech on Twitter. Specifically, operators have been able to observe various user networks, including central users and seed accounts, with the tool enabling NGO operators to discern who these users are within a network, not just 'infamous' accounts acting as central nodes, but also unknown accounts and consistent accounts across time and space. Stop Hate UK's utilisation of the Counter-message generation aspect of the

Platform clearly demonstrated the value of the responses, as a 'catalyst' or starting point for the generation of a counter-message, (which clearly aids speed and efficiency in the generation of a response) and their major potential as a training tool.

The evidence below is a set of separate instances, in which the Hatemeter Platform, and its various functions, have been utilised to identify and combat hate speech pertaining to online Islamophobia. Where possible, the outcome of engagement has also been included.

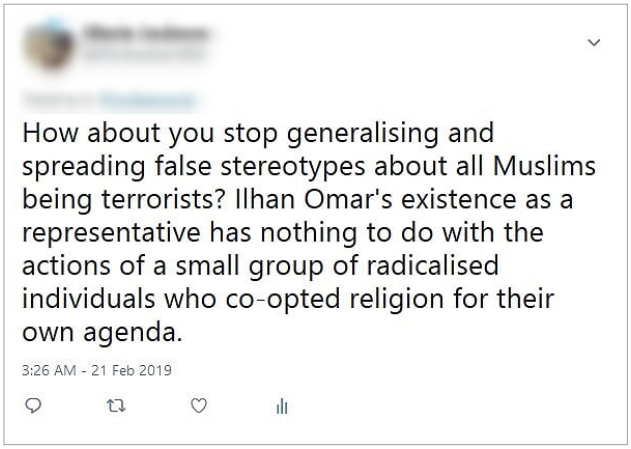
In the first example (Figure 28) utilising the hashtag search function, an account has been located and subsequently reported for Islamophobic post, directed at US Member of Congress Ilhan Omar.

Figure 28 – First example of tweet regarding online Islamophobia

Source: Stop Hate UK's research elaboration – Screenshot from Twitter

In another case, a Twitter user posted a picture which showed the twin towers during 9/11 with the caption '"never forget" they said', and then underneath, a picture of democrat representative Ilhan Omar with the caption 'this is proof that we have forgotten'. This was an early action by a staff member who neglected to screenshot the post they were reacting to. We include it as an example of use of the Platform to identify a 'hateful account' and identify a potential target of counter-messaging activity. NGO operator replied to the image in the form of counter-messaging, as shown in Figure 29. However, within 20 minutes the user had blocked the NGO operator's account.

Figure 29 – Second example of tweet regarding online Islamophobia, example of counter-narrative by an NGO's operator



Source: Stop Hate UK's research elaboration – Screenshot from Twitter

In the third and last instance reported, a Twitter user replied to an image an NGO operator had previously countered, which was posted again separately by another user. The image is of the twin towers during 9/11 and claims Americans said they would 'never forget', but insinuates that the election of democrat representative Ilhan Omar is an insult to this as she is Muslim. The user tweeted "'her face and bending the rules for her hijab makes me angry'" (see Figure 35). While this example is not in itself sourced directly via usage of the Hatemeter Platform, it is directly related to the previous example (the second one above) and demonstrates the practitioner refining/adapting an earlier counter-message. As such, it illustrates the stark difference between 'live interaction' with an account holder, and the safer practice of responding during a training or data-gathering exercise, and therefore the need to place emphasis upon support and supervision of staff and volunteers to address the cumulative impact of such interactions. NGO operator counter-messaged and this provoked the user to send to the operator a string of angry responses (Figure 30). From this interaction, it appears that overtly pointing out racism/Islamophobia can be quite inflammatory, but it may have been useful for other Twitter users to see their narrative exposed.

Figure 30 – Third example of tweet regarding online Islamophobia

Source: Stop Hate UK's research elaboration – Screenshot from Twitter

In general, it is salient to note that, during the detection of hate speech online thanks to Hatemeter Platform, there were positive outcomes by Stop Hate UK's operators, both in regards to Twitter bans and/or account suspensions, and to the removal of inflammatory posts produced and disseminated by hate accounts. NGO operators highly appreciated the ability of the Hatemeter Platform to showcase re-tweets of seed accounts – main anti-Muslim hate accounts – and they felt able to make inferences, having prompts to analyse the meanings behind tweets, which allows them to feel out where and what content to counter in the online fora. Such action strongly suggests that with repeated, prolonged usage of the Hatemeter Platform and its various functions, the Hatemeter Platform can assist significantly in actively identifying potent sources of hate speech in terms of main accounts and highly incendiary posts, and aid in attempting to secure Twitter bans and/or suspensions.

05

Assessment of the Hatemeter Platform

This chapter describes the main results of the evaluation of the Hatemeter Platform. Firstly, it illustrates the validation strategy and the evaluation methodology underlying the assessment, and secondly it demonstrates the results concerning some of the KPIs aimed at measuring the efficacy and efficiency of the Platform and the analysis of the scale of the User Experience Questionnaire (UEQ scales).

5.1. Validation strategy and evaluation methodology

To validate the project, it has been necessary to measure the following expected outcomes: i) efficiency and effectiveness of the Hatemeter Platform; ii) NGOs' capabilities in running the Platform smoothly. In addition, the implemented approach has been evaluated by: iii) feeding the Hatemeter Platform to the NGOs involved; and iv) validating it both for its innovative value and for its usability and quality of experience.

In order to make the project objectives measurable and to validate the project's achievement, Hatemeter proposed two phases of Platform testing within the three NGOs involved in the Partnership. This enabled the investigation of different aspects of the challenge inherent in refining the Hatemeter Platform in countries characterised by different languages and by different attitudes of citizens towards the Muslim community. In this way, was possible to validate the effectiveness of the project results in contexts that differ on the number and heterogeneity of citizens and their social and cultural background.

In order to evaluate the success of the pilots of the Hatemeter Platform the User Experience Questionnaire (UEQ) tool was used. More specifically, the UEQ is a measuring instrument for evaluating the subjective experience of users of interactive tools. The questionnaire is designed in a format that allows users to instantly express the feelings, impressions and attitudes they experience when using a product. The scales of the UEQ cover a wide range of user experience, measuring classical usability aspects (efficiency, perspicuity, dependability) as well as user experience aspects (originality, stimulation) (User Experience Questionnaire, 2017).

The items have the form of a semantic differential, i.e. each item is represented by two terms with opposite meanings. The order of the terms is randomized per item, i.e. half of the items of a scale start with the positive term and the other half of the items start with the negative two term. The scale uses a seven-stage scale to reduce the well-known central tendency bias for such types of items. An example of an item is:

attractive ○ ○ ○ ○ ○ ○ ○ unattractive

The items are scaled from -3 to +3. Thus, -3 represents the most negative answer, 0 a neutral answer, and +3 the most positive answer.

The UEQ contain 26 items as follows (Schrepp, Hinderks, Thomaschewski, 2017):

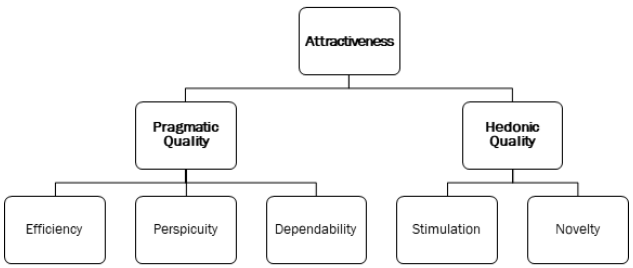
- **Attractiveness**: General impression of the product. Is it liked or disliked by users? Do they perceive it as attractive, enjoyable or pleasing? 6 items: annoying / enjoyable, good / bad, unlikable / pleasing, unpleasant / pleasant, attractive / unattractive, friendly / unfriendly.
- **Perspicuity**: How easy is it to get to know the product? Can it be easily learnt and understood? Is it clear? 4 items: not understandable / understandable, easy to learn / difficult to learn, complicated / easy, clear / confusing.
- **Efficiency**: Do users have to make an effort to solve their tasks? How efficient and fast is the interaction? How immediate is the product's response to user input? 4 items: fast / slow, inefficient / efficient, impractical / practical, organized / cluttered.

- **Dependability**: Does the interaction allow the user to feel in control? Can the system's behaviour be predicted? Does working with the product make the user feel safe? 4 items: unpredictable / predictable, obstructive / supportive, secure / not secure, meets expectations / does not meet expectations.
- **Stimulation**: How exciting and motivating is it to use the product? Does the user have fun? 4 items: valuable / inferior, boring / exciting, not interesting / interesting, motivating / demotivating.
- **Novelty**: Is the product innovative and creative? Is the users' attention captured? 4 items: creative / dull, inventive / conventional, usual / leading-edge, conservative / innovative.

In more detail, Attractiveness rates the overall aesthetics of Hatemeter Platform and how enticed users are by it; Perspicuity shows how easily people understand the Platform; Efficiency investigates how much effort the users put in the resolution of their tasks; Dependability gives an idea about it seeming trustworthy; Stimulation measures the joy of use; and Novelty represents how innovative a tool is perceived to be.

Attractiveness belongs to the dimension of "pure valence". Perspicuity, Efficiency and Dependability represent pragmatic quality, (i.e. they describe interaction qualities that relate to the tasks or goals the user aims to reach when using the product), and Stimulation and Novelty appear as representatives of hedonic quality (i.e. they do not relate to tasks and goals, but describe aspects related to pleasure or fun while using the product) (Schrepp, 2017). It is not assumed that the scales are independent (Figure 31). The general assumption is that the evaluation of the Attractiveness is based on the impression towards the other 5 scales, i.e. that the attractiveness is a result of the perceived perspicuity, efficiency, dependability, stimulation and novelty (Schrepp, Hinderks, Thomaschewski, 2017).

Figure 31 - Assumed scale structure of the UEQ



Source: University of Trento elaboration – Hatemeter Project

5.2. Evaluation of the Hatemeter Platform

This subsection shows in details the results concerning some of the KPIs aimed at measuring the efficacy and efficiency of the Hatemeter Platform and the analysis of the scale of the User Experience Questionnaire (UEQ scales).

5.2.1. KPIs evaluation

Below, some KPIs extracted from the log information of the Hatemeter Platform concerning the users involved so far, as well as statistics on the data extracted for the three languages, processed and stored in the Hatemeter database have been reported.

The users involved are NGO operators and participants to some dissemination events of the project (i.e. deployment days), who had the possibility to access the Platform and test its functionalities. The number of engaged operators using the Platform is 112, but the total number of accesses to the Platform is 953 and it includes also other types of users, for example high-school students, who used the Platform in training events to raise awareness on hate speech countering.

As for the number of persons reached, this includes also participants at dissemination events who did not have access to the Platform, but were reached by information on hate speech and the solutions proposed within Hatemeter. It is estimated that through dissemination, around 800 people have been reached, excluding the interview broadcast by Radio Popolare, whose average listening audience is around 166,000.

As regards the number of posts stored in the Hatemeter database, the numbers confirm the trend observed in the previous phase of the project, i.e. the amount of monitored messages is very different across the three languages, because the number of hashtags and their specificity is different. In particular, Italian hashtags are very specific and unambiguous and therefore fewer messages are retrieved and stored. French hashtags are more generic and after manually inspecting sample messages, it has been observed that a large amount of French tweets were linked to accounts in French Canada and concerned events involving the Islamic community and traditions there. As regards the English language, the high number of messages was an expected result, since this reflects the wide usage of English on Twitter, encompassing different countries.

Table 3 - Most relevant KPIs for the second pilot and their results

Category	KPIs	Results
Raise better understanding of Islamophobia	Number of accesses to the Platform	953 (from January to mid-September 2019)
Hatemeter Platform	N. of social media posts/profiles/ websites automatically identified and stored in the Hatemeter database	622,299 EN 152,025 IT 1,146,552 FR (unique tweets and replies, re-tweets not included)
Efficiency and Effectiveness of NGOs	Number of persons reached	~800 (from January to September 2019)
Stakeholder engagement and acceptance of the Hatemeter Platform	N. of engaged NGO/CSOs operators/ other stakeholders in using the Platform	112 (from January to September 2019)

Source: FBK elaboration – Hatemeter Project

5.2.2. UEQ scales

During the second and last pilot of the Hatemeter Platform, evaluations were collected twice. On the total of 50 users that tested and evaluated the Hatemeter Platform in the first collection during the deployment days, two were eliminated in the UEQ analysis after inconsistent responses.[39] This was repeated for the second collection, when 18 volunteers of the three NGOs involved in the project filled out the evaluation survey. Of these, one respondent was eliminated after analysing inconsistent scores. In total, 65 evaluation responses were collected for this second pilot of the Hatemeter Platform. For typical products evaluated so far, around 20-30 persons already give quite stable results.

Table 4 - UEQ scales, mean and variance

UEQ Scales (Mean and Variance)		
Attractiveness	1.633	0.70
Perspicuity	1.231	0.83
Efficiency	1.392	0.66
Dependability	1.092	0.47
Stimulation	1.762	0.70
Novelty	1.692	0.67

Source: University of Trento elaboration – Hatemeter Project

The standard interpretation of the scale means that values between -0.8 and 0.8 represent a neutral evaluation (yellow area) of the corresponding scale, values > 0.8 represent a positive evaluation (green area) and values < -0.8 represent a negative evaluation (red area). The range of the scales is between -3 (extremely bad) and +3 (extremely good). However, in real applications, only values in a restricted range will be generally observed. Due to the calculation of means over a range of different persons with different opinions and answer tendencies (for example the avoidance of extreme answer categories), it is extremely unlikely to observe values above +2 or below -2.

[39] Specially, if the UEQ is applied as an Online-Questionnaire not all participants will answer all items seriously. To detect such more or less random or not serious answers, a simple heuristic is used. All items in a scale should measure a similar UX quality aspect. The idea to detect random or not serious answers is to check how much differs the best and worst evaluation of an item in a scale. If there is a big difference (>3) this is seen as an indicator for a problematic data pattern. Of course, such situations can also result from random response errors or a misunderstanding of an item. Thus, it makes no sense to consider a response as problematic if this occurs just for a single scale. However, if this is true for 2 or 3 scales, this is a clear hint that the response is somehow suspicious. Consequently, answers from the data set that shows a critical value of 3 or higher have been eliminated from the analysis.

All scales show an extremely positive evaluation (see Table 4), especially the Attractiveness, Stimulation and Novelty scales that represent, respectively, the overall impression of the Hatemeter Platform, how interesting, exciting and motivating the Platform is and, finally, how the design of the product is creative and capacity to catch the interest of users.

Moreover, the scales of the UEQ can be grouped into categories of pragmatic quality (Perspicuity, Efficiency, Dependability) and hedonic quality (Stimulation, Originality). Pragmatic quality describes the tasks related to quality aspects; hedonic quality describes the non-task related quality aspects. Below is a calculation of the mean of the three pragmatic and hedonic quality aspects.

Table 5 - Pragmatic and Hedonic Quality, mean scores

Pragmatic and Hedonic Quality	
Attractiveness	1.63
Pragmatic Quality	1.24
Hedonic Quality	1.73

Source: University of Trento elaboration – Hatemeter Project

The measure of the three qualities of the Hatemeter Platform shows us a positive evaluation given by the volunteers with a mean score higher than 0.8 for each quality, especially for the Hedonic quality (see Table 5).

To get a better picture of the quality of a product it is therefore necessary to compare the measured user experience of the Hatemeter Platform with the results of other established products, for example from a benchmark data set containing quite different typical products (business software, web pages, web shops, social networks). The UEQ offers such a benchmark, which contains the data of 401 product evaluations with the UEQ (with a total of 18.483 participants in all evaluations). The comparison of the results for the evaluated product with the data in the benchmark allows conclusions about the relative quality of the evaluated product compared to other products.

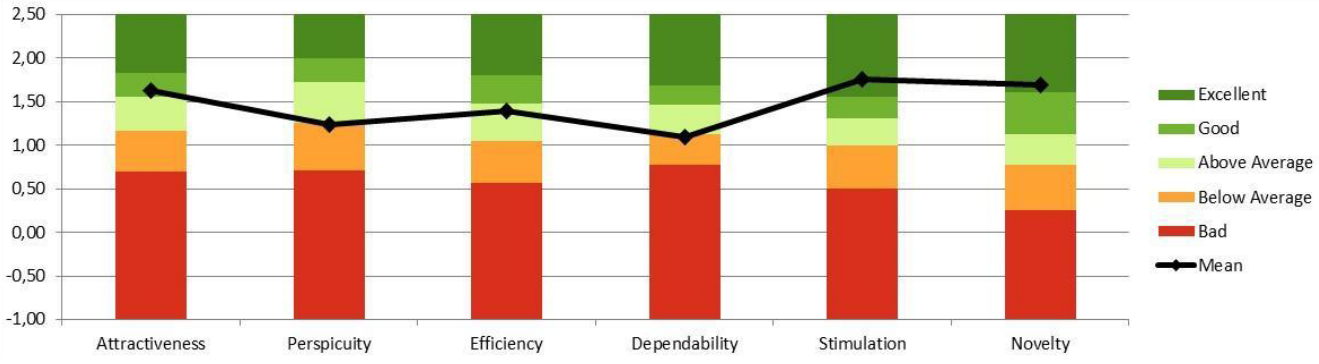
The benchmark classifies a product into 5 categories (per scale):

- Excellent: In the range of the 10% best results.
- Good: 10% of the results in the benchmark data set are better than the result for the evaluated product and 75% of the results are worse.
- Above average: 25% of the results in the benchmark are better than the result for the evaluated product, 50% of the results are worse.
- Below average: 50% of the results in the benchmark are better than the result for the evaluated product, 25% of the results are worse.
- Bad: In the range of the 25% worst results.

The benchmark graph shows how the UX quality of the Hatemeter Platform is. The line represents the results for the evaluated product. The coloured bars represent the ranges for the scales' mean values.

Figure 32 show that almost all of the scales have positive evaluation. Specifically, compared to the benchmark, evaluators gave an excellent score in terms of how interesting, exciting and motivating (i.e. the Stimulation scale), as well as innovative and creative (i.e. Novelty scale) the Platform is. Moreover, users seem to have the impression that it is easy to understand

Figure 32 - Benchmark for the UEQ scales



Source: University of Trento elaboration – Hatemeter Project

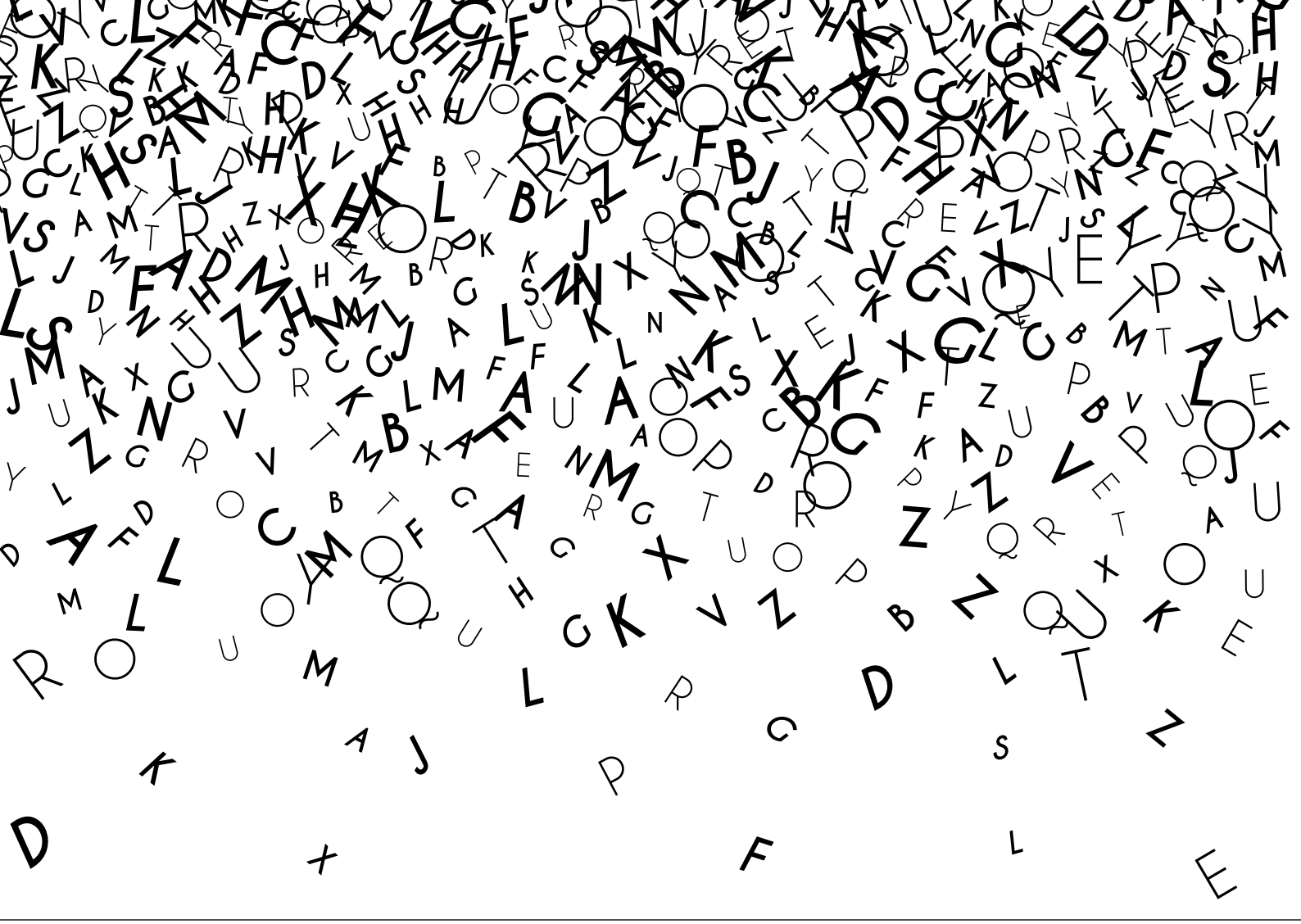
(i.e. the Perspicuity scale) and efficient to use (i.e. the Efficiency scale). The only drawback is that it offers a below average controllable interaction (i.e. the Dependability scale). The release of the next Platform update will indeed include more options for the users in terms of information sources, so that they will extend their interaction possibilities also to YouTube videos and to external sources like news.

5.2.3. Remarks

The second evaluation of the Hatemeter Platform has confirmed that the overall architecture was working well, and that the analyses were quite easy to understand, to use and could support NGO operators in gaining a better understanding of online Islamophobic discourse.

The counter-narrative tool was appreciated by operators and a rough estimate of the time needed to counter a hate message showed that operators using the Hatemeter Platform significantly reduced their response time by half (from 8 minutes to around 4 for each counter-message). This is a key indicator of the effectiveness of the Platform and its impact on the operators' daily activities.

In the final release of the Platform, some improvements and extensions were included, to take into account operators' feedback requesting further flexibility and variety in the information sources. Indeed, the outcome of YouTube monitoring was included in a new tab inside the Platform, showing how some Islamophobic hashtags and keywords can be used to retrieve videos that trigger hateful comments. This is mainly focused on identifying specific channels that spread online Islamophobic messages, and that operators may be interested to follow. A second extension concern the "Hashtag trends" view, in which hashtag peaks are associated with links to external news, where some event is mentioned that may be connected to the peak. This is completely automated and therefore introduces the possibility of some errors. Nevertheless, in most cases this allows operators to quickly identify an explanation for the peak. The functionality is active for all project languages, and is therefore connected to newspaper websites in English, French and Italian.



06

Suggestions and insights on the use of the Platform

The Hatemeter Platform has been implemented with specific objectives in mind, focusing on Islamophobia and to assist NGOs in their daily work by identifying hate speech and dangerous networks, monitoring them and providing adequate responses in a short amount of time. For the NGOs, the tool can be particularly useful in training new employees and volunteers. On the one hand, it can explain the phenomenon of Islamophobia with concrete data and facts and helps to explain and communicate this form of hate with accurate graphs and statistics. On the other hand, it provides counter-narrative suggestions, which can help practitioners to formulate responses to hate speech by providing support and inspiration in generating appropriate responses. For instance, a rough estimate of the time needed to counter a hate message demonstrates that operators using the Hatemeter Platform significantly reduce their response time by half (i.e. from 8 minutes to around 4 for each counter-message). This is a key indicator of the effectiveness of the Platform and its impact upon the operators' daily activities.

Nevertheless, the functionalities of the Platform are general-purpose and leave ample room for future development and extensions that can be implemented on demand, by both new practitioners and the academic world.

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. The analyses presented above shed light on some interesting ideas that can be utilised not only by NGOs, but also in academia and in research works. Indeed, the analytical tools allow 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 Hatemeter 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 the public debate 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.

Secondly, the focus of the Hatemeter Platform was directed towards the monitoring, analysing and tackling anti-Muslim hatred online, but the ICT tool could also be utilised in the investigation of other hate-speech 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, alongside 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 would 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). In addition, this tool could be adapted to and implemented in different domains, such as for education- or investigation-related purposes. The tool could also support 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 pertinent to the topic can be automatically published. However, such usage (i.e. suppression on the social media) may simply shift the problem to private messaging, where no intervention is possible.

Thirdly, and specifically from 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, together with the necessity to strictly comply with General Data Protection Regulation (GDPR) provisions, stakeholders should also acknowledge the fact that there are wider ethical considerations to evaluate in relation to the possibility to control users' networks.

Bibliography

Abbas, T. (2010). 'Islam and Muslims in the UK', Conference '*Islamic Studies in Europe*' held at the British Academy on 23–24 March, 2010.

Alietti, A. and Padovan, D. (2010). 'Il razzismo come legame sociale nella società dell'eccezione giuridica. Alcune note su anti-semitismo e anti-islamismo in Italia dopo l'11 settembre', Research Report '*Permanenza e metamorfosi delle forme del pregiudizio: antisemitismo e islamofobia dopo l'11 settembre*'. DOI: 10.13140/RG.2.2.28868.12164.

Alietti, A. and Padovan, D. (2018). 'Islamophobia in Italy: National Report 2017', in Bayrakli E., and Hafez F. (eds.), *European Islamophobia Report*, Istanbul: SETA. ISBN: 978-975-2459-00-7.

Allen, C., (2014). 'Anti-Social Networking: Findings from a Pilot Study on Opposing Dudley Mosque Using Facebook Groups as Both Site and Method for Research', in *SAGE Open*, *4*(1), pp. 1-12. DOI: 10.1177/2158244014522074.

Amiraux, V. (2015). 'Après le 7 janvier 2015, quelle place pour le citoyen musulman en contexte libéral sécularisé?', in *Multitude*s, 59,(2), pp. 83-93. DOI:10.3917/mult.059.0083.

Amnesty International Italia (2018). *Conta fino a 10. Barometro dell'odio in campagna elettorale*. Available online at: https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2018/02/16105254/report-barometro-odio.pdf

Amnesty International Italia (2019). *Barometro dell'odio. Elezioni europee 2019*. Available online at: https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2019/05/29202706/Amnesty-barometro-odio-2019.pdf

Article 19 (2018). *Responding to 'Hate Speech': Comparative Overview of Six EU Countries*, London: Article 19.

Awan, I. (2014). 'Islamophobia and Twitter: A typology of online hate against Muslims on social media', in *Policy & Internet*, *6*(2), pp. 133-150. DOI: 10.1002/1944-2866.POI364.

Awan, I. and Zempi, I. (2017). 'I Will Blow Your Face OFF - VIRTUAL and Physical World Anti-muslim Hate Crime', in *The British Journal of Criminology*, *57*(2), pp. 362-380. DOI: 10.1093/bjc/azv122.

Awan, I. and Zempi, I., (2016). 'The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts', in *Aggression and violent behavior*, *27*, pp. 1-27. DOI: 10.1016/j.avb.2016.02.001.

Awan, I., (2016). 'Islamophobia on social media: A qualitative analysis of the Facebook's Walls of Hate'. *International Journal of Cyber Criminology*, *10*(1), pp. 1 - 20.

Baehr, P. and Gordon, D. (2013). 'From the headscarf to the burqa: the role of social theorists in shaping laws against the veil', in *Economy and Society*, 42(2), pp. 249-280. DOI : 10.1080/03085147.2012.718620.

Banet, R. and Fauchet, B. (2018). *20 millions de musulmans en France ? Ils sont environ 4 fois moins, selon les estimations les plus sérieuses*, AFP. Available online at: https://factuel.afp.com/20-millions-de-musulmans-en-france-ils-sont-environ-4-fois-moins-selon-les-estimations-les-plus

Bakalis, C. (2018). 'Rethinking cyberhate laws', in *Information & Communications Technology Law*, 27(1), pp. 86-110. DOI: 10.1080/13600834.2017.1393934.

Barlow, C. and Awan, I. (2016). '"You need to be sorted out with a knife": the attempted online silencing of women and people of Muslim faith within academia', in *Social Media + Society*, *2*(4), pp. 1-11. DOI: 10.1177/2056305116678896.

Baubérot, J. (2012). *La Laïcité Falsifiée*, Paris: La Découverte.

Bortone, R. and Cerquozzi, F. (2017). 'L'hate speech al tempo di Internet', in *Aggiornamenti Sociali*, December, pp. 818-27.

Brindle, A., (2016). 'Cancer has nothing on Islam: a study of discourses by group elite and supporters of the English defence league', in *Critical Discourse Studies*, *13*(4), pp. 444-459. DOI: 10.1080/17405904.2016.1169196.

Carter, D.M. (2017). '(De) constructing difference: a qualitative review of the 'othering' of UK Muslim communities, extremism, soft harms, and Twitter analytics', in *Behavioral Sciences of Terrorism and Political Aggression*, 9(1), pp. 21-36. DOI: 10.1080/19434472.2016.1236142.

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. ISBN: 978-92-2-219783-5.

Cleland, J., Anderson, C. and Aldridge-Deacon, J. (2017). 'Islamophobia, war and non-Muslims as victims: an analysis of online discourse on an English Defence League message board', in *Ethnic and Racial Studies*, pp. 1-17. DOI: 10.1080/01419870.2017.1287927.

CNCDH (2017). *Rapport sur la lutte contre le racisme, l'antisémitisme et la xénophobie.* Available online at: http://www.cncdh.fr/fr/publications/rapport-2017-sur-la-lutte-contre-le-racisme-lantisemitisme-et-la-xenophobie

Commission on British Muslims and Islamophobia (2004), *Islamophobia: Issues, Challenges and Action*, Stoke on Trent: Trentham Books.

Copsey, N. (2010). *The English defence league: Challenging our country and our values of social inclusion, fairness and equality*, London: Faith Matters.

Copsey, N., Dack, J., Littler, M. and Feldman, M. (2013). *Anti-Muslim Hate Crime and Cumulative Extremism, and Policy Implications*, Teesside University. Available online at: https://tellmamauk.org/tell-mama-20142015-findings-on-anti-muslim-hate

Dargent, C. (2009). 'L'État et la difficile saisie statistique de la religion : l'exemple des protestants dans les recensements en France au XIXe siècle', in *Population*, 64(1), pp. 215-232. DOI: 10.3917/popu.901.0215.

Decret n°2005-284 (J.O, 30 March 2005), Article 1. Available online at : https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006070719&idArticle=LEGIARTI000006419500

Devriendt, E., Monte, M. and Sandré, M. (2018). 'Analyse du discours et catégories «raciales»: problèmes, enjeux, perspectives', in *Mots. Les langages du politique,* 116(1), pp. 9-37.

Dematteo, L. (2018). 'Marine + Matteo: L'étreinte populiste', in *Vacarme*, 84(3), pp. 17-24. DOI: 10.3917/vaca.084.0017.

Duffy, B. (2016). *Perceptions are not reality: what the world gets wrong*, Ipsos Mori. Available online at: https://www.ipsos.com/ipsos-mori/en-uk/perceptions-are-not-reality-what-world-gets-wrong

Ekman, M. (2015). 'Online Islamophobia and the politics of fear: manufacturing the green scare', in *Ethnic and Racial Studies*, 38(11), pp. 1986-2002. DOI: 10.1080/01419870.2015.1021264.

eMORE, Monitoring and Reporting Online Hate Speech in Europe (2017), *An Overview on Hate Crime and Hate Speech in 9 EU Countries*, Torri di Quartesolo: Research Centre on Security and Crime.

Evolvi, G. (2018). 'Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses', in *Religions*, 9(10), pp. 307-321. DOI: 10.3390/rel9100307.

Fabre, M. (2017). 'Sens et usages contemporains de la laïcité', in Éducation et socialisation, 46. DOI: 10.4000/edso.2754.

Faloppa, F. (2019), 'La mappa delle parole', in Amnesty International Italia, *Barometro dell'odio. Elezioni europee 2019*. Available online at: https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2019/05/29202706/Amnesty-barometro-odio-2019.pdf

Feldman, M. and Littler, M. (2014). *Tell MAMA Reporting 2013/14 anti-Muslim overview, analysis and 'cumulative extremism', Centre for Fascist, Anti-Fascist and Post-Fascist Studies.* Teesside University. Available online at: https://www.tellmamauk.org/wp content/uploads/2014/07/finalreport.pdf

Froio, C. (2018). 'Race, Religion, or Culture? Framing Islam between Racism and Neo-Racism in the Online Network of the French Far Right', in *Perspectives on Politics*, 16(3), pp. 696–709. DOI: 10.1017/S1537592718001573.

Geisser V. (2003). *La nouvelle Islamophobie*. Paris: La découverte. ISBN: 978-2707140609.

Gill, A.K. and Harrison, K. (2015). 'Child Grooming and Sexual Exploitation: Are South Asian Men the UK Media's New Folk Devils?', in *International Journal for Crime, Justice and Social Democracy, 4*(2), pp. 34-49. DOI: 10.5204/ijcjsd.v4i2.214.

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', Conference '*Hate speech e libertà di espressione*' held in Milan on 9 October, 2015.

Gometz, G. (2017). 'L'odio proibito: la repressione giuridica dello hate speech', in *Stato, Chiese e pluralismo confessionale*, (32), pp. 1-39. DOI: 10.13130/1971-8543/9145.

Gupta, C. (2009). 'Hindu women, Muslim men: Love Jihad and conversions', in *Economic and Political Weekly*, 44(51), pp.13-15. DOI: 10.2307/25663907.

Henne, P. (2015). *Latest Trends in Religious Restrictions and Hostilities*, Washington DC: Pew Research Centre.

Horsti, K. (2017). 'Digital Islamophobia: The Swedish woman as a figure of pure and dangerous whiteness' in *New Media and Society*, 19(9), pp. 1440-1457. DOI: 10.1177/1461444816642169.

Kelsey, D. (2017). *Media and Affective Mythologies: Discourse, Archetypes and Ideology in Contemporary Politics*, London: Palgrave Macmillan. ISBN: 978-3319607580.

Kohl, U. (2018). 'Islamophobia, "gross offensiveness" and the Internet', in *Information & Communications Technology Law*, *27*(1), pp. 111-131. DOI: 10.1080/13600834.2017.1393936.

Larsson, G. (2007). 'Cyber-islamophobia? The case of WikiIslam', in *Contemporary Islam, 1*(1), pp. 53-67. DOI: 10.1007/s11562-007-0002-2.

Law on 'liberté de la presse' (J.O, 29 July 1881), Article 24. Available online at: https://www.legifrance.gouv.fr

Law on 'maitrise de l'immigration, l'intégration et l'asile' (J.O, 15 November 2007). Available online at: http://www.conseil-constitutionnel.fr/conseil-constitutionnel/francais/les-decisions/acces-par-date/decisions-depuis-1959/2007/2007-557-dc/decision-n-2007-557-dc-du-15-novembre-2007.1183.html

Leveson Inquiry (2012). *Culture, Practice and Ethics of the Press.* Available online at: https://www.gov.uk/government/publications/leveson-inquiry-report-into-the-culture-practices-and-ethics-of-the-press

Ljujic, V., Engel, V., Castriota, A., and Carter, A. (2016). *Xenophobia, radicalism and hate crime in Europe 2015*, OSCE report - Organization for security and co-operation in Europe.

Longhi, J. (2017). *Tweets politiques: corrélation entre forme linguistique et information véhiculée*, #Info. Partager et commenter l'Info sur Twitter et Facebook. Available online at: https://halshs.archives-ouvertes.fr/halshs-01841132/document 2017

Malchiodi, M. (2016). *L'islam nei social media*, Pavia: Osservatorio di Pavia.

Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., and McClosky, D. (2014). 'The Stanford CoreNLP Natural Language Processing Toolkit', in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60.

Miller, C. and Smith, J. (2017). *Toplines: Anti-Islamic Hate on Twitter.* London: Demos. Available online at: https://www.demos.co.uk/project/anti-islamic-content-on-twitter

Marwan, M. (2014), 'Un nouveau champ de recherche', in *Sociologie*, 5(1), pp. 1-11.

Morgan, G. and Poynting, S. (eds) (2012). *Global Islamophobia: Muslims and Moral Panic in the West*, Abingdon: Routledge.

Oboler, A. (2013). *Islamophobia on the Internet: The Growth of Online Hate Targeting Muslims*. Online Hate Prevention Institute. Available online at: http://ohpi.org.au/islamophobia-on-the-internet-the-growth-of-online-hate-targeting-muslims

Oboler, A. (2016). 'The normalisation of islamophobia through social media: Facebook', in Awan, I. (eds), *Islamophobia in Cyberspace: Hate Crimes Go Viral,* pp. 41-62. New York: Routledge.

Ordonnance n°2018-1125 of the 12 of December 2018 (J.O 12 December 2018), Article 13. Available online at: https://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000026268247&cidTexte=LEGITEXT000006070719

OSCAD (2020), *Elenco normativa antidiscriminatoria*. Available online at: https://www.poliziadistato.it/statics/13/oscad_elenco_normativa_antidiscriminatoria.pdf

Osservatorio sulle Discriminazioni (2010). *Rapporto 2010*, Mantova: Osservatorio sulle Discriminazioni.

Pew Global Attitudes Project (2015). *Faith in European Project Reviving*, Washington DC: Pew Research Centre.

Pugiotto, A. (2012), 'Le parole sono pietre? I discorsi di odio e la libertà di espressione nel diritto costituzionale', Conference '*V Convegno Nazionale dell'Avvocatura per i diritti LGBT-Rete Lenford, Omofobia, Transfobia e Diritto Penale'* held in Bari on 30 November –1 December, 2012.

Roucate, D. (2015). *Quatre questions sur les statistiques ethniques*, LeMonde.fr. Available online at: https://www.lemonde.fr/les-decodeurs/article/2015/05/06/quatre-questions-sur-les-statistiques-ethniques_4628874_4355770.html

Runnymede Trust (1997). *Islamophobia: A Challenge for Us All,* London: Runnymede Trust. ISBN: 0-902397-98-2

Russospena, M. (2009). 'L'uso strumentale delle differenze religiose: l'Islam nelle retoriche pubbliche', in Naletto, G. (eds), *Rapporto sul razzismo in Italia*, pp. 37–46, Roma: Manifestolibri.

Sadique, K., Tangen, J., and Perowne, A. (2018). The Importance of Narrative in Responding to Hate Incidents Following 'Trigger' Events, London: Tell Mama.

Scaramella, C. (2016). *Discorsi d'odio e Social Media. Criticità, strategie e pratiche d'intervento*. Available online at: https://www.arci.it/app/uploads/2018/05/progetto_PRISM_-_bassa.pdf

Sciortino, G. (2002). 'Islamofobia all'italiana', in *Polis* 16 (1), pp. 103–23. DOI: 10.1424/2944.

Schrepp ,M., Hinderks A. and Thomaschewski J. (2017). 'Construction of a Benchmark for the User Experience Questionnaire (UEQ)', in I*nternational Journal of Interactive Multimedia and Artificial Intelligence*, 4(4), pp. 40-44. DOI: 10.9781/ijimai.2017.445.

Schrepp, M. (2017). *User Experience Questionnaire (UEQ)*. Available online at: http://www.ueq-online.org

Selma (2019). 'Hacking Online Hate: Building an Evidence Base for Educators', in *European SchoolNet*. Available online at: https://hackinghate.eu/news/hacking-online-hate-building-an-evidence-base-for-educators

Serdar, K. (2015). 'Islamophobia in Western Europe: A Comparative, Multilevel Study', in *Journal of Muslim Minority Affairs*, 35(3), pp. 450–65. DOI: 10.1080/13602004.2015.1080952.

Simon, P. (2008). 'Les statistiques, les sciences sociales françaises et les rapports sociaux ethniques et de «race»', in *Revue française de sociologie*, vol. 49(1), pp. 153-162. DOI:10.3917/rfs.491.0153.

Spärck Jones, K. (1972). 'A Statistical Interpretation of Term Specificity and Its Application in Retrieval', in *Journal of Documentation*, 28, pp. 11–21.

Strabac, Z. and Listhaug, O. (2008). 'Anti-Muslim Prejudice in Europe: A Multilevel Analysis of Survey Data from 30 Countries', in *Social Science Research*, 37(1), pp. 268–86. DOI: 10.1016/j.ssresearch.2007.02.004.

Strapparava, C. and Valitutti, A. (2004). 'WordNet-Affect: an Affective Extension of WordNet', in *Proc. of 4th International Conference on Language Resources and Evaluation* (LREC 2004), pp. 1083–1086, held in Lisobon on May, 2004.

Strohl, D.J. (2019). 'Love jihad in India's moral imaginaries: religion, kinship, and citizenship in late liberalism', in *Contemporary South Asia*, 27(1), pp.27-39. DOI: 10.1080/09584935.2018.1528209.

Szalek, B. Z. (2013). 'Some praxiological reflections on the so-called Overton window of political possibilities, framing and related problems', in *Reality of Politics. Estimates*, 4, 237-257.

Szmania, S. and Fincher, P. (2017). 'Countering Violent Extremism Online and Offline', in *Criminology and Public Policy*, January. DOI: 10.1111/1745-9133.12267.

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)', in *Mots, Les langages du politique,* 116, DOI: 10.4000/mots.23050.

Tell Mama (2014). *Analysing the Lexicon of Anti-Muslim Prejudice.* Available online at: http://www.tellmamauk.org/analysing-the-lexicon-of-anti-muslim-prejudice

Tell Mama (2016). *A Constructed Threat: Identity, Intolerance and the Impact of Anti-Muslim Hatred.* Tell MAMA Annual Report. Available online at: https://www.tellmamauk.org/constructed-threat-identity-intolerance-impact-anti-muslim-hatred-tell-mama-annual-report-2016

The Public Order Act 1986, Part III, Articles 18-23.

The Public Order Act 1986 - Part 3A Hatred against persons on religious grounds or grounds of sexual orientation (Inserted by The Racial and Religious Hatred Act 2006 (c1).).

The Public Order Act 1986 - Part 3A Hatred against persons on religious grounds or grounds of sexual orientation (Inserted by The Criminal Justice and Immigration Act 2008 (c4).).

User Experience Questionnaire (2017). *10.04.2019*. Available online at: http://www.ueq-online.org

Vitullo, A. (2019). 'Odio e religione: musulmani nel mirino', in Amnesty International Italia, *Barometro dell'odio. Elezioni europee 2019*. Available online at: https://d21zrvtkxtd6ae.cloudfront.net/public/uploads/2019/05/29202706/Amnesty-barometro-odio-2019.pdf

Vox, Osservatorio Italiano sui Diritti (2018). *La mappa dell'intolleranza*. Available online at: http://www.voxdiritti.it/la-mappa-dellintolleranza-anno-3-la-nuova-radiografia-dellitalia-che-odia-online

Wieviorka, M. (2017). *Le Front national, entre extrémisme, populisme et démocratie*, Paris: Maison des Sciences de l'Homme.

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.

Trento, January 2020

ISSN 2284-3302
ISBN 978-88-8443-886-7