MEDIA CONTENT ANALYSIS ON ONLINE HATE SPEECH

Comparative Report

Compiled under the Coalition of Positive Messengers to Counter Online Hate Speech Project

Project reference number: JUST/2015/PRAC/AG/BEST/8931

Massimo Buscema
Guido Ferilli
Giulia Massini
Emma Zavarrone

This publication has been produced with the financial support of the Rights, Equality and Citizenship (REC) Programme of the European Union. The contents of this publication are the sole responsibility of The Languages Company (TLC) and can in no way be taken to reflect the views of the European Commission.
C o n t e n t s

SECTION 1 5
LIST OF ABBREVIATIONS 6
EXECUTIVE SUMMARY 7
INTRODUCTION 7
1. DIGITAL ECONOMY AND SOCIETY 9
2. METHODOLOGY 17
2.1. Research aim 17
2.2. Justification of the research 17
2.3 Definitions used 17
2.4 Research methods 18
2.5. Analysis Tool 22
3. RESULTS OF MEDIA CONTENT ANALYSIS 25
3.1. Main results of the investigation in United Kingdom 25
3.1.1. Overview on national context 26
3.1.2. UK’s Media content analysis 26
3.1.2.1. Keywords Analysis 2017 27
3.1.2.2. Social media Twitter and FB Analysis 2017 28
3.1.2.3. Social media Twitter and FB Analysis 2016 30
3.1.3. Principal hints 34
3.2. Main results of the investigation in Italy 34
3.2.1. National context 34
3.2.2. Italy’s Media content analysis 35
3.2.2.1. Keywords Analysis 2017 36
3.2.2.2. Social media Twitter and FB Analysis 2017 37
3.2.2.3. Social media Twitter and FB Analysis 2016 39
3.2.3. Principal hints 43
3.3. Main results of Bulgaria’s investigation 44
3.3.1. National context 44
3.3.2. Bulgaria’s Media content analysis 45
3.3.2.1. Keywords Analysis 2017 47
3.3.2.2. Social media Twitter and FB Analysis 2017 48
3.3.2.3. Social media Twitter and FB Analysis 2016 49
3.3.3. Principal hints 54
3.4. Main results of Romania’s investigation 55
3.4.1. National context 55
3.4.2. Romania’s Media content analysis 56
3.4.2.1. Keywords Analysis 2017 57
3.4.2.2. Social media Twitter and FB Analysis 2017 58
3.4.2.2. Social media Twitter and FB Analysis 2016 60
3.4.3. Principal hints 64
3.5. Main results of Czech Republic’s investigation 65
3.5.1. National context
3.5.2. Czech Republic’s Media content analysis
3.5.2.1. Keywords Analysis 2017
3.5.2.2. Social media Twitter and FB Analysis 2017
3.5.2.3. Social media Twitter and FB Analysis 2016
3.5.3. Principal hints

3.6. Main results of Croatia’s investigation
3.6.1. National context
3.6.2. Croatia’s Media content analysis
3.6.2.1. Keywords Analysis 2017
3.6.2.2. Social media Twitter and FB Analysis 2017
3.6.2.3. Social media Twitter and FB Analysis 2016
3.6.3. Principal hints

3.7. Main results of Greece’s investigation
3.7.1. National context
3.7.2. Greece’s Media content analysis
3.7.2.1. Keywords Analysis 2017
3.7.2.2. Social media Twitter and FB Analysis 2017
3.7.2.3. Social media Twitter and FB Analysis 2016
3.7.3. Principal hints

4. COMPARISON OF THE COUNTRIES THROUGH THE LENS OF INTERNATIONAL RANKING
4.1 Global competitiveness
4.2 Press Freedom
4.3 Freedom in the World: political rights and civil liberties global scores
4.4 Economic Freedom
4.5 Corruption Perception Index
4.6. Global terrorist attacks
4.7. Comparison between the total population and the immigrants residing in the countries

5. CONCLUSIVE ANALYSIS

REFERENCES

SECTION 2 - NEURAL NETWORKS ANALYSIS OF THE ON LINE HATE SPEECH IN 2016
INTRODUCTION
PRELIMINARY DEFINITION, DATA AND METHODS

METHODS

1. DATA ANALYSIS OF EACH COUNTRY
1.1. Bulgaria
1.1.1. Analysis of the frequency
1.1.2. Similarities in behaviour
1.1.3. Cause-Effect Relationships
1.1.4. Relations between the analyzed days (13.03.2016 - 11.05.2016)

1.2. Czech Republic
1.2.1. Analysis of the frequency
1.2.2. Similarities in behaviour
1.2.3. Cause-Effect Relationships
1.2.4. Relations between the analyzed days (23.07.2016 - 22.09.2016)

1.3. Croatia
1.3.1. Analysis of the frequency
1.3.2. Similarities in behaviour
1.3.3. Cause-Effect Relationships
1.3.4. Relations between the analyzed days (07.02.2016 - 12.04.2016)

1.4. Greece
1.4.1. Analysis of the frequency
1.4.2. Similarities in behaviour
1.4.3. Cause-Effect Relationships
1.4.4. Relations between the analyzed days (07.02.2016 - 12.04.2016)

1.5. Italy
1.5.1. Analysis of the frequency
1.5.2. Similarities in behaviour
1.5.3. Cause-Effect Relationships
1.5.4. Relations between the analyzed days (19.09.2016 - 17.11.2016)

1.6. Romania
1.6.1. Analysis of the frequency
1.6.2. Similarities in behaviour
1.6.3. Cause-Effect Relationships
1.6.4. Relations between the analyzed days (25.07.2016 - 23.09.2016)

1.7. United Kingdom
1.7.1. Analysis of the frequency
1.7.2. Similarities in behaviour
1.7.3. Cause-Effect Relationships
1.7.4. Relations between the analyzed days (23.02.2016 - 23.04.2016)

2. COMPARATIVE ANALYSIS

2.1. Fusion G-IN-OUT
   2.1.1. Fusion G-IN-OUT - Data analysis of each country in the global graph

2.2. Fusion MRG

3. MACHINE LEARNING USED FOR THIS ANALYSIS

3.1. Auto Contractive Map (Auto-CM)
3.2. Maximally Regular Graph (MRG)
3.3. Dynamic Networks Block (DNB)
3.4. Adaptive Fusion System (from: Theory of Impossible Worlds (TIW))

REFERENCES

RECOMMENDATIONS
SECTION 1

Authors*:
Guido Ferilli
Emma Zavarrone

* we thanks: Alessia Forciniti and Mariangela Paolillo for their contribution
# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL</td>
<td>Anti-Defamation League</td>
</tr>
<tr>
<td>FB</td>
<td>Facebook</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>ECRI</td>
<td>European Commission against Racism and Intolerance</td>
</tr>
<tr>
<td>EMC</td>
<td>Electronic Media Council</td>
</tr>
<tr>
<td>GONG</td>
<td>Citizens organize to oversee voting</td>
</tr>
<tr>
<td>HRT</td>
<td>Croatian Radio Television</td>
</tr>
<tr>
<td>IULM</td>
<td>International University of Languages and Media</td>
</tr>
<tr>
<td>LGBTI</td>
<td>Lesbian, gay, bisexual, transgender, and intersex persons</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-governmental organization</td>
</tr>
<tr>
<td>OHS</td>
<td>Online hate speech</td>
</tr>
<tr>
<td>RSF</td>
<td>Reporters Without Borders</td>
</tr>
<tr>
<td>SDA</td>
<td>Sofia Development Association</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
</tbody>
</table>
Executive Summary

This report, composed of two sections, provides a comparative analysis of the online hate speech (OHS) phenomenon in terms of multimedia content analysis. The study is developed under the project "Coalition of Positive Messengers to Counter Online Hate Speech", with the support of seven different organizations: SDA (Sofia Development Association, Bulgaria), the Languages Company (The United Kingdom), Center for Peace (Croatia), People in Need (Czech Republic), Asociația Divers (Romania), Associazione FORMA.Azione (Italy), and Municipality of Agii Anargiri-Kamatero (Greece). Based on the national media content analysis the current report proposes meaningful relations between the online hate speech level - related to the recent refugee crisis and migration flows to Europe - and the tolerance level - towards the immigrants - in seven European Countries. The report proposes a cross - Countries evaluation, in order to create a common European framework in terms of online hate speech phenomenon and in terms of concepts such as integration, tolerance, security perception and hostility climate towards minorities. Finally the report concludes proposing recommendations (after the section 2), on dealing with hate speech at the policy and grassroots level.

Introduction

“Coalition of Positive Messengers to Counter Online Hate Speech” Project has the aim to map the issue of online hate speech targeted against migrants, refugees and asylum seekers in seven EU countries (UK, Italy, Bulgaria, Romania, Czech Republic, Croatia and Greece). It originates from the need for a more effective civil society in response to phenomenon of hate speech and in order to develop active engagement of local communities, in creating and sharing narrative tools able to contrast the online xenophobic discourses. After the mapping of the hostility degree against the migrants in the seven European Countries (based on their multimedia contents), it have been analysed and confronted the common aspects. Which are the similarities or the differences of perception towards the migratory phenomenon? To answer this question, the analysis has been focused on the multimedia contents. The implementation of the entire methodology has required long time and labour -intensive, trying to focus on the opportunity to answer research questions in a coherent and effective way, scientifically speaking. The limits and difficulties faced over time have been many. Limits that have somewhat slowed down this task and led to the starting point the phase...
of processing of data several times. However, these limitations have represented foreseeable aspects, as it is a project that for the first time involved a wide multi-lingual analysis in the strict sense. Being able to manipulate and process different languages - not commons to text analysis as reference methodology – this study proves to be the real challenge of the project.

The time for data processing was longer than expected, considering the specific needs of the project's partners and evaluating the precious commitment provided by individual stakeholders. Each of them has employed enormous efforts for the scientific nature of the analysis and organizational commitment. Despite the absence of statistical background in many partners, they contributed very punctually and collaboratively to the goal of reaching valuable results for the project. In accordance with the methodology established, the media content analysis has covered two time intervals, one in 2016 and the other in 2017. Therefore, in this report it is proposed to take a fresh outlook between Twitter's multimedia content and Facebook’s multimedia contents for each Country. The comparison between 2016 and 2017 is not possible for methodological reason: the respondents in each country are different from occasion to another and the time periods under observation are different. The project focuses both on a Keywords Analysis and on multimedia contents of three selected media outlets, except for the UK with four. Each Country presents the same kind of structural results, so that, it is possible to carry out a coherent comparison. The research, summarized through an automated data search algorithm for monitoring social media content of media outlets, is based on a well-known methodology. IULM Team provides research design, collection and elaboration of data for 2017, whilst for 2016 ELIF LAB provides the data collection. IULM Team carries out the content analysis and interpretation.

The activities were carried out in two steps. The first step of this investigation was related to the understanding of the most relevant and pertinent events within each national context in 2016. In other words, the aim during this step is to identify events for which the media and people have spoken. This analysis of the context has led to the definition of two aspects:

a) Choice of the most relevant events for 2016 and understanding if OHS is increasing around these

b) Extraction of a list of ten keywords for 2017 and 2016 events.

The second step considers the choice of three online media for each Country, based on their popularity on Facebook and Twitter.

The extracted multimedia contents in 2016 and 2017, definable in terms of Big data, represents the unit of the Statistical analysis. The methodologies used are the Text Mining and the Social Network Analysis. The analysis of the data allows the mapping of the hostility degree in each Country.
The aim is to provide the profile of the Hate speech in each of the seven countries through the analysis, and with the support of the project partners’ evaluation and contextualization of the results. Although computational aspects are identical for each Country, the contextual diversity and multimedia culture can change the interpretation of the outcomes. The online hate speech phenomenon is somewhat affected by the level of social and cultural context of each country. In addition, it is crucial to evaluate the degree of multimedia evolution, as it is crucial to understand the level of actual use of the social media. If the traditional media are still the main source of transmission of information, it is possible to be in fallacy affirming that in a country is absent the online hate speech phenomenon. For example, the increasing use of online devices can affect the interpretation of the statistical results obtained. In other words, the national context might influence the interpretation of a simple evaluation of the statistical analysis.
Considering this last assumption and for this purpose the study has been introduced with the multimedia framework of each Country investigated.

1. Digital economy and society

Information and social media affect people’s everyday lives in many ways. The national and European statistical institutes carry out annual studies regarding several aspects of information and communication technologies usage in households and by individuals.

Eurostat every year investigates the ICT usage in the context of the Digital Single Market process available both in the Consumer Conditions Scoreboard and in the Employment Guidelines (e-skills of individuals)⁠¹. The characteristics analysed by Eurostat in order to evaluate the digital economy and society are referred to: access and use of ICTs by individuals and/or in households, use of the Internet and other electronic networks for different purposes by individuals, ICT security and trust, ICT competence and skills, barriers to the use of ICT and Internet, perceived effects of ICT usage on individuals, use of ICT by individuals to exchange information and services with governments and public administrations (e-government).

It also analyses the Internet access, how technologies enabling the connection among networks from anywhere and at any time (ubiquitous connectivity).

¹ http://ec.europa.eu/eurostat/documents/341889/725524/Monitoring+the+Digital+Economy+%26+Society+2016-2021/7df02d85-698a-4a87-a6b1-7994df7fbeb7
The data set is based on sample surveys\(^2\). The reference area is relative to EU-Member States, Candidate countries, Iceland and Norway.

The investigation mentioned above is based on the Regulation 808/2004 of the European Parliament and of the Council of 21 April 2004\(^3\) concerning Community statistics on the information society.

The purpose of this framework regulation is to establish a common framework for the systematic production of statistics of the Community on the Digital economy and society.

Following this path it is possible to see every relevant legal act. Furthermore, it has to be considered also the Regulation (EC) No 223/2009 on European statistics (recital 24 and Article 20(4)) of 11 March 2009 (OJ L 87, p. 164), regarding the need to establish common principles and guidelines ensuring the confidentiality of data used for the production of European statistics and the access to those confidential data.\(^4\)

Finally, the investigation looks at ICT annual model questionnaires and how ICT has become widely available to the general public, both in terms of accessibility as well as in terms of costs.\(^5\)

![Figure 1. Internet access and broadband Internet connections of households, EU-28, 2007–2016](http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals)

\(\text{Figure 1. Internet access and broadband Internet connections of households, EU-28, 2007–2016}\)

\(\text{(}\%\text{ of all households)}\)

\(\text{Source:}\)


\(^{2}\) https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp

\(^{3}\) http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32004R0808


\(^{5}\) http://ec.europa.eu/eurostat/statistics-explained/index.php/Internet_access_and_use_statistics_-_households_and_individuals
In 2016, the share of EU-28 households with Internet access rouses 2 additional percentage points compared with 2015, reaching 85%. Broadband represents by far the most common form of Internet access in all EU Member States: it was used by 83% of the households in the EU-28 in 2016 (Eurostat, 2017).

Households with highest proportion of the Internet access in 2016 was recorded in Luxembourg and in the Netherlands (percentage equal to 97%) while Denmark, Sweden, the United Kingdom, Germany and Finland also reported that more than 9 out of every 10 households had Internet access in 2016. The lowest rate of Internet access among the EU Member States was observed in Bulgaria (64%). However, Bulgaria, together with Spain and Greece, recorded a rapid expansion of the proportion of households having access to the Internet with an increase of 19% between 2011 and 2016.

Figure 2. Internet access of households, 2011 and 2016 (% of all households).

Source:
The figure 2 shows the percentage of households Internet access from 2011 to 2016. In particular, according to the Eurostat's investigation, among monitored Countries, UK represents the Country with highest level of social media access and it occupies the fifth position in the chart shown above. Republic Czech occupies the thirteenth position and Italy the nineteenth place. They are followed by Croatia, Romania, Greece and Bulgaria.

In terms of Internet usage "at least 9 out of 10 individuals in Denmark, Luxembourg, the United Kingdom, Finland, the Netherlands, Sweden and Germany used the Internet. By comparison, slightly more than two thirds of all individuals aged 16 to 74, used the Internet in Portugal (70 %), Greece and Italy (both 69 %). However, this share falling to 60 % in Romania and 59 % in Bulgaria"(Figure 3).
Figure 3. Frequency of Internet use, 2016 (% of individuals aged 16 to 74).


Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
In the Figure 3 the highest proportion of Internet use is related to UK followed by Czech Republic, Croatia, Greece and Italy, Romania and Bulgaria. Bulgaria is the lowest country in terms of multimedia evolution, but it is not true in terms of social networking activities.

During 2016, in UK, the 52% of individuals between 16 to 74 years age used social media such as Facebook or Twitter for social networking. It means that one more time the United Kingdom is the most multimedia Country (Figure 4).

Another investigation is the one about how individuals use Internet in order to participate in social networking.6

---

6 http://ec.europa.eu/eurostat/statistics-explained/index.php/Internet_access_and_use_statistics_-_households_and_individuals

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931

14
Again, this research shows that the British from 16 to 74 years age spend a lot of time on Internet and utilize intensively the social network platforms.

After UK there are Croatia, Greece, Bulgaria, Czech Republic, Romania and finally Italy. It is worth pointing out that this study is not enough satisfying in terms of answer to the multimedia evolution phenomenon.

Another contribution for better understanding the multimedia situation in Europe is the annual report Digital Yearbook in 2017 written thanks to the help of We Are Social and Hootsuite.\(^7\)

This report is a digital yearbook of Internet, social media and mobile data for 239 Countries around the world.

It is divided into several sections. One section is dedicated to the global overview through statistical key indicators for social media and mobile; another section concerns the annual digital growth. According to the aims of the study, we can trace a multimedia overview for each Country involved in the present project report.

Following an alphabetical order the first place is occupied by Bulgaria. The Digital Yearbook for Bulgaria indicates a total population of 7.07 million inhabitants. Between them the 59% of the Bulgarian usually use Internet and the 51% is a very active social media users. Very interesting is that the number of mobile subscriptions exceeds the number of inhabitants (151%) but only a very low percentage of them (41%) is an active mobile social users. On the other hand, in Croatia the 75% of the total population, composed by 4.22 million of people, is Internet users; the 47% of the Croats is activate social media users and, also in this case, the percentage of mobile subscription exceeds the number of inhabitants (113%), while only 38% of the total population is composed by active mobile social media.

The following Country analysed is Czech Republic. The Digital Yearbook indicates that the active social media users are the 46% of the total population (10.55 million of inhabitants). Mobile subscriptions are made by the 136% of the people lived in Czech Republic, but only the 37% are active in mobile social media. The digital multimedia context for Greece indicates a total population composed by 10.91 million of inhabitants. Greeks Internet users compose are 67% and 49% of people is active social media users. In this regard, it is interesting to point out how the percentage of mobile subscriptions is equals almost twice the total population (169%), and only a very low percentage (40%) is composed by active mobile social users. The Italy Digital Yearbook shows a total population of 59.80 million of people. Between them the 66% uses Internet, the 52% is active on social media and the mobile subscriptions, also in this case, amount to almost twice the total of

---

\(^7\) https://wearesocial.com/uk/special-reports/2017-digital-yearbook

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
the population (128%), but the active mobile social users is just composed by the 47% of the Italian people. Romania Yearbook indicates 19.31 million as total population, 58% as Internet users and 49% as active social media users. Again, the number of mobile subscriptions is very high (142%), and again the number of the active mobile social users is much lower than the subscriptions (41%). At last, in the United Kingdom there are 65.31 million of inhabitants. The 92% of them use Internet, the 64% is active on social media, the mobile subscriptions amount to the 113% of people and just the 57% are active in mobile social. To conclude, according to all researches mentioned above, the Eurostat Digital Yearbook and the We Are Social-Hootsuite report, it is possible to affirm that the United of Kingdom is one of the countries with a higher multimedia level both in 2016 and in 2017. These studies have been useful in mapping a general framework about the Internet access, in figuring out the actual and past Internet utilization in terms of activity on social media platforms, but they do not tell us anything about the reason why people use Internet. Tab. 1 shows a summary of these social media characteristics.

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>POPULATION</th>
<th>INTERNET USERS</th>
<th>ACTIVE SOCIAL MEDIA USERS</th>
<th>MOBILE SUBSCRIPTIONS</th>
<th>ACTIVE MOBILE SOCIAL USERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BULGARIA</td>
<td>7.07 MILLION</td>
<td>59%</td>
<td>51%</td>
<td>151%</td>
<td>41%</td>
</tr>
<tr>
<td>CROATIA</td>
<td>4.22 MILLION</td>
<td>75%</td>
<td>47%</td>
<td>113%</td>
<td>38%</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>10.55 MILLION</td>
<td>88%</td>
<td>46%</td>
<td>136%</td>
<td>37%</td>
</tr>
<tr>
<td>GREECE</td>
<td>10-91 MILLION</td>
<td>67%</td>
<td>49%</td>
<td>169%</td>
<td>40%</td>
</tr>
<tr>
<td>ITALY</td>
<td>59.80 MILLION</td>
<td>66%</td>
<td>52%</td>
<td>128%</td>
<td>47%</td>
</tr>
<tr>
<td>ROMANIA</td>
<td>19.31 MILLION</td>
<td>58%</td>
<td>49%</td>
<td>142%</td>
<td>41%</td>
</tr>
<tr>
<td>UNITED KINGDOM</td>
<td>65.31 MILLION</td>
<td>92%</td>
<td>64%</td>
<td>113%</td>
<td>57%</td>
</tr>
</tbody>
</table>

Table 1: Country’s key digital statistical indicators
Source: Source processing on We are social and Hootsuite 2016 data
2. Methodology

2.1. Research aim

The principal aim of the research has to deal out how online hate speech against migrants can be configured on the social media among the project's partners. The research reports on the two-year period: 2016 and 2017. For the first year the study presents the OHS trend around a selected event in 2016, while for 2017 the study presents the trend of OHS in a specific time frame. The research uses several contextual data, through the consultation of different sources and specific tools such as information retrieval and statistical algorithms.

2.2. Justification of the research

A number of studies on online media hate speech have been done in several countries and across the EU, while less attention has been devoted to the online hate speech “co-produced” by media and individual users: the forums with readers’ comments supported by the online platforms of media outlets, and the social media content (with comments) of many media outlets. The increase of social media online hate speech instances against migrant and refugees in the last two-three years does not necessarily go ahead with an increase in self-regulation, legislative and law enforcement efforts. The media content analysis is complementary to a first study of the project that mapped the national context with assessment of the prevention and responses to hate speech phenomenon, analysing the legislative framework of each Country. It aims to deepen understanding the links between occurrences (like planned political events; for example elections, terrorist attacks at home and abroad, flows of refugees) and the popular reactions in terms of OHS.

2.3 Definitions used

The research does not aim to engage in theoretical debates on the definition of the term “hate speech” or debates about the tension between freedom of expression and hate speech. For the purposes of this study, and given the lack of a common international definition of hate speech, the project partners have agreed with the use of the definition proposed by the Council of Europe: “The
term ‘hate speech’ shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin”8. The research will also take into account the article 2.1 of the Additional Protocol to the Convention on Cybercrime (Budapest- 2001), which states that "racist and xenophobic material" means any written material, any image or any other representation of ideas or theories, which advocates, promotes or incites hatred, discrimination or violence, against any individual or group of individuals, based on race, colour, descent or national or ethnic origin, as well as religion if used as a pretext for any of these factors. The project partners will also use as guidelines the definition of cyber hate, and the forms and mechanisms used by those who spread or promote hate online proposed by the Anti-Defamation League (ADL) “ADL defines Cyber hate as any use of electronic communications technology to spread anti-Semitic, racist, bigoted, extremist or terrorist messages or information. These electronic communications technologies include the Internet (i.e., Web-sites, social networking sites, “Web 2.0” user-generated content, dating sites, blogs, on-line games, instant messages, and E-mail) as well as other computer- and cell phone-based information technologies (such as text messages and mobile phones).”9

2.4 Research methods

The methodological architecture was based on textual data available on the web. In the past, events such as civil wars or dictators have been characterized each country involved in the project, highlighting the past and present differences among the citizens. These differences cannot be neglected and the web could be a way for overcoming them. So, under this perspective, the selection of texts from the web could be an agglomerative criteria. Facebook and Twitter have been chosen since they represent the most used social media around the world as confirmed from the statistics realized by Alexa and SimilarWeb. This analysis indicates that Facebook is still the leader in 119 out of 149 Countries analysed in January 2017 (Figure 5).

8 https://books.google.it/Individual empowerment in the international system: Towards development, through freedom
In the world Facebook is the leader and represents the first choice of social network for 1.9 billion of users, followed by Ozone, Instagram, Twitter and in last position Linkedin with 106 millions of users (Figure 6).

Figure 5. "World Map of Social Networks, 2017".  
Source: Alexa/SimilarWeb. License. CC-BY-NC

Figure 6. "Social Media users of the world, 2017".  
Source: Alexa/SimilarWeb. License. CC-BY-NC
Another aspect of the methodology consists in studying how to obtain the data for the analysis. Therefore, the development and implementation of algorithms for information retrieval and data analysis have been a very complex and articulated part of the project. Since our interest is given from the multimedia content in terms of text (textual data), it’s necessary to point out that has been done a multilingual investigation. As far as we know, it is the first time that an investigation involves so many uncommon languages. It is a cross-country perspective, and therefore it represents an ambitious pilot project. It is important to underline the difficulties to manage so many languages that derived from different familiar languages and furthermore that include Cyrillic alphabet. The challenge is considerable and every partner proposed suggestions and requests. The aim is to reach the research objectives and not disappoint proponents' expectation. For this reason was made a process of adjustment, even if very time consuming, in some case, as for UK’s proposals, for verifying the impossibility of using a methodologies and tools proposed by the project partner and already implemented in other different projects. In addition, the IULM Team had face some technical problems, due to the storage and to the Cyrillic alphabet. A very important goal, although very challenges, was the promotion among the partners, with their active involvement, the understanding of these tools of analysis and the interpretation of the hate speech phenomenon, which is part of the problems they face on a daily basis for helping refugees and immigrants. All results provided are characterized by clarity, and coherence with the image that represent each country. This was possible because the extraction of data is geo-referenced for each country. This approach allows us to indicate how the multimedia content is referred to people that live in that given context, at a given time. The extracted data must be treated according to the same data-cleaning operations and the same weighting system in order to achieve comparable outcomes. This approach allows to obtain a structure of the results identical in each country; using the same software, same sampling in terms of extraction time and nature of sampled data; with the same weight for the unit analysis and the same type of outcomes. Then, evaluated the above-mentioned issues, in terms of limits and difficulties, the IULM Team has defined the methodology both for collecting data through information retrieval and sampling, and for the analysis. In order to map what people think and since the interest for the textual data in multimedia messages, IULM Team defined as methodology the textual analysis where, as well known, the unit of analysis is the word. The starting point of this methodology is the occurrence of the words.
The design of the research is depicted in Fig. 7 where two social media, FB and Twitter, have been engaged. As for the keywords used for the web scraping, each partner has provided them to IULM Team, based on internal criteria. The lists are twofold: one written in own language and another one in English, in order to have a framework for the evaluation.

![Figure 7. Research Design](image)

The body of techniques labelled under the information retrieval allows the identification and the selection of the data. In this analysis the IULM Team has preferred to extract data directly from an API.

API stands for Application Programming Interface (which is the set of subroutine protocols, definitions, and tools for building application software)\(^\text{10}\) consists in a set of HyperText Transfer Protocol (HTTP) request messages, along with a definition of the structure of response messages, which is usually in an Extensible Markup Language (XML) or JavaScript Object Notation (JSON). In order to extract data is mandatory to create the API both for Twitter and for Facebook. Tweets posts and messages can be filtered by keywords, users, language, and location. These data become the raw data for developing the analysis.

\(^{10}\) [https://github.com/CognitiveBuilder/HelloCognitiveWorld/blob/master/prerequisites/3-watson_api.md](https://github.com/CognitiveBuilder/HelloCognitiveWorld/blob/master/prerequisites/3-watson_api.md)
The collected data is characterized by volume, velocity, variety and variability and they can be considered as Big Data (Fig. 8).

Big data can be managed in various ways, and there are many software available. However, for this project the IULM Team had two requirements that have led to the choice of R, open source software:

1) to find a software able to analyse all languages;
2) to find a software solution available economically.

2.5. Analysis Tool

R is a language and environment used for statistical computing and graphics developed by Ross Ihaka and Robert Gentleman in the 1992 even if the first release was on 1995 11.

R is highly extensible, being an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes data handling and storage facility, operators for calculations on arrays (matrices), integrated collection of intermediate tools for data analysis, graphical facilities for data analysis and display and a well-developed programming language which includes input and output facilities.

The main package used for web scraping on Twitter is TwitterR\textsuperscript{12} and rtweet\textsuperscript{13}. The command lines has been integrated with a geocode referred to each Country through the indicating of latitude, longitude and the language for the extraction. Both the packages have implemented limits on sample size, the maximum number of tweets that the package is able to return was 3,200 for each request. These conditions are used for the extraction of both keywords and social media as indicated in the second step of L1 (cfr section, “Introduction”). This step has been organized in two separate stages in which R use was intensive for web scraping of data. In details:

- 2017: extraction of data from Twitter and from Facebook of national medias, using selected keywords. The data were collected from May, 22 to June, 3, for 2017.
- 2016: extraction of Twitter and Facebook data from national medias provided by ELIF LAB for the two months chosen by each partners.

The data collection has generated:

1. seven corpus, based on the keywords, each one has been achieved as sum of data extraction for the chosen interval in 2017, in Twitter,
2. twenty-one corpus, based on the social media, each one represents a newspaper (three newspaper x 7 countries) for 2017 in Twitter,
3. twenty-one corpus, based on the social media, each one represents a newspaper (three newspaper x 7 countries) for 2017 in Facebook,
4. twenty-one corpus, based on the social media, each one represents a newspaper (three newspaper x 7 countries) for 2016 in Facebook.

The raw data sets have a textual nature and therefore are unstructured data; that is, not encoded information. Text mining method is helpful in order to transform textual data (unstructured data) in a structured data by a process of structuring or encoding. The coding process is necessary to extract information useful for the knowledge and study of the phenomenon.

The data collection drives to the second phase of the project, L2, characterized by the use of the textual and social network analysis.

The operation of data cleaning has reduced the initial corpus and transformed the corpus in document term-matrix. Inside this matrix, each element is weighted through the weighting scheme

\textsuperscript{12} https://cran.r-project.org/web/packages/twitteR/twitteR.pdf
\textsuperscript{13} https://cran.r-project.org/web/packages/rtweet/rtweet.pdf
TF-IDF (Term frequency–inverse document frequency). TF-IDF is a measure designed to reflect the importance of a word in a document in a collection or corpus. The TF-IDF value increases in proportion to the number of times a word appears in the document, but it is often compensated by the frequency of the word in the corpus, which helps to adjust the fact that some words generally appear more frequently.

In order to understand the online hate speech (OHS) phenomenon, IULM Team has chosen to use simple data visualization, such as barplots. A barplot is a graphical representation based on stacked bars for each term in which the hits height indicates the occurrences: how many times a term is present. More a term occurs and more it is frequent.

Other types of analysis that allow data visualization are wordclouds\(^\text{14}\), bigrams, trigrams, cluster analysis. The representation of wordcloud has the same meaning of the barplot, the difference is the form of representation, which is based on depict of the words in wordcloud. The terms can present different colour and size. Bigger is the size of character and higher is the frequency of correspondent term. The words with same value of occurrences have same colour and same size.

Also, the most frequent bigrams (couple of words) and trigrams (triad of words) have been investigated in order to understand better the meaning of the sentence. If we are talking about the bigram is referring to sequence of two adjacent elements (textual element: two forms). A bigram is an n-gram for n=2. While a trigram is an n-gram for n=3 (that is three adjacent elements). It has been proposed the list or the graph bar of the most frequent bigrams and trigrams. It has been considered a clustering analysis about the words. Cluster analysis or simply clustering is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (homogeneity principle within groups) respect to the objects in the other groups (heterogeneity principle between groups). In the literature there are two main approaches for clustering algorithms: hierarchical and non-hierarchical. In this project the IULM Team has chosen the hierarchical ones and consequently its graphical representation, the dendrogram, based on Cartesian axis. The Y axis represents the distance (Euclidean distance) and the X axis are the terms. Each branch is a cluster. The line connecting two or more branches is the degree of fusion distance of the clusters. The rules of fusion are several and in this analysis is used Ward method. The Ward rule is based on minimal increment of deviation within groups. Furthermore, it has been applied to textual data the Social Network Analysis (SNA). It is a technique of analysis that through the graphs theory highlights the relationships (ties) among the statistical entities of a phenomenon. The

\(^{14}\text{Wordclouds, bigrams, trigrams, barplots, cluster analysis, social network analysis have been presented in the previous reports.}\)
textual network analysis is composed by two stages: the first stage is dedicated to the understanding of the relationship among words and the second stage is dedicated to the study of the ties generated among users. The actors or nodes, which are statistical entities, are placed in a geometric space where they are connected to each other through links. Depending on the position occupied in the geometric space and its degree of interconnection with the entire reticular structure and with other nodes, the statistical entity assumes a role, which is not only defined by the graphical intelligibility of the graph, but also by the calculation of some measures of centrality, such as degree, closeness centrality, between centrality, page rank, etc.

3. Results of Media Content Analysis

The methodology described in the previous chapter has been applied for each Country, generating these scheme:

1) Overview on national context,
2) Description of keywords choice and on-line press,
3) 2017: results of keywords selection both for Twitter and on-line press,
4) 2016: results Twitter and FB Analysis as agreed with each Partners.
5) Principal hints

3.1. Main results of the investigation in United Kingdom

The scheme used for representing UK is the following:

1. Overview on national context,
2. Description of keywords choice and on-line press,
3. 2017: results of keywords selection both for Twitter and on-line press,
4. 2016: results Twitter and FB Analysis as agreed with each Partners.
5. Principal hints
3.1.1. Overview on national context

According to the European Journalism Centre\(^\text{15}\) ‘the media landscape in the United Kingdom is complex and mature, arguably ranking second globally to that of the USA. ... The UK media landscape is a single entity, however there are distinctive English, Scottish, Irish and Welsh dimensions, reflecting the composition of the UK. The UK comprises a vast array of traditional electronic media, multi-television, multi-radio and multi-telephone’. They are worth a total of about £100bn (about 110m euros) a year.

Ofcom\(^\text{16}\) (the English Office of Communications) deals with issues relating to the media, telecommunications and postal services. It regulates TV, radio and video-on-demand sectors, fixed-line telecoms, mobiles and postal services, plus the airwaves over which wireless devices operate. Its role is to make sure that people in the UK get ‘the best from their communications services and are protected from scams and sharp practices, while ensuring that competition can thrive’ according to the powers and duties set for it by Parliament in legislation.

3.1.2. UK’s Media content analysis

In United Kingdom, the keywords selected are reported in Tab.2. The eleven words were been decided by investigating at the user-generated Facebook and Twitter posts related to the online hate speech when refer to refugees, immigrants and asylum seekers.

The UK team included terms which are strongly associated with online hate such as “fakefugees” and “gimmigrants” to mitigate for the phenomenon whereby terms are often shared by posts which are actually combatting hate speech. Research by the think-tank Demos\(^\text{17}\) has developed algorithms which can be trained to detect hateful tweets, but even after many iterations the accuracy rate does not reach 80%.

<table>
<thead>
<tr>
<th>KEYWORDS</th>
<th>Invaders</th>
<th>Fakefugees</th>
<th>Gimmigrants</th>
<th>#NoRefugees</th>
<th>Rapefugees</th>
<th>Vermin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Foreigners</td>
<td>Romanians</td>
<td>Deport</td>
<td>Migrants</td>
<td>Immigrants</td>
<td></td>
</tr>
</tbody>
</table>

\(^\text{15}\) http://ejc.net/media_landscapes/united-kingdom#link_755
\(^\text{16}\) https://www.ofcom.org.UK/about-ofcom/what-is-ofcom
\(^\text{17}\) Centre for the Analysis of Social Media, Demos, Islamophobia on Twitter: March to July 2016 Carl Miller, Josh Smith, Jack Dale.

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
The following step has regarded the multimedia content. For the multimedia contents, four English on-line press have been analysed (Tab.3).

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>TWITTER PROFILE</th>
<th>FB PROFILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE GUARDIAN</td>
<td><a href="https://twitter.com/guardian">https://twitter.com/guardian</a></td>
<td><a href="https://www.FB.com/theguardian/">https://www.FB.com/theguardian/</a></td>
</tr>
<tr>
<td>BBC NEWS</td>
<td><a href="https://twitter.com/bbcnews">https://twitter.com/bbcnews</a></td>
<td><a href="https://www.FB.com/bbcnews/">https://www.FB.com/bbcnews/</a></td>
</tr>
<tr>
<td>BRITAINFIRST</td>
<td><a href="https://twitter.com/britainfirsthq">https://twitter.com/britainfirsthq</a></td>
<td><a href="https://www.FB.com/OfficialBritainFirst/">https://www.FB.com/OfficialBritainFirst/</a></td>
</tr>
<tr>
<td>DAILY MAIL</td>
<td><a href="https://twitter.com/dailymailUK">https://twitter.com/dailymailUK</a></td>
<td><a href="https://www.FB.com/DailyMail/?nr">https://www.FB.com/DailyMail/?nr</a></td>
</tr>
</tbody>
</table>

Table 3. Monitored press on line on Twitter’s and FB’s

The framework for the selection of the media was either the market share of online news consumption, or the different editorial political position relating to refugees and migrants (neutral; left wing; right wing). News outlets were selected as the most likely transmitters of the news, debate and controversies surrounding the events.

3.1.2.1. Keywords Analysis 2017

The automated data search harvested and analysed 48,170 Tweets in a short period of time (from the 28th of May –to the third of June 2017, as has shown in Table 318. There were an estimated 15.8 million Twitter users in the UK in 201619 and over 35 million FB users20 (Table 4).

<table>
<thead>
<tr>
<th>N. OF TWEETS</th>
<th>TERMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>48170</td>
<td>foreigners, migrants, immigrants, stealing, video, Manchester, memorial, victims, bombing, flowers, norefugees, kids, America, Trump, Maga, Potus, foreign, deal, leader, Paris, attacking, proof, accord, pres, people women, attacks, fighting, Europe, deport.</td>
</tr>
</tbody>
</table>

Table 4. Most frequent terms extracted by Keywords Analysis on Twitter 2017

---

The keywords analysis indicates an important debate on the immigration topic because of the high number of tweet considering that the keywords are just eleven. Thanks to this surprising result, it is possible to identify several associations among words and analyse the most frequent words. One of the possible association could be, for example, Terrorism and war (considering words such as bombing; Manchester; Paris; memorial; flowers), Trumpism (looking to MAGA ['Make America Great Again’]; furthermore America; US; Trump; POTUS (President of the United States), Europe (Europe; EU; leaving) and Conflict. Also “war footing” vocabulary (words like attacking; victims; stealing; caught; proof). Given the above association of anti-immigrant lexical fields, it might be argued that the appearance of the word "women" is associated to the perception of misogyny among refugees.

The statistical outcomes allow to compare social media results: Guardian, BBC News and Britain first are represented in the Tab. 5.

### 3.1.2.2. Social media Twitter and FB Analysis 2017

In 2017, 1,657 tweets were selected from The Guardian, from BBC News there are 368 tweets and from British First it is possible to count 238 tweets. The cross study about the Twitter Analysis among all of the three monitored social media, reveals an interesting framework. The common thread is the Manchester attack, but there also references to the Brexit issue and to the election topic.

<table>
<thead>
<tr>
<th>DAILY MAIL</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. OF TWEETS</td>
<td>1657</td>
<td>232/1,2633</td>
<td>368</td>
<td>1141/101,297</td>
<td>238</td>
<td>200/29,554</td>
<td>654054*</td>
</tr>
<tr>
<td>N. POSTS/COMMENTS</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TERMS</td>
<td>Trump</td>
<td>Brexit</td>
<td>Attack</td>
<td>climate</td>
<td>agreement</td>
<td>238 Muslims</td>
<td>Britain</td>
</tr>
<tr>
<td></td>
<td>Review</td>
<td>Leader</td>
<td>Manchester</td>
<td>Paris</td>
<td>Islam</td>
<td>muslim</td>
<td>people</td>
</tr>
<tr>
<td></td>
<td>Election</td>
<td>School</td>
<td>Theresa May</td>
<td>President</td>
<td>Manchester</td>
<td>terror</td>
<td>man</td>
</tr>
<tr>
<td></td>
<td>Jeremy Corbyn</td>
<td>Person</td>
<td>Jeremy Corbyn</td>
<td>world</td>
<td>terror</td>
<td>migrants</td>
<td>police</td>
</tr>
<tr>
<td></td>
<td>Theresa May</td>
<td>Police</td>
<td>Paul Nuttall</td>
<td>children</td>
<td>attack</td>
<td>attack</td>
<td>horrifying</td>
</tr>
<tr>
<td></td>
<td>Manchester</td>
<td>Trump</td>
<td>Death</td>
<td>mother</td>
<td>police</td>
<td>Manchester</td>
<td>heartbreaking</td>
</tr>
<tr>
<td></td>
<td>UK</td>
<td>Britain</td>
<td>Election</td>
<td>father</td>
<td>Russian</td>
<td>signatures</td>
<td>shocking</td>
</tr>
<tr>
<td></td>
<td>Climate</td>
<td>Jeremy Corbyn</td>
<td>Debate</td>
<td>hospital</td>
<td>forces</td>
<td>activists</td>
<td>tiger</td>
</tr>
</tbody>
</table>

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
While The Guardian and BBC News are focused on politics, the Britain First underlines other issue related to religion such as terrorism, Isis and Islam. Moreover, there is an association with right-wing parties: the appearance of Paul Nuttall and UKIP indicates that news-related Tweeters are making an association between negative attitudes to refugees and extreme-right parties. Britain First tweets are predominantly associate to a xenophobic and alarmist language which includes words like death and again terrorist and Isis. For this newspaper the number of Videos detected in tweets is very high.

For Facebook Analysis it is possible to count 232 posts and 12,633 comments for The Guardian, 1,141 posts and 101,297 comments for BBC News and finally, 200 posts and 29,554 comments for Britain First and the results obtained for twitter could be the same for Facebook.

It is necessary to highlight that The Guardian and BBC News reveal the political context of the Country: words related to Brexit and to the General Election and Manchester, but also politicians like Donald Trump and Theresa May were associated to these two newspapers.

On the other hand, Britain First underlines more religious themes (migrants, Isis, Muslims and terror).

There is also an emotive language generated from words like heart breaking, shockinghorrifying.
3.1.2.3. Social media Twitter and FB Analysis 2016

According to the methodology, in 2016 analysis there was also the request to identify and select one significant event that took place during the year related to immigrants and refugees (Tab.6). The UK Team choose the referendum on the prospective withdrawal of the United Kingdom (UK) from the European Union (EU), on 23 June 2016.

Debates on immigration dominated in the months before and following the referendum, taking into account that the monitored period runs from the 23rd of May to the 15th of July 2016 February that is to say one month before and one month after the selected event.

The perspective is the same. The table below (Table 5) shows the most frequent terms in 2016 in Twitter and Facebook: there are 16,439 tweets for The Guardian, 2,369 for BBC News and 426 for Britain First. The most frequent words are referred to the political contest: Brexit, referendum, Theresa, Corbyn, Cameron , Trump, Labour, Tories. Words in The Britain First are associated with inflammatory rhetoric over the period (making associations among words such as terror, Isis, Muslim, army, Syria and refugees). The analysis of tweets in 2016 reveals many common topics with the year 2017. In particular words like Brexit and Theresa May are the same in both years. The word referendum, appears in all the three newspapers: The Guardian, BBC News and Britain First.

<table>
<thead>
<tr>
<th>TERMS</th>
<th>TWITTER GUARDIAN</th>
<th>FB 3,054/1,013,620</th>
<th>TWITTER BBC NEWS</th>
<th>FB 2,396/1,761,055</th>
<th>TWITTER BRITISH FIRST</th>
<th>FB 426</th>
<th>FB 3,395/58,451</th>
<th>FB 3,515</th>
<th>FB 5,490/1,252,617</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theresa May</td>
<td>right</td>
<td>Theresa May</td>
<td>Euref</td>
<td>muslim</td>
<td>immigrants</td>
<td>women</td>
<td>foreigners</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donald Trump</td>
<td>Europe</td>
<td>Referendum</td>
<td>Brexit</td>
<td>referendum</td>
<td>police</td>
<td>British</td>
<td>illegal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People</td>
<td>jobs</td>
<td>David Cameron</td>
<td>Europe</td>
<td>Brussels</td>
<td>islamists</td>
<td>Euref</td>
<td>Trump</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour</td>
<td>Countries</td>
<td>Daily Mail</td>
<td>People</td>
<td>action</td>
<td>vermin</td>
<td>Europe</td>
<td>stop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>David Cameron</td>
<td>vote</td>
<td>Telegraph</td>
<td>kill</td>
<td>Country</td>
<td>Boris Johnson</td>
<td>Britain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>immigration</td>
<td>Guardian</td>
<td>warms</td>
<td>invaders</td>
<td>leadership</td>
<td>France</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vote</td>
<td>campaign</td>
<td>The Sun</td>
<td>Syria</td>
<td>London</td>
<td>labour</td>
<td>refugees</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2016)

With the Facebook analysis The Guardian shows 3,054 posts and 1,013,620 comments regarding political themes like referendum, vote, campaign but also terms like immigrant and foreigners; BBC News 2,396 posts and 1,761,055 comments which include words like immigrants, foreigners, vote, questions, Euref, Brexit, Europe, and finally Britain First indicates 3,395 posts and 584,521 comments with a highest level of hostility towards the immigrants; because its most frequent words are: deporting, Muslim, Sharia, vermin, invaders, attack, Nice, Islamism, migrants.

The data collected is referred to over a two-month period surrounding the key event, the EU referendum in the UK on 22 June 2016.

The people talk of terroristic attacks, Muslims and Islamic religion, political agreements, terror, Brexit, Theresa May and Donald Trump in 2017 and in 2016.

Twitters for Daily Mail also propose figure like Theresa May, Jeremy Corbyn, David Cameron and furthermore vote and referendum is present such as the most frequent terms and so the political context. FB for Daily Mail 2017 permits to observe the most social and cultural context because of the presence of words like immigrants, migrants, deports, foreigners, illegal and therefore the migration issue and the hate speech.

If you consider the period of the event selected, June 2016, it is possible to affirm that political and public speeches in the months previous and following the referendum were dominated by debates on immigration. There is a clear correlation between the EU referendum and the volume of social media posts overall. To demonstrate that the online speech was growing during the selected event, it is possible to indicate a barplot (about The Guardian FB Analysis) showing the number of posts over the selected period of time. The peak is relative to the day after the EU referendum, on the 23th of June 2016.
Figure 9. Barplot of daily counting of keywords for The Guardian FB Analysis 2016

For BBC News FB data there are two positive peaks: one after the EU referendum and another one after the terrorist attack in Nice (France) on the 14th of July 2016:

Figure 10. Barplot of daily counting of keywords for BBC News FB Analysis 2016

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
32
The daily counting of keywords for the FB account of Britain First reveals the peak on the 27th of June and therefore few days after the EU referendum on the 22th of June. However, the date indicates that the peak is referred to another event connected to EU referendum, because some extremists burned the England flag as a symbol of protest (Fig. 12)

Figure 11. Barplot daily counting of keywords for Britain First FB Analysis 2016

Figure 12. Barplot daily counting of keywords for Daily Mail FB Analysis 2016

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
3.1.3. Principal hints

The different barplots above shows that in reference of the selected event there is an effective growth of OHS. There are several peaks distribution:

- after the EU referendum at the 23th of June 2016
- after the terrorist attack in France on the 14th of July 2016
- another event connected to the EU referendum in which some extremists burned the England flag as a symbol of protest.

Regarding Facebook and Twitter, both in 2016 and in 2017 great attention has been paid to political issues, immigration, politics leaders like Theresa May or Jeremy Corbyn and Trump, and alto to religious themes: muslims, islamists.

These conclusions will be confirmed also in the chapter 4, when the profile of each country is declined through the reading of some International rankings of the social, economic and political characteristics of every Country in the World.

3.2. Main results of the investigation in Italy

The scheme used for this paragraph is the following:

1. Overview on national context,
2. Description of keywords choice and on-line press,
3. 2017: results of keywords selection both for Twitter and on-line press,
4. 2016: results Twitter and FB Analysis as agreed with each Partners.
5. Principal hints

3.2.1. National context

Blogging is very popular in Italy. Social-networking sites, especially FB and Twitter, have become important tools for self-expression. The degree of press freedom is determined by pooling the experts replies to questionnaires evaluated media independence, media environment and self-censorship, legislative framework, transparency, and the quality of the infrastructure that supports...
the production of news and information. In 2016, more than 30 journalists were currently receiving police protection in Italy. The level of violence against reporters and politicians is alarming (including verbal and physical intimidation and threats), especially because of the pressure they make on journalists who opt to censor themselves. The broadcasting regulatory body AGCOM (Authority for the Guarantees in the Communications) is subjected to pressure from the government, politicians and large corporate interests. According to the statistics of the UNAR, warnings of hate incitement discourses in the media (including Internet) represent the 34.2% of the whole reports received by the UNAR in 2013, in comparison to the 19.6% in the 2012. For what concern the legislative context. Since “the refugee issue” took a central position on the media agenda in Italy in the last years, the media have often helped to create a negative representation of the refugee as ‘enemies’.

### 3.2.2. Italy’s Media content analysis

In order to support automated social media data search algorithm, the Italian members of the Positive Messengers project have proposed ten keywords according to their daily professional experience and studies and researches on the topic. The identified keywords, referred to immigrants and refugees can be grouped according to the following characteristics: words which contain neutral nouns (such as “migrants”, “islamists”), words containing derogatory or pejorative nouns (such as “people wearing flip-flops”, “clandestine”, “they’re robbing our jobs”), adjectives related to the OHS targets (such as “fake refugees”, “Maghreb people”). The project team opted to select both a representative sample of words already identified as frequent regarding refugees and migrants as online hate speech targets, as well as certain diversity since the automated algorithm not only analyses words frequency but also the relation of selected words to other words. For this reason the word “taking our jobs” was found associated with “NOT” and not alone. In this case the meaning is opposite to the expression of hate against immigrants and refugees. The choice of ten keywords for Italy includes (Table 7).

<table>
<thead>
<tr>
<th>KEYWORDS</th>
<th>ENGLISH TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ciabattanti</td>
<td>people wearing flip-flops</td>
</tr>
<tr>
<td>Profugo / profughi</td>
<td>Refugees</td>
</tr>
<tr>
<td>Migrante / migrant</td>
<td>Migrants</td>
</tr>
<tr>
<td>Immigrato / immigrati</td>
<td>Immigrants</td>
</tr>
</tbody>
</table>
The same approach is used for the choice of the media outlets. Therefore three online media outlets, particularly active on Twitter and FB accounts had been selected for the automated social media data search. The market share of online news consumption and the different editorial political position have guided the rationale behind the choice to refugees and migrants (left wing; catholic area; right wing). The objective was to subject to automated search social media channels of media that are interested in the refugee theme instead of ignoring it.

The Italian research team provided the following (Table 8):

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>TWITTER PROFILE</th>
<th>FB PROFILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>La Repubblica</td>
<td><a href="https://twitter.com/repubblica">https://twitter.com/repubblica</a></td>
<td><a href="https://www.FB.com/Repubblica/">https://www.FB.com/Repubblica/</a></td>
</tr>
<tr>
<td>Avvenire</td>
<td><a href="https://twitter.com/avvenire_nei">https://twitter.com/avvenire_nei</a></td>
<td><a href="https://www.FB.com/avvenire.it/">https://www.FB.com/avvenire.it/</a></td>
</tr>
</tbody>
</table>

According to the same logical perspective the results of the statistical social media content analysis for Italy are shown to follow.

**3.2.2.1. Keywords Analysis 2017**

<table>
<thead>
<tr>
<th>KEYWORDS EXTRACT THROUGH TWITTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS</td>
</tr>
<tr>
<td>YEAR=2017</td>
</tr>
<tr>
<td>TERMS</td>
</tr>
</tbody>
</table>

Table 9. Most frequent terms extracted by Keywords Analysis on Twitter
The extraction of the ten keywords from Twitter is equal to 2,149 in a period of time from the 28th of May 2017 to the 3rd of June 2017. The Keywords Analysis allows us to understand the whole political contest. This analysis reveals the most frequent terms like "First the Italians" (one of keywords for the extraction), Gianfranco Fini (he is an Italian politician belonging to centre right), immigrants, "fake refugees" (other keyword for the extraction), public housing, Matteo Salvini (leader of Lega Nord), support, project, Lega, illegal immigrants, law, government. These words represent a picture of the actual Italian situation showing who are the main political figures. Everyday, the Italian media content presents a section that talks about immigration, disembarks of illegal immigrants, public housing assignment that views to give the priority to immigrants. Therefore the motto: "First the Italians". The Fini political figure is connected to the law “Bossi-Fini”. In fact On the 11th of July 2002 the Italian government passed by decree Law No. 177, the "Bossi Fini" law, introducing criminal sanctions for people caught who illegally came to the country or who returned after being expelled. Under this law an immigrant who is stopped without a residence permit will be accompanied to the border and expelled immediately. Immigrants are also subjected to arrest and detention from six to twelve months, which is followed by the immediate deportation if caught attempting to re-enter Italy before the expiry of a re-entry ban. A second offence is punishable by up to four years imprisonment. The residence permit for immigrants has been strictly linked to a work contract. The time of seclusion in detention centres whilst waiting for extradition has been extended from thirty days to sixty days and asylum seekers will be placed in detention while awaiting asylum review, in contravention of Article 5 of the European Convention on Human Rights which states: "Everyone has the right to freedom and security of the person". Finally it is possible to notice the presence of the politician Matteo Salvini, leader of the party called “Lega”. It could also be highlighted that the most used devices for tweeting are predominantly smart-phones (smart-phones with Android and with IOs): for example, social media users wrote “Italian people first” primarily through Twitter for Android (299) and Twitter for Iphone (144). This seems to be coherent with the data analysis of the National context which reports the high number of smart-phones in Italy compared with other EU Countries.

3.2.2.2. Social media Twitter and FB Analysis 2017
Table 10 shows for the year 2017, the twitter analysis indicates 101 tweets for Il Giornale, 1,606 tweets for La Repubblica and for Avvenire there are 254 tweets. The first consideration emerged from the analysis of the most frequent words is that there are two social media that reveal the political aspects of the country and another one that is less representative. Il Giornale has been characterized by: Donald Trump, Berlusconi, Alfano (Minister for Foreign Affairs and International Cooperation), Electoral law and "No Boldrini (President of Chamber of Deputies) government". La Repubblica shows terms like Electoral law (such as Il Giornale), Rome, Paris, Milan, climate, agreements, Renzi. Avvenire presents words like: Pope Francesco, consultation, Church, life, work, dead, hope, world, children, God, family. Contrariwise Avvenire is concentrate towards religious and humanitarian issues. It is the proof that Il Giornale has a very political cut as well as La Repubblica. For FB Analysis we can count 42 posts and 2,467 comments for Il Giornale, 312 posts and 12,891 for La Repubblica and finally for Avvenire we have 41 posts and 839 comments. FB Analysis indicates that, the topics proposed from Il Giornale on Twitter are not the same proposed on the FB profile. In 2017 comments were supported by video, photo and links to several blogs. There is a recall to the terrorism because of words contain feelings of terror far from the political

Table 10. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2017)

<table>
<thead>
<tr>
<th>NEWSPAPERS</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL GIORNALE</td>
<td>101</td>
<td>42/2,467</td>
<td>1,606</td>
<td>312/12,891</td>
<td>254</td>
<td>41/839</td>
</tr>
<tr>
<td>LA REPUBBLICA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVVENIRE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Terms:
- Donald Trump
- Berlusconi
- Alfano
- Electoral law
- No Boldrini government
- Video
- Photo
- Company
- Blog
- words
terror
world
- electoral
law
- Rome
- Paris
- Milan
- climate
agreements
- Renzi
- Republic
Italy
- Business
- Pope
- Francesco
- history
- State
- economy
- insiders
TV
Control
- Pope
- Francesco
- hearing/consultation
- Church
- life
- work
- dead
- hope
- world
- children
- God
- family
- Pope
- reflection
dialogue
interview
community
people
mass
world
themes of the Country. La Repubblica shows terms like Republic, Italy, Business, Pope Francesco, history, State, Economy, invaders. Therefore, we have not total political aspects but several themes that include economic and religious topics.

3.2.2.3. Social media Twitter and FB Analysis 2016

The event that has been selected to analyze the year 2016 is not linked to immigrants. In fact, the Italian Team wanted to verify if an unforeseeable event, that not concerns immigrants people could originated the OHS or in a some way lead to the migratory issue.

![Earthquake information]

Source: Italian Team description of the event chosen.

The selected event is the earthquake that affected part of central Italy. The number of victims was particularly high in the affected areas, which were inhabited by a small number of residents, were at that time at the peak of tourist season, causing a higher number of victims. Indeed, the earthquake caused 299 deaths, 388 injured and 4,500 homeless (in the area around Accumoli – Rieti). The Italian intuition is not so wrong, indeed the earthquake in central Italy can be considered a crucial event for comments referring to immigrants. The Italian people feel the necessity to underline the emergency of the evacuees. The multimedia content indicates the necessity to put the focus on the immigrants’ flows (Table 11).
Table 11. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2016)

For the year 2016, Il Giornale presents 817 tweets, La Repubblica 19,971 tweets and Avvenire 1,136 tweets.

For the year 2016, the Twitter Analysis indicates, also in this case, that the newspapers Il Giornale and La Repubblica were concentrated on the political aspects of the Country. However, it is important to underline that the most important topic is the earthquake occurred in Umbria. In fact, the term earthquake is most frequent in all three newspapers. Il Giornale shows as the most
frequent terms Renzi, migrants, referendum, Virginia Raggi (Rome Mayor), Isis, Merkel, Germany, government, USA, Donald Trump. So, despite the earthquake, politics themes are in first position among the news of the day. The newspaper La Repubblica follows Il Giornale with terms like Renzi, Rome, USA, dead, Virginia Raggi, terrorism, France. Both newspapers are focused on national and international political aspects.

On the other hand, Avvenire, talks about religious and humanitarian issues: Pope, hearing, earthquake, world, dead, God, prayer, migrants, Church, Syria. But certainly highlights factors connected to migratory phenomenon putting words like migrants and Syria.

For the year 2016, Il Giornale presents 781 on line posts and 468,586 comments, La Repubblica presents 7,811 posts and 866,073 comments and Avvenire shows 987 posts and 10,039 comments. FB analysis highlights terms such as illegal immigrants, immigrants, Islamic, terrorists, Italy, government, Muslims, Christians, faithless for Il Giornale. La Repubblica indicates: immigrants, illegal immigrants, Islamic, people, Italy, migrants, territory, Country, first the Italians. Avvenire shows: terrorists, immigrants, illegal, Islamic, Church, Christians. FB analysis in 2016 underlines an important thread among social media and newspapers, talking about the same topics (Fig. 13).

![Barplot of daily counting of keywords for Il Giornale FB Analysis 2016](image)

*Figure 13. Barplot of daily counting of keywords for Il Giornale FB Analysis 2016*
There is more than one-distribution peak. One of these is referred to the end of July because of the murder of a French person by the hand of an organization related to Isis. Other peaks are on August 23rd, August 24th and August 25th.

The correspondence to the selected event is showed in the barplot below (Figure 6), that represents the Facebook analysis of La Repubblica. On the contrary, Il Giornale shows people talking about electoral law, Trump and Boldrini. In 2016 in spite of the earthquake, politics is on first place topics with migrants, referendum, Isis, Libya. Same situation for La Repubblica in which the words founded on Twitter were: electoral law, agreements, Renzi and for FB they are: economy, Pope Francesco (Fig.14).

![Figure 14. Barplot of daily counting of keywords for La Repubblica FB Analysis 2016](image)

La Repubblica shows in a clear way a peak related to the to selected event in a multimodal distribution. The use of the selected ten keywords is low during the monitoring period of time and on the 24th of August there is a considerable increase of words (Figure 15).
The same appraisal is possible for Avvenire in which the distribution of the words is fragmented due to the lack of references on the searched topics.

The FB analysis highlights terms such as illegal, immigrants, Islamic, terrorists, Italy, government, Muslims, Christians, faithless for Il Giornale. La Repubblica, indicates: immigrants, illegal, Islamic, people, Italy, migrants, territory, Country, “first the Italians”. Avvenire shows: terrorists, immigrants, illegal, Islamic, Church, Christians.

The FB analysis of 2016 is very interesting because it underlines an important relation among the three newspapers. All of them address issues related to immigration.

In 2016 there was a great hostility against immigrants.

A final interesting remark for the Italian case is the absence of the word “taking our jobs”. This word appears twice preceded by NOT. In other words, it does not appear as Hate speech, but on the contrary, it is in favour of immigrants and refugees.

### 3.2.3. Principal hints

The Keywords Analysis for 2017 allows us to understand the whole political contest. In fact Il Giornale newspaper and La Repubblica newspaper talk about political themes and for this reason the words related to these two newspapers are: Trump, USA, Germany, Salvini.
On the contrary, Avvenire has a more social mode and presents words like Pope, earthquake and God.

The event selected to analyze the year 2016 is not linked to immigrants in order to verify if an unforeseeable event, that does not concern immigrants people could originate the OHS or in some way lead to the migratory issue.

The results are the same of the previous years, thus it is possible to affirm that, despite the earthquake, politics themes are always at the bottom of the daily news. Also in 2017 Il Giornale and La Repubblica paid attention to international political aspects and Avvenire talks about religious and humanitarian issues.

### 3.3. Main results of Bulgaria’s investigation

The scheme used for this paragraph is the following:

1. Overview on national context,
2. Description of keywords choice and on-line press,
3. 2017: results of keywords selection both for Twitter and on-line press,
4. 2016: results Twitter and FB Analysis as agreed with each Partners.
5. Principal hints

#### 3.3.1. National context

According to the Council for Electronic Media and the National Statistical Institute - 2014, the number of active media outlets in Bulgaria is as follows: print: 295 newspapers, 635 magazines; radio stations: 85; television stations: 112. The total annual newspaper circulation is 324,310,000. The media environment in Bulgaria, as the other post-communist countries, has dramatically deteriorated since they joined the EU. The “the refugee issue” took a central place in the Bulgarian scenario in September 2013, event highlighted by the media which have contributed to create a negative consideration about immigrants. Sova Harris, polling agency, affirms that at the beginning of 2016 the 60% of the Bulgarian citizens considered the refugees as a threat to the national security. Government complaints, politicians, and large corporate interests pressure the broadcasting regulatory body, affecting the solution of problems such as

---

*Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931*
hate speech\textsuperscript{21}. The ECRI Bulgaria Report concludes that the method in place for violations of relevant legislation sanctions is ineffective, and the Recommendation 17 proposes to encourage the Council for Electronic Media to take part in all the cases of dissemination of hate speeches. The 58\% of the respondents in 2016 declared they witnessed hateful statements towards members of minority groups and refugees. The hatred or verbal aggression against Muslims is growing rapidly - from 10.6\% in 2014 to 38\% in 2016. Television still remains the most influential media by which people associate the spread of hate speech – three quarters of the respondents who during the last year had heard hate speeches had heard it on television, the report said. The combination of catastrophic, sensational news and hateful language has been turned into a commercial media practice since it provokes reactions from the readerships and boosts advertising.

3.3.2. Bulgaria’s Media content analysis

The Bulgarian project team has identified the following ten words (Table 12) in order to apply the automated social media data search algorithm:

<table>
<thead>
<tr>
<th>KEYWORDS</th>
<th>ENGLISH TRADUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>мигрант, мигранти</td>
<td>migrant, migrants</td>
</tr>
<tr>
<td>бежанец, бежанци</td>
<td>refugee, refugees</td>
</tr>
<tr>
<td>заплаха</td>
<td>Threat</td>
</tr>
<tr>
<td>терорист, терористи</td>
<td>terrorist, terrorists</td>
</tr>
<tr>
<td>исламист, исламист</td>
<td>Islamist, Islamists</td>
</tr>
<tr>
<td>нелегален, нелегални</td>
<td>Illegal</td>
</tr>
<tr>
<td>талибани</td>
<td>Talibans</td>
</tr>
<tr>
<td>на сапун</td>
<td>turned into soap, brutal expression “turn Gypsies/Arabs”</td>
</tr>
<tr>
<td>чернилка</td>
<td>nigger, offensive for a person with dark skin</td>
</tr>
<tr>
<td>вън</td>
<td>out</td>
</tr>
</tbody>
</table>

\textit{Table 12. Table about Bulgarian keywords}

The ten keywords have been identified based on a reviewed national research and media monitoring studies occurred from 2016 and 2017 and related to online hate speech, channels, targets, messages, refugees and immigrants:

\textsuperscript{21} Freedom of the Press 2016 – Bulgaria, \url{http://www.refworld.org/docid/582ac6dbf.html}

\textit{Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931}
• The survey of the Mission Salvation Foundation\textsuperscript{22} of fourteen anti-refugee and xenophobic FB groups, listing the most frequent offensive qualifications of refugees and migrants (“Talibans, traitors, aliens, slugs, parasites”);

• “Representation of Roma and Refugees in the Bulgarian Online Media (Monitoring of the Media Coverage during the Official Election Campaign 24 February 2017 – 24 March 2017”, Teodor Spasov\textsuperscript{23}, pointing out the semantic circles of online anti-refugee rhetoric linked with national disasters and calamities, “plague of grasshoppers”, “influx of invaders”, “contract killers”, “non-humans”, “sexual assaults”);

• “Refugees. Media image and Cultural projections”, Milko Petrov, which outlines several phases of online representation of refugees: from personal drama stories through fear and juxtaposition, to geopolitics and the interests of the great powers, to Islamic terrorism and trafficking\textsuperscript{24}.

The choice of the online media outlets with active Twitter and FB accounts depends on the market share of the online news consumption, and the different editorial political position to refugees and migrants (neutral; left wing; right wing). The project team supported by the Advisory Board decided to base their online media selection on the popular appeal measured by number of visitors/followers (Table 13).

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>TWITTER PROFILE</th>
<th>FB PROFILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOVA.TV</td>
<td>twitter.com/NoviniteNaNova</td>
<td>FB.com/Nova.bg/</td>
</tr>
<tr>
<td>NOVINI.BG</td>
<td>twitter.com/Novini.bg</td>
<td>FB.com/Novini.bg</td>
</tr>
<tr>
<td>BLITZ.BG</td>
<td>twitter.com/blitz.bg</td>
<td>FB.com/blitz.bg</td>
</tr>
</tbody>
</table>

\textit{Table 13. Monitored social media on Twitter’s and FB’s profiles}

\textsuperscript{22} Mission Salvation.
\textsuperscript{23} See T. Spasov, op. cit., \url{http://www.aej-bulgaria.org/bul/p.php?post=8112&c=328}

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931  
46
Several media have a very well implemented Facebook accounts but have not a Twitter account and were therefore not appropriate for the purposes of the automated research algorithm. Between the 1th and the 10th of March 2017 the most popular 28 media outlets with available online content and their social media profiles were reviewed: televisions, radio stations, newspapers, news agencies and purely online news portals. In addition, data from “Gemius” metrics available at www.audience.bg was used as well.

### 3.3.2.1. Keywords Analysis 2017

The Bulgarian Keywords Analysis indicates 182 selected tweets on those themes in 2017 for the monitoring period from May, 28 to June, 3 (Tab.14). The main topics are: migrants, traffic, airport, Sofia, threat, terrorist, level, critically, Great Britain, Rumen Radev (Bulgarian Major General of the Reserve and the current President of Bulgaria since January 2017), Country, Islamic, Putin, sea, Mediterranean, illegal, police officers, borde, refugees, danger, Islamists, Europe, security, arrested, Erdogan (Recep Tayyip Erdoğan: current President of Turkey), terrorist, Manchester.

<table>
<thead>
<tr>
<th>KEYWORDS EXTRACT THROUGH TWITTER 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS</td>
</tr>
<tr>
<td>TERMS</td>
</tr>
</tbody>
</table>

*Table 14. Most frequent terms extracted by Keywords Analysis on Twitter*

These words are the expression of a climate of terror, fear, danger and of the terroristic perception. The topics are referred to Islamism, security, police, danger, arrests, and illegality. The fear condition is confirmed by the Manchester and London attacks presence between the most frequent terms. The relevance of this international episode is a constant trend in a
frequency study of the words about much investigated Nations. Furthermore, a considerable political framework becomes increasingly evident due to the presence of political figures like President of Bulgaria (Rumen Radev), President of Turkey (Erdogan) and Vladmir Putin.

3.3.2.2. Social media Twitter and FB Analysis 2017

Table 14 reports the number of tweets and posts/messages obtained through the webscraping phase. The Twitter Analysis for 2017 counts 4,999 tweets for Nova Tv, 5,000 tweets for Blitz.bg and 767 for Novini.bg. Nova tv presents most frequent terms as vote, guest, children, game, house, peaceful, transmission, deal. The contextual analysis is so fragmented that does not permit to understand the social and political climate of the Country. However It is possible to suppose a request of a peaceful climate both in political terms (thinking of the elections) and in social terms (thinking of the several terrorist attacks in Europe).

It is also possible to indicate a danger transmission of viral sickness from the immigrants. Blitz.bg analysis presents a different framework of Bulgaria showing political figures like Putin, Trump and Hillary Clinton. The words “elections” and “President” indicate how the contextual situation is addressed to a probable political propaganda and elections. In addition also the international context is present through terms like Russia, Germany, Syria, Mosul (Iraqi city) and Aleppo (city of northern Syria) or Borisov (Belarusian city). These words indicate an argumentative climate on the terrorism and on violent actions by Isis or other extremist groups. Hostility is more evident because of the presence of the word “Islamic”. Novini.bg through a survey, represents the political aspects thanks to figures like Trump, Rumen Radev and Putin,. In addition, issues such as terrorist attacks, Borisov (Belarusian city) and the Manchester attack that indicates the common thread of the entire transnational investigation emerge on their Twitter account.

<table>
<thead>
<tr>
<th>NEWSPAPERS</th>
<th>TWITTER NOVATV</th>
<th>FB 2017</th>
<th>TWITTER BLITZ BG</th>
<th>FB 2017</th>
<th>TWITTER NOVINI BG</th>
<th>FB 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMENTS</td>
<td>4,999</td>
<td>77/2,149</td>
<td>5,000</td>
<td>45/464</td>
<td>767</td>
<td>120/1,032</td>
</tr>
</tbody>
</table>

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
Table 15. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2017)

Table 15 shows that the 2017FB analysis for Nova tv indicates 77 posts and 2,149 comments, for Blitz.bg 45 posts and 464 comments and finally for Novini.bg 120 posts and 1,032 comments. The FB Analysis does not permit to define a contextual framework in terms of immigration, of hostility against immigrants or in terms of political aspects.

3.3.2.3. Social media Twitter and FB Analysis 2016

For the Analysis of the 2016 in Bulgaria, the Bulgarian Team selected one of the cases of “refugee hunters” as a first example of providing a platform for hate speech. This case concerns a group of citizens self-organized as militia to catch and arrest illegal immigrants in the mountain of Strandzha, close to the Turkey border. The para-military group was first reported on bTV on the 18th of February 2016 as the story of Dinko Valev from Yambol who has captured several times the refugees on the border with Turkey. In May, Dinko and his group “arrested” 3 young migrants.

That action has been registered with a video became viral in which the group leader Dinko openly expressed a hostile attitude and intent to protect the country from immigrants. Between March and May 2016, Nova TV published more than 50 information units regarding migrants and refugees. The word “migrant” was mentioned 48 times, “refugees” 47 times, “terrorists” 30 times, and “threat” 61 times. These results indicate that at least once a day there was a connection between migrants/refugees and threat. The news coverage of Bulgarian events is
more detailed: the activity of Dinko (the selected event) was reported in 12 information units and 2 “My news”, provided by viewers. The readers/viewers’ comments are highly divided pro and against Dinko. During the monitored period in 2016 Blitz published 13 materials about the selected event (almost identical with Nova TV), which, however, were read 156,624 times and encouraged 857 readers’ comments. The most popular articles with over 10,000 readers have sensational headlines, containing slang and offensive language in the form of a quote.

Table 16 shows 433 tweets for Nova Tv in 2016, 1606 tweets for Blitz.bg and 3818 tweets for Novini Bg always in 2016. The Twitter analysis of Nova Tv presents terms like Brussels, children, refugees, displacement, Borisov (Belarusian city), protest, home, anthrax, died, police. There is a strained atmosphere as demonstrate the presence of word like refugees, protest, anthrax, died, police. In particular, anthrax is a well-known bacterial disease of sheep and cattle, typically affecting the skin and lungs. It can be transmitted to humans, causing severe skin ulceration or a form of pneumonia (also called wool-sorter's disease).

In the Novini.bg are mentioned cities and Countries like Borisov (Belarusian city), Brussels, Germany, Russia, Greece, Europe, and people like Ministers and children. There are also references to crashes and attacks.

It is possible to observe that Brussels attack represents an important event that involved the European Countries during the 2016.
<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>TWITTER NOVATV</th>
<th>FB BLITZ BG</th>
<th>TWITTER NOVINI BG</th>
<th>FB NOVINI BG</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMENTS</td>
<td>433</td>
<td>1,078/11,836</td>
<td>1,606</td>
<td>2,016/11,005</td>
</tr>
</tbody>
</table>

**TERMS**
- Theme
- Sofia
- Brussels
- Bulgaria
- children
- woman
- refugees
- displacement
- Borisov (Belarusian city)
- Easter
- money
- cars
- protest
- home
- anthrax
- died
- bus
- police
- souls
- airport
- danger
- red
- picture
- profile
- cut
- spooks
- numbers
- buddy
- silent
- coast
- anonymous
- Bulgarian
- migrants
- English
- place
- return
- hands
- blends
- back
- swinging
- voices
- West
- Native
- Swear
- authorities
- encourage
- Arrests
- Civilian
- terrorist
- Country
- Atlantic
- Bosporus
- presence
- Target
- categorical
- Military
- Syria
- Bulgarian
- journalist
- Train
- migrants
- terrorists
- students
- Europe
- People
- Bulgaria
- Border
- Children
- Sofia
- Press
- Bulgaria
- Borisov (Belarusian city)
- Brussels
- Minister
- Russia
- Europe
- Children
- Woman
- Leading
- People
- Germany
- Easter
- coat of arms
- President
- Police
- Greece
- Crash
- Attacks
- Souls
- migrants
- migration
- maintenance
- scandalously
- Bulgaria
- terrorists
- people
- attacks
- Dinko Valev Committee
- Court
- questioned
- judged
- appeared
- attitude
- stormy
- Helsinki
- complaint
- application
- Brussels
- Europe
- Turkey
- police
- security

*Table 16. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2016)*
However, Novini.bg also highlights other aspects, not national, through the use of words such as Russia, Germany, Europe, and Greece. The Facebook Analysis for 2016 comprehends 1,078 posts and 11,836 comments for Nova Tv, 2016 posts and 11,005 comments for Blitz.bg and 3,838 posts and 169,240 comments for Novini.bg. The counting of occurrences reveals concepts not related to a social and political scenario. Just one term expresses a social consideration with the word “migrants”. Blitz.bg indicates more closeness to the immigration phenomenon. There are words like arrests, terrorists, Syria, migrants, Europe, border which reveal a dangerous perception of the social climate full of terrorism and migrants. Novini.bg analysis is interesting in terms of the dangerous perception of the migration phenomenon in European Countries. There are words like migrants, terrorists, Dinko Valev, people, judged, attacks, complaint, police, Brussels, Turkey, Europe, security. The security issue is evident and so like the Brussels bombings and the terroristic situation. The topics in 2016 were stronger compared to those belonging to the analysis of the year 2017. Hostility against the immigrants is focused on issues such as refugees, terrorism, danger, attacks and security. On the other hand the FB analysis for 2017 does not indicate the same speeches. The climate has absolutely changed. FB reveals profound changes in the Bulgarian society. Novini.bg represents the only media able to capture the figure of Dinko Valev, which indicates the main person of the event selected in 2016. In order to check how the incitement to online hatred has increased or decreased during the selected event or just before or shortly after; it could be necessary to consider the Barplot of the post count and comments. As a prerequisite, the research has shown a clear diffusion between online and offline. The 2016 event identified for refugees and migrants and the two-month period observed allowed us to see some dynamics in the online hate speech (Figure 16).

![Figure 16.Barplot of daily counting of keywords for Nova Tv FB Analysis 2016](image)

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931

52
Within the distribution is possible to individuate the first peak in relation to the Brussels bombings. The event is occurred on the 22th of March. The second peak is on the 11th of April and so in correspondence of the selected event about the phenomenon of "refugees hunters". The same consideration in terms of the biggest number of keywords utilization around the selected event is possible for the Novini.bg.

For Novini.bg the first peak is on the 22th of March and so precisely the date of Brussels bombings. The second peak is referred to the period of the "refugees hunters" phenomenon. Therefore, the highest bar in the distribution is on 22 April (Figure 17).

![Barplot of daily counting of keywords for Novini.bg FB Analysis 2016](image)

Figure 17. Barplot of daily counting of keywords for Novini.bg FB Analysis 2016

To confirm the concordance of the online hate speech around the selected event is also the analysis done for the Blitz.bg. The first peak of words in on the 22th of March for the Brussels bombings and the second peak is on the 22th of April (Figure 18). The initial hypotheses are totally confirmed, according to some important and popular events condition the online hate speech phenomenon.
3.3.3. Principal hints

The 2017 analysis for Nova.Tv newspaper shows words related to several social aspects like children, vote or peaceful, for the Blitz.bg it is possible to read something more related to politics such as President, Putting, and the same is possible for the Novini.bg with the words “attack”, "Manchester", "Trump" and “freedom”.

In 2016 the selected event comes from one of the cases of the “refugee hunters”, first example of providing a platform for hate speech. This case concerns a group of citizens self-organized as militia to catch and arrest illegal immigrants in the mountain of Strandzha, close to the Turkey border. For this year there are three very relevant peaks in correspondence of:

- Brussels bombings, event occurred on the 22nd of March;
- The selected event about the phenomenon of "refugees hunters" (11th of April)
- The "refugees hunters" phenomenon (22nd of April).
3.4. Main results of Romania’s investigation

The scheme used for this paragraph is the following:

1. Overview on national context,
2. Description of keywords choice and on-line press,
3. 2017: results of keywords selection both for Twitter and on-line press,
4. 2016: results Twitter and FB Analysis as agreed with each Partners.
5. Principal hints

3.4.1. National context

According to the daily media consumption routine, TV and Internet are the most used media channels. As said by Manuela Preoteasa and Andrei Schwartz: “Media owners becoming politicians, politicians connected to media outlets, imprisoned media owners, prosecuted media outlets, tax evasion, blackmail, and a nearly bankrupt state TV station, but also award-winning investigative media centers and a vivid online community of investigative journalists and media civil society organizations – Romania has them all.”

Although the attitude towards free movement of citizens seems to be positive, in fact only 36% of Romanians appreciate the contribution made by immigrants and, among them, a large part prefers immigrants who come from other Europe States. Two out of five Romanians consider that measures to combat illegal immigration of people outside the EU should preferably be taken at national level well above the European average of 24% but also increasing compared to 2015 (24%). From the Romanian point of view, immigration and terrorism are very important problems to be faced, although the number of those who spoke about immigration and terrorism is decreasing compared to 2015, when the 43% told that terrorism was a problem and the 47% talked about immigration as a real problem. One of the most relevant findings of the “Report on the perception of Romanians about the refugee crises” case study is represented by the response saying that if only 2.5% of Romanian citizens with Internet access would expel a foreigner belonging to a European country, 18% would expel a refugee.

3.4.2. Romania’s Media content analysis

Romanian project team selected the ten keywords listed in Table 17:

<table>
<thead>
<tr>
<th>KEYWORDS</th>
<th>ENGLISH TRADUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Islamiști</td>
<td>Islamists</td>
</tr>
<tr>
<td>Musulmani</td>
<td>Muslims</td>
</tr>
<tr>
<td>Migrant/Migrațiți</td>
<td>Migrant/Migrants</td>
</tr>
<tr>
<td>Refugiați</td>
<td>Refugees</td>
</tr>
<tr>
<td>Teroriști</td>
<td>Terorists</td>
</tr>
<tr>
<td>Cioară/Ciori</td>
<td>Crow/Crows</td>
</tr>
<tr>
<td>Tigan</td>
<td>Gypsy</td>
</tr>
<tr>
<td>Poponar</td>
<td>Fag</td>
</tr>
<tr>
<td>Jidan</td>
<td>Jew</td>
</tr>
<tr>
<td>Bozgor</td>
<td>can not translate it is a rude word for Hungarians (hatefull name for Hungarians)</td>
</tr>
</tbody>
</table>

Table 17. Table about Romanian keywords

For the choice of the ten keywords it was necessary the contribution of journalists, members of the Project Advisory Board project was necessary, in addition to national research and the monitoring of social media in 2016 and 2017.

The monitoring includes the search of the keywords related to online hate speech, channels, targets, messages, refugees and immigrant.

To select the ten keywords related to the issue of immigration, was used a list containing the words most often used with the aim of inciting online hate speech, as "migrants" or "refugees", rather than as "terrorists", "gypsy" and "bozgor". The Advisory Board consists of three experts with different knowledge in different areas that are human rights, freedom of expression, hate speech, media monitoring, training and education.

In order to select the online media outlets with active Twitter and FB accounts for the automated social media data search it has been used the market share of online news consumption.

The three online media selected for the analysis through the use of social platform such as Facebook and Twitter Social are reported in Table 17 and they are: Adevărul, TV Antena 3 and PROTV (Table 18).
3.4.2.1. Keywords Analysis 2017

The analysis of Romanian Keywords on Twitter for the 2017 shows 89 tweets (Table 19).

It is possible to group the most frequent words as follows:

- Names of Countries/States (Bulgaria, Europe, Germany, Manchester, Turkey)
- Words related to Religion (islamists, Muslims, Ramadan, gypsy, terrorists)
- Bad words/bad themes (dead, offenders, arrested, police, tragedy)
- Words related to immigration (refugees, save, sea)
- Political themes (Süleyman Soylu – a Turkish politician, leaders).

These results could reflect a tension climate because of the presence of words like Islamist, migrants and refugees, terrorism and gypsy. In addition to the six themes indicated above, it is possible to divided the list of the words in three other semantic areas:

- geographical (“Mediterranean” “sea”)
- geopolitical (“Germany”, “Polonia”, “Bulgaria”, “Manchester”, “Europe”)
- and descriptive verbs (“death”, “attempt”, “saved”, “arrested”).

Furthermore, it is always present the Manchester attack which seems to have profoundly shaken the European scenario because of almost all monitored Countries reveals high values connected to this event. The devices used for tweets production are in the more case FB (40 tweets produced from this source), Twitter Web Client (13) and Google (7 tweets for this source).
3.4.2.2. Social media Twitter and FB Analysis 2017

The 2017 Twitter Analysis includes 16 tweets for Pro Tv, 200 for Adevarul and 1150 for Antena 3 (Tab.19). The monitoring period of time is from the 28th of May to the 3rd of June.

2017 ProTv shows frequently words of political figures like Donald Trump, Hillary Clinton, Iohannis Klaus (Romanian President), Prime Minister, and also related with countries (Romania, Country), and society like live, dead, people.

Adevarul expresses the national political framework through figures like Klaus Iohannis and Liviu Dragnea (Romanian politician). There are also references to some aspects related to the international context: Madrid, Emmanuel Macron and Trump. The European politics is involved thanks the political personality of Emmanuel Macron (France President) and Donald Trump.

Antena Tv is the media with a lower number of terms than the others media but the words obtained with the analyses are related to the immigration phenomena (death and refugees). Also in this case the most frequent issues are also religious (Islam).

Many tweets also contain the term “video” and so it is possible to suppose the presence of a video of a particular event that became viral.

The 2017 FB analysis shows for Pro Tv seven posts and 1,310 comments and this time the topics are not referred to the immigration theme or a racist climate and not permits to deep into the political situation.

The FB analysis is different from the Twitter Analysis for the same social media. In other words, the Twitter analysis indicated the political national and international context counter wise the FB
analysis does not permit a contextual study. Adevarul counts 15 posts and 1,370 comments. The topics are again policies, related to immigration, social but it doesn’t explain the political or social context but is not present any form of racism; however, the context is ascribable with concepts like death, State and Trump (Table 20).

<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>16</td>
<td>7/1,310</td>
<td>200</td>
<td>15/1,370</td>
<td>1150</td>
<td>131/7,335</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TERMS</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donald</td>
<td>Trump</td>
<td>Hillary</td>
<td>Clinton</td>
<td>Iohannis</td>
<td>Klaus</td>
<td>Live</td>
</tr>
<tr>
<td>Trump</td>
<td>Hillary</td>
<td>Clinton</td>
<td>Iohannis</td>
<td>Klaus</td>
<td>Live</td>
<td>Dead</td>
</tr>
<tr>
<td>Hillary</td>
<td>Trump</td>
<td>Hillary</td>
<td>Trump</td>
<td>Hillary</td>
<td>Klaus</td>
<td>Live</td>
</tr>
<tr>
<td>Clinton</td>
<td>Iohannis</td>
<td>Klaus</td>
<td>Live</td>
<td>Dead</td>
<td>People</td>
<td>Prime</td>
</tr>
<tr>
<td>Iohannis</td>
<td>Klaus</td>
<td>Live</td>
<td>Dead</td>
<td>People</td>
<td>Prime</td>
<td>Minister</td>
</tr>
<tr>
<td>Klaus</td>
<td>Live</td>
<td>Dead</td>
<td>People</td>
<td>Prime</td>
<td>Minister</td>
<td>Little</td>
</tr>
<tr>
<td>Live</td>
<td>Dead</td>
<td>People</td>
<td>Prime</td>
<td>Minister</td>
<td>Little</td>
<td>Wounded</td>
</tr>
<tr>
<td>Dead</td>
<td>People</td>
<td>Prime</td>
<td>Minister</td>
<td>Little</td>
<td>Wounded</td>
<td>Romania</td>
</tr>
<tr>
<td>People</td>
<td>Prime</td>
<td>Minister</td>
<td>Little</td>
<td>Wounded</td>
<td>Romania</td>
<td>Country</td>
</tr>
<tr>
<td>Prime</td>
<td>Minister</td>
<td>Little</td>
<td>Wounded</td>
<td>Romania</td>
<td>Country</td>
<td></td>
</tr>
<tr>
<td>Minister</td>
<td>Little</td>
<td>Wounded</td>
<td>Romania</td>
<td>Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little</td>
<td>Wounded</td>
<td>Romania</td>
<td>Country</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wounded</td>
<td>Romania</td>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 20. Most frequent terms extracted by Twitter and FB Analysis: year 2017

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
59
3.4.2.2. Social media Twitter and FB Analysis 2016

Also in this case it was important to identify an event: the selected event for Romania was one of the most mediated and commented cases of incitement and racial hatred took place in early April 2016 in Bucharest. Two Muslim high school girls were assaulted by a group of young people because of the traditional Muslim veil they wore. The incident occurred in a busy area from Bucharest (the Moșilor area). Few days after the incident, the witnesses were refractory to admit they had witnessed an assault. The incident did not have serious physical consequences, the aggressors only confined to pulling the scarf. However, according to the press, one of the two girls went to the police to report the incident, but no formal complaint was filed. It seems, however, that in a press release, the representatives of the District 2 District Prosecutor's Office told the press that they opened a file for offenses and other violence. The trigger event is the intention to build a mosque in Bucharest. The conditions for its realization has exacerbated the attention of the national press and of course the attention of the readers. The reaction of the population was totally negative, also thanks to the way the news presented the event, it means associating Muslims with terrorists. This event has attracted the attention of the media and many users of the web and therefore proves to be a very useful event in order to monitor the hate phenomena that are activated during an event and to test the peaks immediately after the event or immediately before it.

During the research time, the Adevărul newspaper published 200 articles containing one of the keywords selected for the analysis, ProTv broadcasted 59 video and Antena3 television, contained 107 materials. It is interesting to note how in reality the most virulent comments and posts were triggered by the way the news were presented. However, one of most important announcement came after the Brussels bombings on the 22nd of March 2016. The event has been presented after an important episode: Brussels bombings. The Analysis of the content for the year 2016 is summarized in Table 21: 86 tweets for Pro Tv, 2,423 for Adevarul and finally for Antena 3 we have 3,572 tweets.

The point of concordance among all the three social media is the Brussels bombings, the event told above and occurred on the 22th of March 2016, because of its presence in each occurrence calculation. Pro tv has been characterized by terms referred to countries (Romania, Brussels, Country), criminality (prison, died, criminal) and by the word “euro”. Adevarul also shows the same words with the addition of Brussels as country, and of terms like money, people, children,

27 https://positivemessengers.net/images/library/pdfs/MCA-1.2-RO-eng.pdf
Iohannis, For Antena 3 the most frequent terms are died, Ioannis Klaus, accident, Bucarest and again Brussels.

<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADEVARUL</td>
<td>2,423</td>
<td>18/71,492</td>
<td>4,600</td>
<td>3,572</td>
<td>10,118/260,396</td>
<td></td>
</tr>
<tr>
<td>ANTENA3</td>
<td>86</td>
<td>92/18</td>
<td>2,423</td>
<td>4,600/71,492</td>
<td>3,572</td>
<td>10,118/260,396</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TERMS</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romania</td>
<td>hour</td>
<td>Dead</td>
<td>shameless</td>
<td>罗马尼亚</td>
<td>Story</td>
<td>罗马尼亚</td>
</tr>
<tr>
<td>Brussels</td>
<td>follow</td>
<td>odd</td>
<td>ruthless</td>
<td>布鲁塞尔</td>
<td>photo</td>
<td>布鲁塞尔</td>
</tr>
<tr>
<td>announcement</td>
<td>criminal</td>
<td>head</td>
<td>expedite</td>
<td>公布</td>
<td>Bucarest</td>
<td>страны</td>
</tr>
<tr>
<td>died</td>
<td>fight</td>
<td>structures</td>
<td>regime</td>
<td>show</td>
<td>Life</td>
<td>refugees</td>
</tr>
<tr>
<td>prison</td>
<td>ruthless</td>
<td>people</td>
<td>policy</td>
<td>money</td>
<td>migrants</td>
<td>host</td>
</tr>
<tr>
<td>plane</td>
<td>children</td>
<td>money</td>
<td>people</td>
<td>Romans</td>
<td>Muslims</td>
<td>Muslims</td>
</tr>
<tr>
<td>euro</td>
<td>degree</td>
<td>people</td>
<td>children</td>
<td>Country</td>
<td>Muslims</td>
<td>Romans</td>
</tr>
<tr>
<td>Country</td>
<td>family</td>
<td>world</td>
<td>Iohannis</td>
<td>refugees</td>
<td>establishment</td>
<td>people</td>
</tr>
<tr>
<td>Rome</td>
<td>corruption</td>
<td>Europe</td>
<td>months</td>
<td>attacks</td>
<td>disaster</td>
<td>world</td>
</tr>
<tr>
<td>place</td>
<td>investigation</td>
<td>blogs</td>
<td>Brussels</td>
<td>accident</td>
<td>Bucarest</td>
<td>images</td>
</tr>
<tr>
<td>dollar</td>
<td>Europe</td>
<td>euro</td>
<td>Iohannis</td>
<td>euros</td>
<td>euros</td>
<td>euros</td>
</tr>
</tbody>
</table>

Table 21. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2016)

All the three newspapers try to reproduce the national situation talking about Ioannis Klaus, a political figure of reference, and presenting the accident that occurred. The accident is, probably, referred to assault of the Muslim girls. The 2016 Twitter Analysis indicates a context characterized
by politics and criminality presenting words like “died” or “criminal” and so it is clear how the climate is not so positive.

The 2016 Facebook Analysis report 92 posts and 18 comments for Pro Tv, 4,600 posts and 71,492 comments for Adevarul, and finally 10118 posts and 260,396 comments for Antena 3.

The content analysis of FB for Pro tv is a little bit different from the previous one and includes very bad words like “shameless”, “fight”, “ruthless” but also more social terms as “children” and political words likes “regime”, “policy”, “expedite”, “corruption”, “investigation”.

So it is again possible to see that the climate is not positive both the social one and the political one, but fortunately there aren’t any racist affirmation. The Adevarul not indicate a so negative climate but proposes terms related to the immigration phenomenon with the two word “migrant” and “refugees”. On the contrary Antena 3, indicates the important topics linked to immigrants: refugees, attacks, immigrants, Muslims, Merkel, ISIS, Iohannis Klaus, terrorists. In a spite of a comparison among three selected media Antena 3 represents the most indicative ones in terms of hate speech than f the three-monitored social media.

The case study event relevant to refugee and migrants from 2016 observed during February 23 – April 23 clearly confirms this situation (Fig. 19).

Indeed, considering the barplot above (Figure 20) shows the counting of posts and comments per day, it is possible to observe that the distribution peak is referred to the two successive days of Brussels bombings (22 March) because the highest bar is on the 24th of March. Furthermore, it is
important to consider that the selected event indicated by the Romanian Team took place on the 24th of March 2016.

The confirmation of the impact of the selected event derives also by observing the bar graph of Antena 3, which shows the first peak on the 22nd of March. This date represents the suicide attacks on the Brussels airport and subway in which died 32 people and injured hundreds.

The successive days we have a little decrease but not less important. Indeed, on the 23rd of March there was the declaration of building of mosque in Bucharest.

Same consideration for Adevarul (Fig. 21). The peak is referred to the Brussels bombings on the 22nd of March. However, there is not a peak on 23 and 24 March in relation of building of mosque in Bucharest. This situation indicates a difference compared to the other two social media (Pro tv and Antena 3). There is another peak on 28 March, that is some days after the episode of mosque and the Brussels bombings. It is interesting, also, to observe the second important peak on the 12th of April, that is 12 days after the assault of Muslim girls happened on 30 March.
3.4.3. Principal hints

The 2017 analysis shows as social media expresses the national and international political aspects talking about immigration, politics and social aspect “through” words like children or “life”, words always accompanied by video.

The 2016 selected event took place on the the 23rd of April in Bucharest and it was the opening of a new mosque.

This event has provoked the wrath of the citizens also thanks to the way the news presented the event, it means associating Muslims with terrorists.

For Pro tv There were bad words like “shameless”, “fight”, “ruthless” but also more social terms as “children” and political words likes “regime”, “policy”, “expedite”, “corruption”, “investigation”.

The Adevarul indeed proposes terms related to the immigration phenomenon (“migrant”, “refugees”). The same thing could be said for Antena 3.

The two peaks of chatting and posting were related to the of building of mosque in Bucharest (23rd of March) and of the suicide attacks on the Brussels airport (22nd of March).

The following days we have a little decrease but not less important. There was another peak on the 28th of March, that is some days after the episode of mosque and the Brussels bombings.
3.5. Main results of Czech Republic’s investigation

The scheme used for this paragraph is the following:

1. Overview on national context,
2. Description of keywords choice and on-line press,
3. 2017: results of keywords selection both for Twitter and on-line press,
4. 2016: results Twitter and FB Analysis as agreed with each Partners.
5. Principal hints

3.5.1. National context

The topic of migration begins to emerge in Czech Republic in 2015 and then intensifies rapidly. According to some public opinion polls conducted by the Center for Public Opinion Research (CVVM), Median and STEM and others, most Czechs perceive migration in rather negative way. The 60 % of respondents said that the Czech Republic should not accept refugees from war-torn regions and one-third supports their acceptance until they're able to return home (CVVM data survey). Only 3 % think that the Czech Republic should accept refugees. The qualitative research conducted by Anthropictures focused on information sources and how the Czech public perceived them. It confirmed that media play an important agenda-setting role. The focus on the issue of migration was dealt with online through social media and discussion forums (such as the news portal parlamentnlisty.cz, together with idnes.cz and to a certain extent novinky.cz), determining the spread of dangerous and often untrue online hate speeches. Desperate were the attempts to maintain a certain online standard and not violate ethical and civil codes, even eliminating bad and unjust comments. For example the Seznam Zprávy (Seznam News) online media outlet has recently introduced an innovative feature: readers newly cannot comment to published articles in writing, but have to record their comments on a maximum 60-second long video. They all provide an important and useful insight into the state of online hate speech in the Czech Republic.

28 “Kontexty a kořeny percepcí migrace".
However, it is good to underline that 2015 is the year characterized by the most intense political debate linked to the theme of migration, surely the situation has changed in the following years even if the debate has remained very heated. 29

3.5.2. Czech Republic’s Media content analysis

Table 22 shows the ten selected keywords for the automated social media data search and analysis for Czech Republic, keywords linked to online hate speech:

<table>
<thead>
<tr>
<th>KEYWORDS</th>
<th>ENGLISH TRADUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uprchlíci</td>
<td>refugees</td>
</tr>
<tr>
<td>Migranti</td>
<td>migrants</td>
</tr>
<tr>
<td>Islám</td>
<td>islam</td>
</tr>
<tr>
<td>Islamisté</td>
<td>islamists</td>
</tr>
<tr>
<td>Terorismus</td>
<td>terrorism</td>
</tr>
<tr>
<td>Pražská kavárna</td>
<td>Prague coffee club – the term for Czech/ Prague intellectuals, admirers of Vaclav Havel and human rights defenders</td>
</tr>
<tr>
<td>Vlastizrádci</td>
<td>traitor to own country</td>
</tr>
<tr>
<td>Sluničkáři</td>
<td>“Shiny happy people” – pejorative term for migrant defenders</td>
</tr>
<tr>
<td>Pravdoláškaři</td>
<td>m for Vaclav Havel admirers, it is very often use to ridicule intellectuals and migrant defenders</td>
</tr>
<tr>
<td>Multikulturalismus</td>
<td>multiculturalism</td>
</tr>
</tbody>
</table>

Table 22. Table about keywords

Also in this case the experts’ opinion was necessary regarding the online migration-related discourse to select ten words, and also a literature review and secondary data review, including, among others:

- Hate Speech in the Online Environment and Social Networks (People in Need, 2015);
- Hordes of Scum, Niggers and Primitives are Coming to Overtake us, HateFree Culture, Office of the Government, 2016;
- Hate Speech Online: Internet Discussions in the Czech Republic (Multicultural Centre Prague, 2015);

29 https://positivemessengers.net/images/library/pdfs/MCA_1.2_CzR-eng.pdf
• Media Campaign Against Racism Based on Hate (HateFree Culture, Office of the Government, 2016);
• Fairies, Pansies and Dykes on the Czech Web (HateFree Culture, Office of the Government, 2016);
• An Analysis of Media Coverage of the Refugee Crisis (Masaryk University, 2015)
• Why Do Refugees Stir Up Our Emotions? Migration Narratives in Czech Society and a Glance Beyond Them (Glopolis, 2016);
• Homos, Hodgies and Darkie Czechs Online (Coordinated by: HateFree Culture, Office of the Government, 2016).

The selected ten keywords include some characteristics typically associated with migrants such as "Islam", "Islamist (s)", "terrorism", "sluněčkáři", "pravdoláskáři" various names for groups associated with liberal opinions on migration - "welcome" stereotropically depicted those who would unconditionally accept any number of refugees / migrants.

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>TWITTER PROFILE</th>
<th>FB PROFILE</th>
</tr>
</thead>
</table>

*Table 23. Monitored social media on Twitter’s and FB’s profiles*

Furthermore, the research indicated the selection of three online media outlets with active Twitter and FB accounts (Tab. 23).

According to the way the words deal with the problem that leads to the attribution of words to three types of polarity: negative (against), positive (pro) and neutral (measured by number of followers). In case of the Czech Republic, whether a specific outlet tends to be neutral, pro, or against the issue was determined based on an analysis of the headlines of main Czech daily outlets conducted by Jiří Linhart in the fall of 2015 between 1 September and 30 October 2015. The study has been shown the following key words: “refugee”, “immigrant”, “migrant” and their Czech varieties - “běženec”,

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
“utečenec”, “přistěhovalec”. Again, as seen for Romania, the way in which the news are told by the media influences the common thought and the analysis classifies the dailies' headlines according to whether they give a neutral, positive, or negative impression about migration. Headlines are classified as “negative” if they express criticism of the system of “quotas”, call in question the refugees' benefit, legitimacy/legality of their stay and fairness of their conduct, include claims regarding the dangers/damages they bring about or evoke negative emotions: fear, helplessness, oppression of the in-group. Headlines are classified as “positive” if they refer to the benefits of immigration, contain an appeal to help or evaluate such behaviour positively, invoke compassion, defend the EU relocation system of “quotas” or criticize negative views of migration. Since most headlines are classified as “neutral” in case of all media outlets, the sample contains the three social media in the Table above.

3.5.2.1. Keywords Analysis 2017

The 2017 Keywords Analysis for Czech Republic comprehends 268 tweets during the monitoring period of time (Table 24). The content analysis point out words related to:

- Religion (Islam, islamists, terrorism)
- Politics (Zeman (Czech President Miloš Zeman), speech)
- Cities (Prague, Mosul, Lidicích)
- Organizations (ONU)
- Feelings (existence, fleeing)
- Immigration (people, migrants)
- Bad words (cowards, attack, waring)

<table>
<thead>
<tr>
<th>TERMS</th>
<th>268</th>
<th>YEAR=2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Islam, people, Zeman (Czech President Miloš Zeman), speech, cowards, Mosul, Lidicích, migrants, Prague, islamists, terrorism, existence, attack, ONU, waring, fleeing.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 24. Most frequent terms extracted by Keywords Analysis on Twitter*

30 Jak psaly české deníky o uprchlické krizi*.
Therefore, it supposed the terms like Zeman, speech, cowards and Lidicich. Furthermore, it is present the Mosul city; city of Iraq since has been protagonist of several attacks. Finally, the Keywords Analysis highlights the terroristic attacks and the UN working for peace. However, the friction towards immigration is tangible.

### 3.5.2.2. Social media Twitter and FB Analysis 2017

<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMENTS</td>
<td>7</td>
<td>14/189</td>
<td>150</td>
<td>100/2,382</td>
<td>170</td>
<td>161/2,385</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TERMS</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>astrophysicist</td>
<td>Not available</td>
<td>States</td>
<td>crowns</td>
<td>smoking</td>
<td>President</td>
<td>decision</td>
</tr>
<tr>
<td>universe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>complaint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Republic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blow up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interrupted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forbid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exception</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fade away</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 25. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2017)

The 2017 media content analysis (table 25) starts from the Twitter analysis. Lidové Noviny shows 7 tweets, Hospodářských Noviny indicates 150 tweets and Blesk presents 170 tweets. Counting the occurrences it is possible to find terms, which demonstrate a not racist climate. Hospodářských Noviny is not connected to words related to immigrants, but certainly indicates an interest in politics with words like minister, Macron (French president), protocol and Emmanuel. There are also words related to the economic situation such as money, companies and finance. The 2017
Facebook analysis indicates for Lidové Noviny 14 post and 189 comments, for Hospodářských Noviny 100 posts and 2,382 and for Blesk presents 161 posts and 2,385 comments. The results for the Twitter analysis show the same results of the Facebook Analysis and both of them indicate a non-hostile climate towards immigrants. Hospodářských Noviny presents few political factors not indicative for this study about the online hate speech. The same consideration is possible for the content analysis about Blesk. It could be that the media context has caused and influenced the results that do not show particular attention for migrants. First of all it is necessary to underline the low number of tweets, comments and posts, a number that is much higher in the other monitored countries. This result is perhaps driven by the small monitoring period.

This information coincides with the official statistics presented in the First chapter. Indeed, Czech Republic indicates a not too high penetration of media users (46% of penetration and so 4.9 million of people on 10.55 million of total population). There is a standard deviation of 5.65 social media users between the population and the people that use the social media when connected to Internet.

3.5.2.3. Social media Twitter and FB Analysis 2016

The choice of an event is necessary to understand how the incitement to hate online changes before and after the selected event. For Czech Republic the project team selected an event occurred on August 2016 (on the day of the 1968 Warsaw Pact invasion of Prague anniversary) when a Czech anti-Muslim and anti-refugee group called “We don't want Islam in the Czech Republic” staged a fake Isis terrorist attack at the Old Town's square, a popular tourist destination. The group, including the head of the event, the populist rightist Martin Konvička, drove to the scene on trucks dressed as jihadists, while waving the Isis flag and fake submachine-style BB guns. Having involved the firing of fake weapons and shouts of “Allahu Akbar”, the demonstration caused panic before the police managed to abort the “performance”.

Lidové Noviny shows 40 tweets, Hospodářských Noviny presents 825 and Blesk indicates 826 tweets.

The 2016 Twitter Analysis for the social media outlets not reveals such as crucial aspects relative to immigration: Kolín and Rýnem [Cologne; a symbol of the perceived predatory nature of refugees due to sex attacks on New Year's Eve of 2015 in Cologne] (also appearing in the bigram analysis), “Samková [a Czech public figure opposing migration/Islam] and the words like “ageing” and “threatening”. Hospodářských Noviny, so like for the Twitter Analysis 2016, is focused on economic aspects through words, such as: companies, boss, million, billions, money, Prague, State,

31 Fake ISIS attack in Prague by anti-immigrant protesters sparks panic | The Independent".
police, world, Court, dollar, government. Also, the Twitter Analysis for the 2016 not reveals themes linked to immigrants.

The FB Analysis for 2016, registers for Lidové Noviny 329 posts and 81 comments, for Hospodářských Noviny 1,298 posts and 36,264 comments and for Blesk 1,378 posts and 28,805 comments. Therefore, it is not surprising not to find the topics related to hate speech against the migratory phenomenon. However, Lidové Noviny's outcomes indicate political factors, perhaps linked to the immigration question. The most frequent terms are: people, interest, unlocking, National, dispute, conformity, politically and Merkel.

<table>
<thead>
<tr>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIDOVÉ NOVINY</td>
<td>HOSPODÁŘSKÉ NOVINY</td>
<td>BLESK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMENTS</td>
<td>40</td>
<td>329/81</td>
<td>825</td>
<td>1,298/36,264</td>
<td>826</td>
</tr>
<tr>
<td>TERMS</td>
<td>Rýnem Kolin (district of the Czech Republic)</td>
<td>people interest unlocking National dispute spectrum conformity politically square airport names search theater Merkel</td>
<td>flight companies boss million billions money Prague State police world Court dollar government Germany found out threaten asylum countries problem conductors society freight</td>
<td>Not available</td>
<td>refugees Africa connect place simulated occupation August army migrants world Isis Miloš Zeman Europa Islam people</td>
</tr>
</tbody>
</table>

Table 26. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2016)
This data it could be confirmed through the bar graph of daily counting the posts and comments. (Fig. 22).

Figure 22. Barplot of daily counting of keywords for Lidové Noviny FB Analysis 2016

The posts and comments distribution is not so high. The unique peak is on the 21st August as shown in the barplot (Figure 23) While Lidové Noviny indicates Angela Merkel, Hospodářských Noviny mentioned Germany and therefore, it is possible to suppose a ties to the topic during the monitored period. Hospodářských Noviny presents, in addition, found out, threaten, asylum, countries, problem, society, freight.

Hospodářských Noviny on 1,298 posts and 36,264 comments proposed 99 documents containing the selected keywords. However, through the bar graph of the daily counting we can observe that there is not the peak during the selected event. The highest bar within the distribution is referred to another event; on 31st July. The interesting observation is that on 21st August. The interesting observation is that on 21st August, that is the selected event, the height of the bar not indicate a peak and so the selected keywords utilization.
Finally, the analysis of Blesk FB account indicates on 1378 posts and 28805 comments documents containing the keywords. Blesk represents, in absolute, the social media that permits to see the immigration aspects. The most frequent terms are: refugees, Africa, connect, occupation, army, migrants, Isis, Miloš Zeman (President of the Czech Republic), Europa and Islam. (Fig. 24).
Data gathered from the FB account of Blesk demonstrate that main peaks in the distribution is referred to others event and not to the selected event. The two main peaks are on August 26th and 28th. Around the selected event the value of occurrence is relatively higher compare to other days but it is sure that the peak is not in correspondence on 21st August. The two media events that on the FB page of Blesk provoked a more robust response (that contained the selected keywords) was a) news regarding criminal prosecution of individuals who publicly approve of an act of terrorism and b) news from two days before regarding a photomontage of a fake advertisement published on the Block Against Islamization FB account.

### 3.5.3. Principal hints

The 2017 analysis show how the most frequent words could be grouped as follows:

- Religious themes (Islam, islamists, terrorism)
- Politics themes (Zeman (Czech President Miloš Zeman), speech)
- Cities (Prague, Mosul, Lidicích)
- Organizations (ONU)
- Feelings (existence, fleeing)
- Immigration (people, migrants)
- Bad words (cowards, attack, warning)
- terms which demonstrate a not racist climate.

The migration area is not so predominant because of the influence of the media context on the individual capacity: it determines the way of thinking of citizens who therefore do not perceive the problem of immigration.

The event chose for the 2016 analysis occurred on August 2016 (on the day of the 1968 Warsaw Pact invasion of Prague anniversary) when a Czech anti-Muslim and anti-refugee group called “We don't want Islam in the Czech Republic” staged a fake Isis terrorist attack at the Old Town's square. The choice it is necessary understand how the incitement to hate online changes before and after the selected event. For Czech Republic the project team selected an event. However, the two main peaks are on August 26th and 28th. Around the selected event the value of occurrence is relatively higher compare to other days but it is sure that the peak is not in correspondence on 21st August.
3.6. Main results of Croatia’s investigation

The scheme used for this paragraph is the following:

1. Overview on national context,
2. Description of keywords choice and on-line press,
3. 2017: results of keywords selection both for Twitter and on-line press,
4. 2016: results Twitter and FB Analysis as agreed with each Partners.
5. Principal hints

3.6.1. National context

The Croatia media contextual analysis reveals an important issue linked to media freedom. According to the World Press Freedom Index compiled by Reporters Without Borders (RSF), in 2016 Croatia was downgraded from place 54 to 63 (out of 180 countries) in comparison to the previous year. In 2017 Croatia has been ranked 74 with the score equal to 29.59, which means media freedoms were even more aggravated. The hate speech rhetoric is directed especially towards ethnic Serbs, Roma, refugees and migrants, LGBTI persons and women. According to Gemius Audience, 10 most popular online media channels are 24sata.hr, vecernji.hr, net.hr, dnevnik.hr, rtl.hr, tportal.hr, express.hr, telegram.hr, novilist.hr and direktno.hr.

In January 2016 there has been a protest march in which 7,000 people took part. The march induced by the newly elected Vice-President of the Parliament Ivan Tepes was a protest march against the decision took by the Electronic Media Council (EMC) to suspend broadcaster Z1 three days for airing hate speech directed at ethnic Serbs. On the 10th of March 2016 the government decided to reject the annual CEM report, to revoke the agency's mandate and dismiss its director. The decision of the Parliament led the director of the EMC, Marijana Rakic, to resign on charges of "unbearable pressures." In those days this event was the focus of the media attention: the restrictions on the material deemed as source of hate speech had been applied to print and to other broadcast media.
While the freedom of expression of the private media was very high and the government seemed not to interfere, the journalists complained about the lack of transparency on the part of the government. In fact, a great number of journalists complained about the presence of strong complaints to avoid reporting negatively on advertisers or those politically linked to key advertisers. On August the Union of Electronic Media urged the government to provide for HRT's independence following the September parliamentary elections, saying it was "essential that the HRT be allowed to maintain its editorial independence to provide impartial and relevant news, especially in these important weeks leading up to the elections in Croatia. Freedom of the media is critical to a healthy democracy". Another hostile initiative linked to the Croatian nationalists was linked to the news of a group of veterans based in Split have brought against Novosti journalists Boris Dezulovic, Victor Ivancic and Milorad Krstulovic. The complaint was based on a series of articles that the newspaper published in 2016. Newspapers were accused of inciting hatred and in particular of "promoting intolerance towards the Croatian people and the Croatian state." In February, an ultra-conservative nationalist group called "In the Name of the Family" held a news conference outside parliament and launched a petition for Novosti's closure. Other organizations have come out in support of Novosti and have condemned the newspaper's stigmatization by nationalists as "enemy of the state". Furthermore, another example of problems with freedom of media is the HRT case. Recently, HRT decided to cancel the show "Hrvatska uživo" ("Croatia live"), without any explanation, after 13 years of broadcasting. The show dealt with topics regarding minorities, people with special needs, culture, education, etc. Some sources claim that the show and its main editor Maja Sever were too critical about the government and the state bodies and had been warning about irregularities of the system too often, so the state television decided to shut it down.

3.6.2. Croatia’s Media content analysis

Table 27 shows the ten keywords referred to immigrants and refugees selected by the Croatian Team.

Through comprehensive research of media outlets and readers’ comments, the Advisory Board has selected 10 words, which are predominantly associated to online hate speech. It was challenging task, as Croatian language includes several rude words.

34 Available at: https://rsf.org/en/news/croatian-nationalists-want-serbian-minority-newspaper-close
The consulted source is the “Report on observing hate speech, discriminatory, stereotypical and inflammatory speech in the news media in the August of 2016” (Dražen Hoffman, Zagreb, 7 September 2016)\textsuperscript{35}.

The selected keywords referred to immigrants and refugees can be grouped as follows:

- neutral nouns (such as “refugees”, “kriza”)
- negative nouns (such as “balije”, “kozojebi”)

The Croatian Team opted to select both groups of words as well as words not only linked to refugees and migrants due to high presence of hate speech towards national minorities in Croatia, especially Serbs and Roma ("Četnici", "ustaše"). The Advisory Board- formed by Ms Cvijeta Senta, Project coordinator in Centre for Peace Studies, expert in Anti-Discrimination Policies, Hate crime, Hate/discriminatory speech; Ms Maja Munivrana, associate professor at the Faculty of Law in Zagreb and Ms Natalija Havelka, legal adviser, expert in provision of free legal assistance to marginalized population e.g. women, children, persons with disabilities, members of Roma population and migrants- has agreed that context plays a crucial role in identifying hate speech because it ensures that limitations on freedom of expression remain justifiable in a free and democratic society.

\begin{table}[h]
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{KEYWORDS} & \textbf{ENGLISH TRADUCTION} \\
\hline
izbjeglička kriza & refugee’s crisis \\
terori & terrorists \\
kozojebi & very rude word goat fuckers \\
ustaše & members of Nazi collaborators – Croats during the 2nd World war in Croatia \\
pederi & rude word for gay population \\
balije & rude word for Muslims \\
četnici & members of Nazi collaborators – Serbs during the 2nd World war in Croatia \\
antife (derisive term for liberal or left-wing members of community) & derisive term for liberal or left-wing members of community \\
bodljikava žica & razor fence \\
izbjeglice & refugees \\
\hline
\end{tabular}
\caption{Table 27. Table about Croatian keywords}
\end{table}

\textsuperscript{35} Available at: http://www.gong.hr/media/uploads/2016_08.pdf
According to the GONG\textsuperscript{36} report the online hate speech was monitored in a period of time from October 2015 to September 2016, period in which a lot of articles and sentences are in dnevno.hr, which has repeatedly shown the tendency to discriminate with intoned, humiliating and insulting speech, displaying non-European refugees and migrants as criminals and rapists, and affirming homophobic attitudes\textsuperscript{37}. Other media outlets chosen for the analysis have been also included in the GONG report as subjects in which journalistic presentation in discourse of social hazards and dangers have been experienced in different periods, more or less explicitly, by national minorities, LGBTIQ population, refugees and migrants.

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>TWITTER PROFILE</th>
<th>FB PROFILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNENO.HR</td>
<td><a href="https://twitter.com/portaldnevno">https://twitter.com/portaldnevno</a></td>
<td><a href="https://www.FB.com/PortalDnevno.hr/">https://www.FB.com/PortalDnevno.hr/</a></td>
</tr>
<tr>
<td>DIRECKTNO.HR</td>
<td><a href="https://twitter.com/direktnohr">https://twitter.com/direktnohr</a></td>
<td><a href="https://www.FB.com/direktno.hr/">https://www.FB.com/direktno.hr/</a></td>
</tr>
<tr>
<td>MAXPORTAL.HR</td>
<td><a href="https://twitter.com/maxportahr">https://twitter.com/maxportahr</a></td>
<td><a href="https://www.FB.com/maxportal.hr/">https://www.FB.com/maxportal.hr/</a></td>
</tr>
</tbody>
</table>

*Table 28. Monitored social media on Twitter’s and FB’s profiles*

The other criterion used to select the media was the number of publications containing the words “migrant”/”refugee”. The objective was to subject to automated search social media channels of media that are interested in the refugee theme instead of ignoring it. Croatian research team provided the following media outlets for the analysis (Tab.28).

**3.6.2.1. Keywords Analysis 2017**

For the 2017 Croatian Analysis have been extracted 2140 tweets (Table. 29). The emerged atmosphere underlines clear references to the immigration phenomenon. The most common terms of this investigation are bad words, which underline opposition to immigrants (refugees, terrorists, villains and motherfuckers).

\textsuperscript{36} Acronym for “Citizens organize to oversee voting” (Gradani Organizirano Nadgledaju Glasanje).

\textsuperscript{37} http://hnd.hr/najvise-govora-mrznje-na-portalu-dnevno-hr [accessed 1 October 2017].
In addition the political and historical topic emerge of the Country, because the most frequent words present are: Nikola Kalabić (Serbian Chetnik commander during World War Second), Ustaša (Croatian revolutionary organization) and Chetniks (selected keywords for the analysis).

<table>
<thead>
<tr>
<th>KEYWORDS EXTRACT THROUGH TWITTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS</td>
</tr>
<tr>
<td>TERMS</td>
</tr>
</tbody>
</table>

Table 29. Most frequent terms extracted by Keywords Analysis on Twitter

The relation between Nikola Kalabić and Chetniks is very interesting because it is related to an historical period very far from us: the Second World War.

It is important to point out that Croatia was allied with the Kingdom of Yugoslavia and the Chetniks. Indeed, on the 26th of November 1943, Kalabić and the Colonel Simic, General Inspector of Chetnik Troops, concluded a formal collaboration agreement (German: Waffenruhe-Verträge) together with the General der Infanterie (Lieutenant General) Hans Felber, who represented the German Military Commander in Southeast Europe. Other frequent terms are related to communists. Probably, these terms are associated to Nikola Kalabić, since he was anti-communist. In history, the Anti-Communist Volunteer Militia has been composed by local armed auxiliary units of Chetniks and Slovene anti-Partisans in Italian-occupied parts of Yugoslavia. The reference to the historical period of the Second World War and of the commander Kalabić and so to Ustaša, indicates that some topics went back in the current speeches. Past returns. Migrant crisis stirs historical Croatia-Serbia enmity. In 2015, Zagreb is trying to use its border controls as leverage to persuade Belgrade to direct the flow of migrants somewhere other than Croatia - understandable after around 50,000 arrivals in a week.

3.6.2.2. Social media Twitter and FB Analysis 2017

<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>Dnevno.hr 2017</th>
<th>Direktno.hr 2017</th>
<th>Maxportal.hr 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMents</td>
<td>207 TWITTER</td>
<td>28/1,124 FB</td>
<td>346 TWITTER</td>
</tr>
<tr>
<td>TERMS</td>
<td>time, faith, Parliament, message, Plenković, peace, people, election, Croatian</td>
<td>message, shepherd, peace, comment, statement</td>
<td>government, Trump, first, against, President, Plenkovic, million, people, code, Croatian, Croatia, dynamo</td>
</tr>
</tbody>
</table>

**Table 30. Most frequent terms extracted by Twitter and FB Analysis: 2017**

The 2017 Twitter Analysis (Table 30) shows 207 tweets for Dnevno.hr, 346 tweets for Direktno.hr and 355 tweets for Maxportal.hr. The analysis highlight a political factors related to the migratory phenomenon in Europe: The most frequent terms are political terms like peace, election, Parliament, and an important Croatian politician named Plenković. Andrej Plenković who is also a diplomat.
serving as the Prime Minister of Croatia since the 19th of October 2016. Also the Prime Minister Plenkovic has been involved in the immigration question in this last period. During the 2017 conference the Prime Minister has ruled out any fear of deportation of numbers of migrants to Croatia from some other EU countries, following the decision of the European Court of Justice (ECJ) to uphold a European Union rule that requires refugees to apply for asylum in the first EU country they step foot in. Direktno.hr propose the same political framework of Dnevno.hr because mentioned the government, the President and the figure of Plenkovic. Through the MaxPortal.hr the Twitter analysis has extracted information prevalently of political nature. There is the figure of Vesna Pusić (she is a Croatian sociologist and politician who served as a First Deputy Prime Minister and Minister of Foreign and European Affairs in the centre-left Cabinet of Zoran Milanović), of Ivo Josipović (President of Croatia from 2010 to 2015) and of Milan Bandić (Croatian politician that has served his sixth term as a mayor of the Croatian capital, Zagreb). Other terms linked to online hate speech are “Chetniks” and “motherfuckers”. The 2017 FB Analysis reveals 28 posts and 1124 comments for Dnevno.hr, 18 posts and 1810 comments for Direktno.hr and 139 posts and 3990 comments for Maxportal.hr. The content analysis of FB for Dnevno.hr doesn’t show any form of hate speech against immigrants because of the presence of terms like peace and because of the reference to religious or humanitarian topics. Regarding FB analysis, Direktno.hr shows a content analysis much more focused on the social and contextual situation:

- Bad words (violence, criminality, armed forces, fear and a social violent climate)
- Politics (conference, candidate, statements, elections and figures like Prime Minister, President, Krstičević (Damir Krstičević, Croatian Minister of Defence and Deputy) and Nikola Pašić).

This last person was a Serbian and Yugoslav politician and diplomat who was the most important Serbian political figure for almost 40 years and Prime Minister of the Kingdom of Serbia before World War One. Furthermore, it is present Dubrovnik, since it is the destination of an increasing number of refugees.

Finally, Maxportal.hr does not express a clear interest only on political questions. There are also two cities: one is the capital of Croatia, Zagreb and the other one is a city of Serbia, which is Tomislavgrad. Words that do not indicate good topics are also present, such as: gun, deceived, weapon and lied. It is not present a reference to immigrants but there is for sure a climate of tension.
3.6.2.3. Social media Twitter and FB Analysis 2016

Because of the period of political instability, the Croatian Team decided to conduct an analysis in August and September 2016 in order to investigate the causes of the instability. In July 2016, the Croatian President made the decision to call snap elections of members of Croatian Parliament. During the government's training attempt, by the end of 2015, the ruling coalition encountered difficulties in finding a compromise solution. However he failed the attempt to form the parliamentary majority within the legal terms. Therefore, the President officially called for early elections on the 11st of September 2016. According to the GONG report, the communication context was influenced by the starting election campaign on 16th August.

Based on the observed media, it was possible to show the tendency to discriminate immigrants. Over a different period of time, journalists portrayed national minorities, LGBTQ people, refugees and migrants as a danger to society, in more or less explicit terms. The 2016 analysis was conducted from 9th February to 9th April. One of the objectives of the research was to find a possible increase around the event, which took place in 2016, especially on the 9th of March.

The 2016 Twitter Analysis (Table 31) does not reveal political, social and racist aspects of the Croatian society but the above-mentioned political framework is referred to a period of time prior to the monitored period. The burning debates are relative to August and July 2016, while the monitored period goes from February to April 2016.

<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMENTS</td>
<td>1175</td>
<td>2,124/3,5895</td>
<td>725</td>
<td>2,472/4,638</td>
<td>Not available</td>
<td>321/1,224</td>
</tr>
<tr>
<td>TERMS</td>
<td>faith</td>
<td>glory</td>
<td>Chetniks</td>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>message</td>
<td>river</td>
<td>Ustasha</td>
<td>Minister</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>serious</td>
<td>refugees</td>
<td>Veterans</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>migrants</td>
<td>needs</td>
<td>media</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>priests</td>
<td>terrorist</td>
<td>Prime</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>civil</td>
<td>people</td>
<td>Minister</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>attacked</td>
<td>Country</td>
<td>needs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>provokes</td>
<td>Unguided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>voice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>gesture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 31. The Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2016)
The Facebook analysis shows more meaningful outcomes. Dnevno.hr presents 2,124 posts and 35,895 comments, Direktno.hr shows 2,472 posts and 46,387 comments and finally Maxportal.hr indicates 321 and 1,224 comments. Dnevno.hr refers to migrants and civil attacks and thus there could be a connection between the migrants and the attacks in Europe.

The documents containing the keywords of the research on Dnevno.hr are 413. However, the distribution peak is not referred to the selected event occurred on 9th March but results highlighted that this is a period with a growth of illegal entries of migrants into Croatia across the border with Serbia. The article broadcasted in the portal dnevno.hr appears with the title "Police on the hunt for migrants in Slovenian fields and villages" triggering a surge of the hate comments. The peak is on 21st February and on 21st and 22nd March. Other interesting day is on 22nd March for the Brussels bombings (Figure 25).

![Figure 25.Barplot of daily counting of keywords for Dnevno.hr FB Analysis 2016](image)

Direktno.hr is the portal that indicates in a stronger way the perception of immigrants (Figure 26). There are recalls to Chetniks and Ustasha and so to the past history of the Croatia. The analysis individuates also the words refugees, needs and terrorist. It is a social context where there are the refugees like a priority, the negation of the popular needs and a fear of terrorism given by the immigrants’ presence.
In order to test if around the selected event there is a growth of hate speech we can observe the bar graph of the daily counting. The highest peak in the distribution is on 11th February and the second is on 22nd March (Brussels bombings). Therefore, it is not in correspondence of the selected event, but before and after it. Finally, for MaxPortal.hr appear terms like Serbs and conflicts and so not in the strictest sense of the word refereed to immigrants.

![Barplot of daily counting of keywords for Direktno.hr FB Analysis 2016](image)

**Figure 26.** Barplot of daily counting of keywords for Direktno.hr FB Analysis 2016

The bar graph below (Figure 27) for MaxPortal.hr shows a little presence of the selected keywords both for posts and comments. The Croatian results indicate that the social fabric is not so developed in terms of media users, just think to the online penetration equal to 47%, equal to 2 millions of people on 4.22 millions of population. The figures that emerge are referred to a far past and so the social and cultural fabric is linked to this ideological context.

---

*Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931*
3.6.3. Principal hints

The 2017 analysis show words related to an historical period very far from us: the Second World War.

Regarding FB analysis, Direktno.hr and Maxportal.hr show a content analysis much more focused on the social and contextual situation:

- **Bad words** (violence, criminality, armed forces, fear and a social violent climate)
- **Politics** (conference, candidate, statements, elections and figures like Prime Minister, President, Krstičević (Damir Krstičević, Croatian Minister of Defence and Deputy) and Nikola Pašić).
- **Words related to the criminal environment** (gun, deceived, weapon and lied).

The 2016 Twitter Analysis don’t ‘t reveal political, social and racist aspects of the Croatian society.

The burning debates are relative to August and July 2016; while the monitored period goes from February to April 2016 and the event selected was the Prime Minister decision to call snap elections of members of Croatian Parliament.

The attention is focused on the death of Constantinos Mitsotakis and international issues.
3.7. Main results of Greece’s investigation

The scheme used for this paragraph is the following:

1. Overview on national context,
2. Description of keywords choice and on-line press,
3. 2017: results of keywords selection both for Twitter and on-line press,
4. 2016: results Twitter and FB Analysis as agreed with each Partners.
5. Principal hints

3.7.1. National context

The number of active media outlets has been extremely reduced from 2010 to 2017: in 2010 there were 82 national newspapers in Greece among which 22 Sunday newspapers and 607 of local or regional type. Magazines also have a strong market in Greece, with around 174 magazines publishers (Kontochristou and Mentzi in 2010). The financial crisis has negatively affected print media. Newspapers circulation has fallen significantly and several outlets were forced to shut down (Iosifidis and Boucas, Media Policy and Independent Journalism in Greece, 2015). According to Kontochristou and Mentzi, during 2010, 3.7 million Greek households were equipped with television sets. Currently, there are four national state-owned networks, three state-owned national digital television networks, a state-owned satellite broadcast network, and several national private television networks, in addition to approximately 150 local and regional television stations broadcasting across the country. The online news market in Greece is characterized by extreme fragmentation. News websites that regularly engage in conspiracy theories about health and political issues are especially popular. The internet in general, and social media in particular, have become the most popular means of getting the news for Greek people; far more popular than traditional media. Levels of trust toward traditional media are very low while there is higher trust in internet and social media news outlets. Audio visual media content as well as electronic editions of print media and broadcasting on the internet are subjected to state regulations (and self-regulations). There seems to be an adequate legal framework to combat the problem of racist rhetoric and hate speech in regards of both traditional media and online media. As a result, hate speech is widespread in the media and on the internet; it goes largely unchecked and unpunished (ECRI Report, 21). The problem has increased substantially since 2009, in particular in the context of the rise of Golden...
Dawn (ECRI Report, 17). The media have been fuelling anti-immigration sentiments over the years and supported Golden Dawn propaganda indirectly (FIDH/HLHR, 46). Negative stereotypes regarding immigrants and refugees are prevalent in the media (ECRI Report, 21). Media coverage of the migrants’ issue has tended to revolve around a polarization of an “us” versus “them”, in which negative characteristics are ascribed to the image of “the other”. At the same time, the media have systematically represented the immigrant as a “problem” or a “threat”, systematically linking immigrants to anomie, crime, insecurity, morbidity (both literally and metaphorically) (Pantzou). On top of that, the anti-terrorism discourse in the media often targets immigrants and refugees, as well as the Muslim community in general. Thus the media have fuelled hate speech and Islamophobia (ECRI Report, 21). With regard to the Internet in particular, one consequence is that it has turned into a platform where racist and xenophobic messages can be produced and find an audience (Iosifidis and Boucas, 31).

3.7.2. Greece’s Media content analysis

For the automated social media data algorithm, the Greek project team, in collaboration with the advisory board to the project, have decided to focus on the following ten words (Table 32).

<table>
<thead>
<tr>
<th>KEYWORDS</th>
<th>ENGLISH TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Τρομοκράτης</td>
<td>Terrorist</td>
</tr>
<tr>
<td>Μουσουλμάνος</td>
<td>Muslim</td>
</tr>
<tr>
<td>Εγκληματίες</td>
<td>Criminals</td>
</tr>
<tr>
<td>Ισλάμ</td>
<td>Islam – islam-</td>
</tr>
<tr>
<td>Λάθρο-</td>
<td>Smuggle-</td>
</tr>
<tr>
<td>Τζιχαντιστές</td>
<td>Jihadists</td>
</tr>
<tr>
<td>Εισβολή</td>
<td>Invasion</td>
</tr>
<tr>
<td>Βιάζουν</td>
<td>Rape</td>
</tr>
<tr>
<td>Πρόσφυγες</td>
<td>Refugees</td>
</tr>
<tr>
<td>Μετανάστες</td>
<td>Immigrants</td>
</tr>
</tbody>
</table>

Table 32. Table about Greek keywords

The ten keywords were identified after discussing with the Advisory Board to the project. The words “refugees” and “immigrants” were chosen as neutral terms that would lead to the identification of relevant data. The terms “terrorist”, “Muslim”, “Islam”, “Jihadists” were chosen on
the basis of the link that is often made between terrorism and migrants/refugees and Muslims in general (see relevant comment above).

About the choice of media outlets (Table 33 and Table 34), the discussion with the project Advisory Board has been relative to a selection of media outlets with active Twitter and FB accounts. Therefore, for the Greece the choice of social media has fallen on:

**Table 33. Monitored social media on Twitter’s and FB’s profiles**

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>TWITTER PROFILE</th>
<th>FB PROFILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proto Thema</td>
<td><a href="https://twitter.com/prothema?lang=el">https://twitter.com/prothema?lang=el</a></td>
<td><a href="https://www.FB.com/pages/%CE%A0%CF%81%CF%8E%CF%84%CE%BF-%CE%98%CE%AD%CE%BC%CE%B1/906411859380636?ref=ts&amp;fref=ts">https://www.FB.com/pages/%CE%A0%CF%81%CF%8E%CF%84%CE%BF-%CE%98%CE%AD%CE%BC%CE%B1/906411859380636?ref=ts&amp;fref=ts</a></td>
</tr>
</tbody>
</table>

**Table 34. Most frequent terms extracted by Keywords Analysis on Twitter**

<table>
<thead>
<tr>
<th>KEYWORDS EXTRACTED THROUGH TWITTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.OF TWEETS</td>
</tr>
<tr>
<td>TERMS</td>
</tr>
</tbody>
</table>

3.7.2.1. Keywords Analysis 2017

The analysis of the Keywords for the monitored period of time from the 28th of May to the 3rd of June shows 4234 tweets and the most frequent terms are referred to (Table 33):

- Immigration theme (refugees)
- Words related to criminal events or bad people (hatred, prosecutor, dead, terrorist, criminals)
- Religions (Islam)
• Other words like transporters, discharge, transported and children.

Therefore, this time, there seems to be a climate of intolerance: Greeks are aware of the continuous incoming of refugees and of the matured level of hate. It shows, also, a fear context in terms of death, persecution, criminality and terrorism. The condition is that of an invasion from of immigrants/refugees during the last few years.

### 3.7.2.2. Social media Twitter and FB Analysis 2017

<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMENTS</td>
<td>52</td>
<td>30/351</td>
<td>537</td>
<td>8/339</td>
<td>1,345</td>
<td>14/630</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TERMS</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA withdrawal</td>
<td>time here</td>
<td>Greece sketch</td>
<td>solution</td>
<td>Trump</td>
<td>Greece debt</td>
<td></td>
</tr>
<tr>
<td>climate agreement</td>
<td>premiere</td>
<td>Greece long</td>
<td>EU</td>
<td>Macron</td>
<td>euro</td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td>solution</td>
<td>death</td>
<td>debt</td>
<td>Merkel</td>
<td>arrested</td>
<td></td>
</tr>
<tr>
<td>sketch</td>
<td>time</td>
<td>EU</td>
<td>child</td>
<td>survivor</td>
<td>government</td>
<td></td>
</tr>
<tr>
<td>solution</td>
<td>here</td>
<td>Greece</td>
<td>death</td>
<td>Putin</td>
<td>Russia</td>
<td></td>
</tr>
<tr>
<td>debit</td>
<td>Monday</td>
<td>Britain</td>
<td>father</td>
<td>alert</td>
<td>new</td>
<td></td>
</tr>
<tr>
<td>Putin</td>
<td>June</td>
<td>Britain</td>
<td>survivor</td>
<td>dollar</td>
<td>Greek</td>
<td></td>
</tr>
<tr>
<td>company</td>
<td>morning</td>
<td>million</td>
<td>lenders</td>
<td>Immigrants</td>
<td>Merkel</td>
<td></td>
</tr>
<tr>
<td>dirty</td>
<td>look</td>
<td>euro</td>
<td>women</td>
<td>decision</td>
<td>climate</td>
<td></td>
</tr>
<tr>
<td>fast</td>
<td>emission</td>
<td>crisis</td>
<td>solution</td>
<td>Greece</td>
<td>Paris</td>
<td></td>
</tr>
<tr>
<td>company</td>
<td>alloys</td>
<td>death</td>
<td>debt</td>
<td>attack</td>
<td>agreement</td>
<td></td>
</tr>
<tr>
<td>fast</td>
<td>meets</td>
<td>discussion</td>
<td>euro</td>
<td>together</td>
<td>attack</td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>here</td>
<td>USA</td>
<td>Turkey</td>
<td>theme</td>
<td>Manchester</td>
<td></td>
</tr>
</tbody>
</table>

*Table 35. Most frequent terms extracted by Twitter and FB Analysis: 2017*
The 2017 Twitter Analysis (Table 35) has register for ANT1TV 52 tweets, for Kathimerini 537 tweets and for Proto Thema 1,345 tweets. ANT1 TV indicates the themes related to the TV station’s news programs. Through the content analysis it is possible to track the organization of entertainment shows. The topics are related to: time, morning, company, meets and also there are references to the migratory phenomenon. The 2017 analysis show political questions, thanks to words like:

- Countries (USA, Britain, Greece)
- Politics (agreement, Putin, withdrawal)
- Economic situation (Crisis, climate, solution, debt)

Another focus is on the Donald Trump’s intention to pull out of the Paris Agreement.

Finally, the most commonly found words revolve around the death of Constantinos Mitsotakis, a prominent (and for some, controversial) right-wing politician who died on 29th May 2017 (aged 99). The last media is Proto Thema. Also in this case there are words about the death of Constantinos Mitsotakis, as well as Trump and Macron, the Paris agreement on climate change also seems to be a central focus. Other words such as “Merkel”, “Greece”, “euro” and “debt” indicate a focus on the Greek economic crisis. The Manchester attack is also an important topic because a lot of words are linked to this event, already described at the beginning of the chapter.

For the FB analysis it is possible to observe 30 posts and 351 comments for ANT1TV, 8 posts and 339 comments for Kathimerini, and finally 14 posts and 630 comments for Proto Thema. The 2017 FB analysis for ANT1TV show the same topics of the 2017 Twitter Analysis. The attention is focused on the death of Constantinos Mitsotakis and international issues. The Proto Thema FB analysis contains words regarding the death of Constantinos Mitsotakis, Greek debt, current news affairs with a possible emphasis on individual drama and the popular reality show “Survivor”. In addition, it has shown the Manchester bombing (the commonly found word “attack” could be pointing at that; since Manchester is also present). Finally there is something related to immigrants, Merkel, decision, climate, the Paris agreement, important decision and the figure of Angela Merkel. Two events are significant for the examined period from the 28th of May to the 3rd of June 2017. The main one is Manchester attack occurred on the 22nd May 2017. A secondary event is the immigrants’ eviction from the facilities of the former airport in Elliniko, Ahtens, which occurred on 2 June 2017.
3.7.2.3. Social media Twitter and FB Analysis 2016

Again, as in every 2016 analysis shown in this report, the event that the Greek team has selected took place in 2016 and managed to trigger broad discussions on migrants and refugees: that is an event occurred in Oreokastro in the northern Greece on the 16th of October 2016. Oreokastro is one of the Greek places where “Hospitality Centers” for refugees were established amid protests. On October a fatal car accident occurred in which a 76-year old Greek driver hit three refugees with his car. The three people, a mother and her underage son, were killed while the second child was injured. Things got really intense in the “Hospitality Center” for refugees in the area. Angry refugees from the Center attempted to lynch the 76-old driver (Table 36).

Riot police went to the area to calm things down but the refugees have reacted with violence and consequently the police threw stun grenades at them. This event caused heated debates.

The research hypothesis is that online hate speech increases around a relevant event. In this case, the selected event was unexpected (although there was already some tension in the area due to the existence of a “Hospitality Center” for refugees).

<table>
<thead>
<tr>
<th>VARIABLES/YEAR</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. OF TWEETS - N. POSTS/COMMENTS</td>
<td>770</td>
<td>923/5,679</td>
<td>4,914</td>
<td>2,626/19,553</td>
<td>16,537</td>
<td>5,742/71,026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TERMS</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
<th>TWITTER</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>TERMS</td>
<td>Episode look today</td>
<td>episode premiere exclusive company time</td>
<td>Obama USA suffering burning government Clinton chairman Greece</td>
<td>Children Immigrants refugee Church difference Islam East wars jihadists stabilization signed video agreement Syria Russia remains West Clinton public rhetoric</td>
<td>USA Greece photos attack government Obama Turkey empty television Erdogan Clinton shock elections chairman</td>
<td>refugees European Union Greece Country children lynch reasons Islam</td>
</tr>
</tbody>
</table>

Table 36. Most frequent terms extracted by Twitter and FB Analysis: Social media content (year 2016)
The 2016 Twitter Analysis shows: for ANT1TV 770 tweets, for Kathimerini 4914 tweets and for Proto Thema 16537 tweets. ANT1 TV for the content analysis is not representative of political, economics and social themes. The most common words are referred to the shows (for example: episode, look, today). The number of tweets has increased two days after the chosen event in Oreokastro. The third day after the event was also a day of increased Twitter activity. However, during the two-month period there are additional points where an increase in the number of tweets appears, both prior to and after the event. It could be possible to point out the international political situation thanks to the content analysis that lets emerge the Political scenario described by words like Obama, USA, suffering, government and Clinton.

On Twitter there is a great turnout at those days. However, there is a peak on the 8th of November 2016. In addition, the political situation has been mapped through frequent terms such as USA, Obama, Turkey government, Clinton, Erdogan, elections. The number of Tweets related to Proto Thema begins to increase the day after the selected event of Oreokastro. The 2016 FB Analysis shows 923 posts and 5679 comments for ANT1TV, 2626 posts and 19553 comments for Kathimerini and 5742 posts and 71026 comments for Proto Thema. The topics are always linked to the event above and not connected to the immigration scenario. However, the number of Facebook documents containing the selected keywords increased during the days before the event in Oreokastro. The day with the highest number of documents is the 4th of November. This consideration is confirmed by the barplot of daily counting of posts and comments for the FB Analysis 2016 of the ANT1 TV (Figure 28).

![Figure 28.Barplot of daily counting of keywords for ANT1 TV FB Analysis 2016](image-url)
Kathimerini analysis for the monitored period (16 September to 16 November) demonstrates how the most common words are linked to immigration through words such as Jihadists, refugees, immigrants and Islam.

Words that also seem to be related to refugees include “Syria”, stabilization and war. Maybe the words “rhetoric” and “children” are related to the refugee issues. In addition, the political context has been involved and this is proved by the presence of terms like agreement, Syria, Russia and Clinton. However the word rhetoric assumes another meaning if paired with the term public, because it could seem a political question in terms of rhetoric public.

The graphic representation of the FB Analysis for Kathimerini (Figure 29), doesn’t confirm the hypothesis about an increase of online hate speech around the selected event.

In fact, the barplot shows a positive peak of comments and posts on the 19th of September and therefore, before the selected event of Oreokastro. Proto Thema shows as the most frequent words terms like refugees and islam which are directly connected to immigrants.

Furthermore, the word lynch that might be related to the event in Oreokastro since there are reports in which refugees attempted to lynch the driver of the vehicle who killed the two refugees and injured one. Finally, the most frequent terms are: European Union and children; both referred to the immigration topic. In order to test the research hypothesis of the correlation from high level of online hate speech and Oreokastro event it is necessary to observe the bar graph of daily counting of posts and comments that include the keywords for Proto Thema.

![Barplot of daily counting of keywords for Kathimerini FB Analysis 2016](image)
The Barplot of the daily counting of posts and comments for the 2016 FB Analysis (Figure 30) shows:

- A peak of words on the 17th of October, day after the selected event
- A peak of words on the 25th of October

Greece has a penetration equal to 49% and so 5.30 million of person as media users. If it increases the penetration of media users, the results of multimedia content are more indicative. Greece can highlight aspects linked to terrorism and other terms referred to tension climate.

![Figure 30.Barplot of daily counting of keywords for Proto Thema FB Analysis 2016](image)

3.7.3. Principal hints

The 2016 analysis for the monitored period of time from the 28th of May to the 3rd of June shows 4234 tweets and the most frequent terms are referred to:

- Immigration theme (refugees)
- Words related to criminal events or bad people (hatred, prosecutor, dead, terrorist, criminals)
- Religions (Islam)

Other words are transporters, discharge, transported and children, “Merkel”, “Greece”, “euro” and “debt”, indicating a focus on the Greek economic crisis.
Therefore, this time, there seems to be a climate of intolerance: Greeks people are aware of the continuous incoming of refugees and of the matured level of hate. It is also shown a fear context in terms of death, persecution, criminality and terrorism.

The 2016 event selected is the one occurred in Oreokastro in the northern Greece on the 16th of October 2016.

The number of tweets has increased two days after the chosen event in Oreokastro. The third day after the event was also a day of increased Twitter activity.

The number of Tweets related to Proto Thema begins to increase the day after the selected event of Oreokastro and there are words like Jihadists, refugees, immigrants and Islam that indicate the terror climate.

4. Comparison of the countries through the lens of International ranking

4.1 Global competitiveness

We define competitiveness as the set of institutions, policies, and factors that determine the level of productivity of a country. The level of productivity, in turn, sets the level of prosperity that can be reached by an economy.  

The Global competitiveness indexes measure the capacity of countries to provide their citizens opportunities to prosper. This indexes used in business and permits to measure the management of resources, bureaucracy, and regulation institutions of countries.

The Global competitiveness index is calculated by the World Economic Forum and includes the average of different components grouped into twelve pillars that embrace technological and non-technological knowledge. They are Institutions, Infrastructure, Macroeconomic environment, Health and primary education, Higher education and training, Goods market efficiency, Labor market efficiency, Financial market development, Technological readiness, Market size, Business sophistication and Innovation. 

In order to better understand the impact of the above mentioned pillars on the economic systems of the seven countries analysed, and in order to figure out an exhaustive analysis able to show any economic and technological changes, we have chosen five recent years, that are 2012, 2013, 2014, 2015, 2016.

The 12 pillars can be traced back to three different areas:

1. **Basic requirements** (Institutions, Infrastructure, Macroeconomic environment, Health and Primary education);
2. **Efficiency enhancers** (Higher education and training, Goods market efficiency, Labor market efficiency, Financial market development, Technological readiness, Market size);
3. **Innovation and sophisticated factors** (Business sophistication, Innovation).

Looking at Table 37 showing the position of the different countries in the global ranking of competitiveness, it is possible to affirm that the most competitive country is the United Kingdom, which currently occupies the seventh position of the Global ranking.

It is really surprising the increasing competitiveness of Bulgaria from 2012 to 2016, which has risen by 8 positions.

Also Czech Republic has been reliable to develop the three areas such as Basic requirements, Efficiency enhancers, Innovation and sophisticated factors, moving from the forty-sixth position in 2012 to the thirty-first in 2016.

The competitiveness of Croatia and United Kingdom has been remained almost the same during the five years. Greece has increased his competitiveness especially in 2013, jumping from the ninety-first place of the global rank to the eighty-first place in 2013 and 2014, position decrease in years 2015 and 2016. The same reasoning could be made for Romania that has also lost the advantage of competitiveness acquired in 2013, year during that Romania took the fifty-ninth position.

Italy moved from the forty-ninth position in 2012 and 2013 to the forty-third in 2014, and it is still remaining the same.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BULGARIA</td>
<td>57</td>
<td>54</td>
<td>54</td>
<td>50</td>
<td>49</td>
</tr>
<tr>
<td>CROATIA</td>
<td>75</td>
<td>77</td>
<td>77</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>46</td>
<td>37</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>GREECE</td>
<td>91</td>
<td>81</td>
<td>81</td>
<td>86</td>
<td>87</td>
</tr>
<tr>
<td>ITALY</td>
<td>49</td>
<td>49</td>
<td>43</td>
<td>44</td>
<td>43</td>
</tr>
</tbody>
</table>
The following barplot (Figure 31) allows to compare the position of every countries and of course low values are equivalent to higher positions, so the United Kingdom is the most competitiveness country in the follow ranking and it is follow by Czech Republic and Italy. At the bottom of the ranking there are Greece and Croatia.

![Barplot Comparison of Competitiveness](image)

**Figure 31: “Comparison of competitiveness for the seven analyzed countries”**

Source: “Global competitiveness Report”

### 4.2 Press Freedom

In this context of analysis it is really important to look at the Reporters Without Borders. Reporters Without Borders (RWB) is an international non-profit and non-governmental organization with the aim of promoting and defending freedom of press and freedom of information.

RWB works on two spheres of activity: one focused on Internet censorship and the new media, and, on the other hand, the organization provides material, financial and psychological assistance to
journalists. As for the first activity, the RWB organization publishes every year since 2002 the World Press Freedom Index, which is an important advocacy tool based on the principle of emulation between states.

The Index ranks 180 countries according to the level of freedom of press available to journalists, measured in degrees determined by pooling the responses of experts to a questionnaire. The results obtained thanks to the questionnaire is combined with quantitative data on abuses and acts of violence against journalists during the period evaluated and originated from several tallies of abuses and from a network of correspondents in 130 countries.

The table below (Table 38) shows the position associated to each of the seven countries based on their freedom of press calculated from 2012 to 2017 and it is possible to notice that every countries take the same place during time except in the last year, 2017, when Italy overtaken Croatia.

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>BULGARIA</td>
<td>87</td>
<td>87</td>
<td>100</td>
<td>106</td>
<td>113</td>
<td>109</td>
</tr>
<tr>
<td>CROATIA</td>
<td>64</td>
<td>64</td>
<td>65</td>
<td>58</td>
<td>63</td>
<td>74</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>16</td>
<td>16</td>
<td>13</td>
<td>13</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>GREECE</td>
<td>84</td>
<td>84</td>
<td>99</td>
<td>91</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td>ITALY</td>
<td>57</td>
<td>57</td>
<td>49</td>
<td>73</td>
<td>77</td>
<td>52</td>
</tr>
<tr>
<td>ROUMANIA</td>
<td>42</td>
<td>42</td>
<td>45</td>
<td>52</td>
<td>49</td>
<td>46</td>
</tr>
<tr>
<td>UK</td>
<td>29</td>
<td>29</td>
<td>33</td>
<td>34</td>
<td>38</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 38: “Comparison of press freedom for the seven analyzed countries”

Source: “Reporters without borders”

As shown in Figure 32 in the ranking of the Press Freedom, the country with the least possibility of freely expressing itself through traditional media is Bulgaria, placed at the bottom of the ranking, just under Greece. Romania, Italy and Croatia are in the middle of the ranking and at the top there is Czech Republic followed by the United Kingdom which surprisingly is in fortieth place at a global level.

This analysis allow us to affirm that certainly the material recovered through the scraping of the most used social media, such as Twitter and Facebook, surely cannot be generalized because of the restrictions imposed by government policies, policies in continuous development and able to lead the economic and social structure of each country and consequently the manner and the modalities of expression through writing.
4.3 Freedom in the World: political rights and civil liberties global scores

The table 39 shows the scores associated with the seven analyzed countries (Bulgaria, Czech Republic, Croatia, Greece, Italy, Romania and the United Kingdom).

The scores in the table below are the result of the total scores given to different indicators (25) that can be divided into two groups: 10 indicators used to measure political rights and the 15 indicators used for civil liberties.

The scores could range from zero to one hundred: 0 = least free, 100 = most free.

Table 39 shows the aggregate score for every country from 2012 to 2017.

All of the countries analyzed are considered very free from the political and civil point of view because all the scores range from 78 to 97.

All the seven countries are very free taking into consideration political rights, including free and fair elections and, on the other hand, they also have an established fair legal system that ensures the rule of law, allow free economic activity, and tend to strive for equality of opportunity for everyone, including women and minority groups.\(^{41}\)

It is also interesting to note how the freedom score is always the same for each country, that is it has not changed so much during the six years analyzed.

\(^{41}\) https://freedomhouse.org/report/freedom-world-2015/methodology
<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>BULGARIA</td>
<td>81</td>
<td>81</td>
<td>78</td>
<td>79</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>CROATIA</td>
<td>86</td>
<td>86</td>
<td>86</td>
<td>86</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>95</td>
<td>95</td>
<td>94</td>
<td>95</td>
<td>95</td>
<td>94</td>
</tr>
<tr>
<td>ITALY</td>
<td>89</td>
<td>88</td>
<td>90</td>
<td>89</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>GRECIA</td>
<td>85</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>89</td>
</tr>
<tr>
<td>ROMANIA</td>
<td>83</td>
<td>81</td>
<td>84</td>
<td>83</td>
<td>83</td>
<td>84</td>
</tr>
<tr>
<td>UK</td>
<td>96</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>95</td>
<td>95</td>
</tr>
</tbody>
</table>

Table 39: “Civil&political Freedom”
Source: “Freedom House”

The following bar chart (Figure 33) summarized what already said. Clearly the values overlap each other because of their similarity, but it is still possibly to see that the most free countries are United Kingdom, Italy, Croatia and Czech Republic, indeed Bulgaria, Romania and Greece are the least free country among those in our list.

Figure 33: “Civil&political freedom in 2017”
Source: “Freedom House”
4.4 Economic Freedom

The Index of Economic Freedom documents the positive relationship between economic freedom and a variety of positive social and economic goals. The ideals of economic freedom are strongly associated with healthier societies, cleaner environments, greater per capita wealth, human development, democracy, and poverty elimination.

Economic freedom has been measured on the basis of 12 quantitative and qualitative factors, grouped into four broad categories, or pillars, of economic freedom:

- Government Size (government spending, tax burden, fiscal health)
- Rule of Law (property rights, government integrity, judicial effectiveness)
- Open Markets (trade freedom, investment freedom, financial freedom)
- Regulatory Efficiency (business freedom, labour freedom, monetary freedom)42

According to the ranking it is possible to affirm that the United Kingdom is a very free country thanks to the freedom in law, government politics and labor, in fact occupied the fourteenth position in 2014, position that has been changed in 2015 and again in 2017 year in which UK occupied the twelfth position of the economic freedom ranking (Table 40).

Romania and Bulgaria could actually seen as “moderately free”, in particular the first was at the sixty-second position in 2012 and the second was in 61th position in 2012. Both undergo an advancement in the rankings in 2015 (respectively 57th and 55th position) and in 2017 (respectively 39th and 47th positions). Also Italy is now positioned after Romania and Bulgaria, but it is always considered a “moderately free” country. However, it is curious as the change made between 2016 and 2017, jumping from the 86th position to the 79th position.

Croatia and Greece could be considered as “mostly unfree” countries occupying respectively the 95th and the 127th positions. It is possible to point out how Croatia has jumped from the 103rd position of the ranking in 2016 to the 95th position in 2017, showing a better freedom in law, trade and taxes. In the opposite way, could be highlight the negative trend of Greece in 2015 (from 119th of 2015 to 130th place in 2015) and the positive trend in 2017 (from 138th to the 127th).

Czech Republic is still remained one of the most free countries considering the seven countries analyzed in this report and it is now in 28th position.

42 https://www.heritage.org/index/about
Table 40: “Economic Freedom”
Source: “Economic Freedom Index”

<table>
<thead>
<tr>
<th>YEARS</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>BULGARIA</td>
<td>61</td>
<td>60</td>
<td>61</td>
<td>55</td>
<td>60</td>
<td>47</td>
</tr>
<tr>
<td>CROATIA</td>
<td>83</td>
<td>78</td>
<td>87</td>
<td>81</td>
<td>103</td>
<td>95</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>30</td>
<td>29</td>
<td>26</td>
<td>24</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>GREECE</td>
<td>119</td>
<td>117</td>
<td>119</td>
<td>130</td>
<td>138</td>
<td>127</td>
</tr>
<tr>
<td>ITALY</td>
<td>92</td>
<td>83</td>
<td>86</td>
<td>80</td>
<td>86</td>
<td>79</td>
</tr>
<tr>
<td>ROMANIA</td>
<td>62</td>
<td>59</td>
<td>62</td>
<td>57</td>
<td>61</td>
<td>39</td>
</tr>
<tr>
<td>UK</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>12</td>
</tr>
</tbody>
</table>

4.5 Corruption Perception Index

Transparency International refers to Corruption as “grand, petty and political, depending on the amounts of money lost and the sector where it occurs. Grand corruption consists of acts committed at a high level of government that distort policies or the central functioning of the state, enabling leaders to benefit at the expense of the public good. Petty corruption refers to everyday abuse of entrusted power by low- and mid-level public officials in their interactions with ordinary citizens, who often are trying to access basic goods or services in places like hospitals, schools, police departments and other agencies”.43

Table 41 and Figure 34 show that the United Kingdom is the country with the lowest degree of corruption among the seven countries analyzed, in particular the corruption perception has decreased over time from 17th in 2012 to 14th in 2013 and from 14th in 2014 to 11th in 2015 and to 10th in 2016.

On the contrary Greece and Bulgaria are the countries with the higher degree of corruption which for Greece has greatly decreased over time: in fact, Greece has moved from ninety-fourth position in 2012 to fifty-first in 2015 and then goes back to sixty-ninth position in 2016. On the other hand, Bulgaria has occupied more or less a constant position, always around an average equal to the seventy-third position in the ranking (75th in 2016).

43 https://www.transparency.org/what-is-corruption#define
Italy and Romania seemed to be in an equal political situation that leads them to occupy respectively the sixtieth and fifty-seventh position in 2016 and the sixty-second and sixty-sixth position in 2012.

Also, Croatia and Czech Republic seemed to have the same degree of corruption perception which is reflected in the occupation of similar positions in the ranking over the five years analyzed which is equal to 55th in 2016 for Croatia and in 47th for Czech Republic, ranking significantly lower than in 2012 (equal to 62nd and 54th).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BULGARIA</td>
<td>75</td>
<td>77</td>
<td>69</td>
<td>69</td>
<td>75</td>
</tr>
<tr>
<td>CROATIA</td>
<td>62</td>
<td>57</td>
<td>61</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>54</td>
<td>57</td>
<td>53</td>
<td>38</td>
<td>47</td>
</tr>
<tr>
<td>GREECE</td>
<td>94</td>
<td>80</td>
<td>69</td>
<td>58</td>
<td>69</td>
</tr>
<tr>
<td>ITALY</td>
<td>72</td>
<td>69</td>
<td>69</td>
<td>61</td>
<td>60</td>
</tr>
<tr>
<td>ROMANIA</td>
<td>66</td>
<td>69</td>
<td>69</td>
<td>58</td>
<td>57</td>
</tr>
<tr>
<td>UNITED KINGDOM</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 41: “Corruption Perception Index”
Source: “Transparency International”

Figure 34: “Corruption Perception Index”
Source: “Transparency International”
4.6. Global terrorist attacks

The Global Terrorism Database (GTD) is a database comprehending incidents of domestic and international terrorism and the aim is to understand the phenomenon of terrorism. Depending on availability of information, the database records up to 120 separate attributes of each incident, including approximately 75 coded variables that can be used for statistical analysis. These are collected under eight broad categories, as identified in the GTD Codebook, and include, whenever possible:

- incident date
- region
- country
- state/province
- city
- latitude and longitude (beta)
- perpetrator group name
- tactic used in attack
- nature of the target (type and sub-type, up to three targets)
- identity, corporation, and nationality of the target (up to three nationalities)
- type of weapons used (type and sub-type, up to three weapons types)
- whether the incident was considered a success
- if and how a claim(s) of responsibility was made
- amount of damage, and more narrowly, the amount of United States damage
- total number of fatalities (persons, United States nationals, terrorists)
- total number of injured (persons, United States nationals, terrorists)
- indication of whether the attack is international or domestic
- Other variables provide information unique to specific types of cases, including kidnappings, hostage incidents, and hijackings.44

Using “The Global Terrorism Database (GTD)” it could be possible to create a table (Table 42) showing for each of the seven analyzed countries the total attacks from 2012 to 2016 which comprehends several types of attacks that are “bombing/explosion”, “assassination”, “hijacking”, “hostage taking”, “assault”, or “unknown” attack.

44 http://www.start.umd.edu/gtd/using-gtd/
To make comparison possible, every numbers has been returned to base 100, that is to say that to make it possible to compare different measurements, it is necessary to return them to the same base, in this case equal to a total of 100% (Figure 35).

*For Uk we did not consider the attacks in Ireland

---

Table 42: “Terrorist attacks from 2012 to 2016”

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BULGARIA</td>
<td>4,88</td>
<td>3,53</td>
<td>1,61</td>
<td>4,00</td>
<td>1,64</td>
</tr>
<tr>
<td>CROATIA</td>
<td>0,00</td>
<td>2,35</td>
<td>0,00</td>
<td>0,00</td>
<td>0,00</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>0,00</td>
<td>1,18</td>
<td>8,06</td>
<td>8,00</td>
<td>3,28</td>
</tr>
<tr>
<td>GREECE</td>
<td>53,66</td>
<td>62,35</td>
<td>41,94</td>
<td>62,00</td>
<td>50,82</td>
</tr>
<tr>
<td>ITALY</td>
<td>24,39</td>
<td>8,24</td>
<td>11,29</td>
<td>10,00</td>
<td>18,03</td>
</tr>
<tr>
<td>UK</td>
<td>17,07</td>
<td>22,35</td>
<td>37,10</td>
<td>16,00</td>
<td>26,23</td>
</tr>
<tr>
<td>TOT</td>
<td>100,00</td>
<td>100,00</td>
<td>100,00</td>
<td>100,00</td>
<td>100,00</td>
</tr>
</tbody>
</table>

---

*For Uk we did not consider the attacks in Ireland

---

Figure 35: “Terrorist attacks from 2012 to 2016”
On this total, the attacks that hit Bulgaria are a little percentage during years (5% in 2012 and almost 2% in 2016).

The percentage of attacks for Croatia is equal to 0% in every year, counter wise for Chez Republic, it could be possible to see a 0% of attacks in 2012 that increases by 1,18% in 2013 and in 2013 it jumps to 8% then decrease in 2016 when the percentage of attacks is equal to 3%.

Greece and Italy have been attacked much more than the others countries except the United Kingdom.

The percentage of the total attacks for Greece is equal to almost 54% in 2012 then increases in 2013 (62%) and decreases in 2014 when attacks are almost equal to 42%. Another big jump is registered in 2015 when the percentage of terrorists’ attacks is equal to 62/100 decreasing in 2016 (almost 51%).

The total attacks registered for Italy are 24% in 2012. In 2013 there is a decrease of attacks (8%) that increase in 2014 arriving to 18% in 2016.

Finally UK is the second most attacked country in our list of countries, after Greece.

Infact, it has been registered a percentage of terrorist attacks equal to 17% in 2012, 22% in 2013, then there is an increase of terrorist attacks in 2014 (37%) and a decrease in the following year (26%) and another decrease in 2016 (22%).

So taking into consideration the period from 2012 to 2016, numerous attacks hit the United Kingdom during years.

4.7. Comparison between the total population and the immigrants residing in the countries

Comparing population and immigrants residing in the various countries in the years of observation, it is possible to see that the total number of long-term immigrants arriving into the reporting country is not so high considering the total population of each country.

As shown in Table 43 the presence of immigrants is much higher in United Kingdom that in 2015 are 632 thousand, also there are lots of immigrants in Italy and Romania.

Starting from Bulgaria, it is possible to notice that the immigrants arriving into the country are very lower than the total population, as it is for Croatia, country with the lowest presence of immigrants that has been remained the same during years except the jump from 2012 to 2013.
YEARS

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>POP (in Million)</th>
<th>IM</th>
<th>POP (in Million)</th>
<th>IM</th>
<th>POP (in Million)</th>
<th>IM</th>
<th>POP (in Million)</th>
<th>IM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2012</td>
<td></td>
<td>2013</td>
<td></td>
<td>2014</td>
<td></td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>BULGARIA</td>
<td>7.2</td>
<td>14,103</td>
<td>7.2</td>
<td>18,57</td>
<td>7.2</td>
<td>26,615</td>
<td>7.1</td>
<td>25,223</td>
</tr>
<tr>
<td>CROATIA</td>
<td>4.3</td>
<td>8,959</td>
<td>4.2</td>
<td>10,378</td>
<td>4.2</td>
<td>10,638</td>
<td>4.2</td>
<td>11,706</td>
</tr>
<tr>
<td>CZECH REPUBLIC</td>
<td>10.5</td>
<td>34,337</td>
<td>10.5</td>
<td>30,124</td>
<td>10.5</td>
<td>29,897</td>
<td>10.6</td>
<td>29,602</td>
</tr>
<tr>
<td>ITALY</td>
<td>60.9</td>
<td>350,772</td>
<td>60.8</td>
<td>307,454</td>
<td>60.8</td>
<td>277,631</td>
<td>60.7</td>
<td>280,078</td>
</tr>
<tr>
<td>GREECE</td>
<td>10.8</td>
<td>58,200</td>
<td>10.9</td>
<td>57,946</td>
<td>10.9</td>
<td>59,013</td>
<td>10.9</td>
<td>64,446</td>
</tr>
<tr>
<td>ROMANIA</td>
<td>21.4</td>
<td>167,266</td>
<td>20</td>
<td>153,646</td>
<td>19.9</td>
<td>136,035</td>
<td>19.8</td>
<td>132,795</td>
</tr>
<tr>
<td>UK</td>
<td>63.2</td>
<td>498,04</td>
<td>64.6</td>
<td>526,046</td>
<td>65.1</td>
<td>631,991</td>
<td>65.6</td>
<td>631,452</td>
</tr>
</tbody>
</table>

*POP= POPULATION
IM= IMMIGRANTS

Table 43: “Comparison between population and immigrants”
Sources: “Eurostat”, “Global competitiveness Index”, “World Population Data Sheet 2012”

In Czech Republic the number of immigrants decreased from 34.337 in 2012 to 30.124 in 2013, then to around 30,000 in 2014 and 2015.

Also in Italy the number of immigrants decreased from 350,772 in 2012 to 280,078 in 2015, on the contrary to what it is said by newspapers and television.

In Greece, on the contrary, the number of immigrants increases from 58,200 in 2012 to 64,446 in 2015.

In Romania there is again a decrease of immigrants, decreased by almost 35,000 people from 2012 to 2015.

Contrariwise The United Kingdom hosted an increasing number of immigrants that reach 631,452 in 2015.
5. Conclusive Analysis

The project has been developed as an analysis based on Big data use. This approach represents a strength point: a multi-languages comparison, localized in a European belt of countries traditionally considered as “emigrations lands” (except for the UK), on the hate speech theme. The following table (Table 44) provides a comparison of the keywords on which was based the analysis.

<table>
<thead>
<tr>
<th>UK</th>
<th>BULGARIA</th>
<th>ITALIA</th>
<th>ROMANIA</th>
<th>REP. CECA</th>
<th>CROATIA</th>
<th>GRECIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invaders</td>
<td>мигрант, мигранти</td>
<td>Pocket money</td>
<td>İslamişti</td>
<td>Uprihilci</td>
<td>izbjeglička kriza</td>
<td>Тромокрάτης</td>
</tr>
<tr>
<td>Fake fugees</td>
<td>бежанец, бежанци</td>
<td>Prima gli italiani</td>
<td>Musulmani</td>
<td>Migranti</td>
<td>teroristi</td>
<td>Μουσουλμάνος</td>
</tr>
<tr>
<td>Gimmigrants</td>
<td>заплаха</td>
<td>”Non” ci rubano il lavoro</td>
<td>Migrant/Migranti</td>
<td>İslám</td>
<td>kozojebi</td>
<td>Εγκληματίες</td>
</tr>
<tr>
<td>#/NoRefugees</td>
<td>терорист. терористи</td>
<td>Magrebini</td>
<td>İslamisté</td>
<td>ustaše</td>
<td>Ислам</td>
<td></td>
</tr>
<tr>
<td>Rapefugees</td>
<td>伊斯兰, islamist</td>
<td>Fini profughi</td>
<td>Teroristi</td>
<td>Terorismus</td>
<td>pederi</td>
<td>Λάθρο-</td>
</tr>
<tr>
<td>Vermin</td>
<td>нелегален, нелегални</td>
<td>Immigrati</td>
<td>Pražská kavárna</td>
<td>balije</td>
<td>Τζιχαντιστές</td>
<td></td>
</tr>
<tr>
<td>Foreigners</td>
<td>талибани</td>
<td>Clandestini</td>
<td>Četnici</td>
<td>Etobolh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romanians</td>
<td>на сапун</td>
<td>Islamici</td>
<td>Poponar</td>
<td>antife</td>
<td>Війсівн</td>
<td></td>
</tr>
<tr>
<td>Deport</td>
<td>чернилка</td>
<td>Barconi</td>
<td>Jidan</td>
<td>Pravdolaskaři</td>
<td>bodljikava žica</td>
<td>Πρόσφυγες</td>
</tr>
<tr>
<td>Migrants</td>
<td>вън</td>
<td>Bozgor</td>
<td>Multikulturalismus</td>
<td>izbjeglice</td>
<td>Метаанάστες</td>
<td></td>
</tr>
<tr>
<td>Immigrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 44: “Comparison of keywords”

Usually, the common imaginary places the founding countries of the European community as host countries, and hate speech, with reference to illegal migrants, find in these territories niches of supporters. In fact, the results show that the hate speech phenomenon does not develop only in the host territories but also in those traditionally appointed to the departure or emigration. Obviously, the so-called emerging countries channel their hate speech towards more political issues of both international and domestic relevance. These reflections are the result of what emerges from the study conducted, taking into account that the sampling observation was limited to only two weeks in 2017, so the results cannot be generalized to the whole year and especially to the entire population, but only to that part of the population using social media.
With reference to this aspect, it has to be considered that the countries involved have different rates of digitization, thus the difference in the number of tweets, posts and comments among the different countries must be evaluated with weighting. A note of reflection, however, should be placed on the keywords chosen, given that for some countries has not been reflected in the observation of the two weeks.

It would be desirable to analyse data of longer observation times, and a wider range of keywords to further evaluate these first results obtained from an exploratory analysis. Despite the limited outlook offered by the study, a total of 5 million (5,535,237) messages were analysed in 2016 and 2017, in the form of tweets, posts and comments, as shown in the tables below (Tables 45 and 46).

<table>
<thead>
<tr>
<th>TWITTER</th>
<th>BULGARIA</th>
<th>CROTIA</th>
<th>GRECIA</th>
<th>ITALY</th>
<th>REP. CECA</th>
<th>ROMANIA</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>182</td>
<td>2,140</td>
<td>4,234</td>
<td>2,149</td>
<td>268</td>
<td>89</td>
<td>48,170</td>
</tr>
<tr>
<td>Newspaper</td>
<td>1,076</td>
<td>908</td>
<td>1,934</td>
<td>1,961</td>
<td>327</td>
<td>1,366</td>
<td>2,263</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper posts</td>
</tr>
<tr>
<td>Comments</td>
</tr>
</tbody>
</table>

Table 45: “Size of tweets, posts and comments in 2017”

<table>
<thead>
<tr>
<th>2016</th>
<th>BULGARIA</th>
<th>CROTIA</th>
<th>GRECIA</th>
<th>ITALY</th>
<th>REP. CECA</th>
<th>ROMANIA</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper posts</td>
<td>6,932</td>
<td>4,917</td>
<td>9,291</td>
<td>9,579</td>
<td>3,005</td>
<td>14,810</td>
<td>7,335</td>
</tr>
<tr>
<td>Comments</td>
<td>192,081</td>
<td>83,506</td>
<td>96,258</td>
<td>1,344,698</td>
<td>65,150</td>
<td>331,906</td>
<td>3,026,863</td>
</tr>
<tr>
<td>Tweets</td>
<td>5,857</td>
<td>1,900</td>
<td>22,221</td>
<td>21,924</td>
<td>1,691</td>
<td>6081</td>
<td>22,749</td>
</tr>
</tbody>
</table>

Table 46: “Size of tweets, posts and comments in 2016”
With this Big data perspective two classes of suggestions have been obtained. The first one has to do with technical aspects, whilst the second one with the cultural and sociological aspects. The technical details, on the other hand, can be distinguished in information retrieval limits and in methodological aspects. The problems encountered on a web scraping level have shown a rising, albeit important, question about the regulation of the use of online information. At the time of data collection, the legislation on web scraping was rather nuanced, while it was well defined with regard to providers. This aspect is certainly to be monitored, because the usability of the information changes the analytical profile. Closely related to the issue of intellectual property of data that passes over the platforms, there is the topic of the level of availability of the data: it is not always clear the methods of sampling of tweets, given that the maximum quantity for each query is equal to 3,200 tweets. This leads to the limits indicated for the generalization of the results, even if in the explorative perspective that has been pursued, the amount of data analysed can be considered adequate.

A further aspect is related to the complexity of languages, having chosen lesser-diffused languages and above all deriving from Cyrillic, whose comparison required greater attention, and in some cases the development of special codes.

The sociological aspects have instead allowed to point out that the phenomenon of HIT speech exists in all emerging countries and, above all, that it is possible to affirm that the hate speech can be modelled on two different levels: the first referred to what happens in general in the world. In fact, in all analyses we can find references to foreign events and the use of inappropriate terms for the events commented. The second level, on the other hand, is oriented towards internal dynamics, and therefore it is observed that in the emerging countries the historical background is still very vivid, and therefore the expressions are strongly conditioned to the past history. Specifically, with reference to the first level, certain events are attacking the individual safety, inducing fear, danger perception lead to uncontrolled reactions. These trigger events have the potential to emphasize and reinforce the online hate speech phenomenon. Through the study of the media content has been possible to note how some events related to terrorist attacks (or similar acts) have been mentioned in each national analysis. In particular, in the group of the most frequent words were presented Manchester attack or the Brussels bombing that have influenced the opinion against refugees and migrants. However, also political and social events can be the driving force of the racist phenomenon. It is the case of UK investigation that has seen, among other events with similar effects, the EU referendum as the trigger event for the increase of OHS. Finally, there are
international events that represent the sounding board for some expressions against the minorities (e.g. the above-mentioned Brussels or Manchester attacks).

At this point it is possible to carry out a comparative analysis among all seven Countries. The cross-analysis leads to affirm that we have some common topics of the online hate speech in the European countries involved in the project. The following table presents the summary of this analysis. In the rows there are the main common words and in the columns there are the seven Countries. In the intersection between a row and a column, if the cell is coloured, it means that the term is present in the analysis of that Country.

Table 47. “Summary of the transnational Analysis”

<table>
<thead>
<tr>
<th>FREQUENT TERMS</th>
<th>UK</th>
<th>ITALY</th>
<th>BULGARIA</th>
<th>ROMANIA</th>
<th>CZECH REP.</th>
<th>CROATIA</th>
<th>GREECE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTACK/S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOMBING/S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISLAM/ISLAMIST-S/ISLAMIC/ISLAMISM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TERRORISM/TERRORIST-S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SECURITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REFUGEES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMMIGRANTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIGRANTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILLEGAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUSLIM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DANGER/DANGEROUS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MANCHESTER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRUSSELS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGREEMENTS/S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRUMP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MERKEL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUTIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Looking at the table above (Table 47) it is possible to affirm that the issue of the attacks is present on each Country, except for Italy. This word is not one of most frequent terms in the Italian investigation, but the issue is present in UK, Bulgaria, Romania, Czech Republic, Croatia and
Greece. Terrorism is a common thread for all the analysis. The topic emerges in every context and also in every European Country. Islam is an important issue for each country, but not for Croatia. The security is a problem detected only from Bulgaria, but the refugees theme has the same importance of the terrorism, because it appears in each country. The Brussels bombing is a relevant issue for UK, Bulgaria and Romania. For the same countries the Manchester attack is also important. In this last case, to the other countries, it is included Greece in terms of relevance. In addition to that, the ISIS topic is a frequent concept for UK, Italy, Romania and Czech Republic.

Finally, another common thread in the comparative analysis is an international political figure, Donald Trump. This political personality was very intrusive in the social media, also for his engagement to obstacle refugees. Since the headlines of several newspapers indicated: "Mexican Border Wall to Be Built and Plans to Block Syrian Refugees" and also, "Donald Trump's policies, not Mexicans, are the real immigration problem in America" and so on, he could have contributed to widespread the attention and hate speech on the immigration theme.

Finally, with the use of different kinds of indicators that represent the social, economic and political profile of each Country and the level of immigration, it was possible to further evaluate the results of the social media Facebook and Twitter in 2016 and 2017.

The countries are not too competitive in training, market size and in technological devices except for the United Kingdom and that could explain the difference based on the number of tweets, posts and comments between UK and the rest of the Countries: large quantities for the first one and low quantities for the latter ones.

Regarding the Press Freedom, the freest country seemed to be Czech Republic and the worst Bulgaria, but besides this evidence, there are not collected terms linked to this dimension. So, the collected terms do not demonstrate a racist climate in Bulgaria, while online hate speech is certainly high.

Taking into consideration the Global Competitiveness ranking (Tab. 37), that is constructed using indicators that can be grouped into three areas, that is basic requirements, efficiency enhancers and innovation and sophisticated factors, the United Kingdom is the most competitive country of the seven analysed and it is followed by Czech Republic and Bulgaria.

From the Economic point of view (Tab. 40), in 2017 the highest country in the Economic Freedom ranking is even the United Kingdom (12th place) followed by Czech Republic (28th place),


Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
Romania (39th place) and Bulgaria (47th place). The less free countries are Italy (79th place) and Greece (127th place).

All of the seven countries are very free from the political and civil point of view because all the scores range from 78 to 97 (The scores range from zero to one hundred: 0 = least free, 100 = most free). That means they are free taking into consideration political rights, including free and fair elections and, on the other hand, they also have an established fair legal system that ensures the rule of law, allow free economic activity, and tend to strive for equality of opportunity for everyone, including women and minority groups (Tab. 39).

Again, considering the Corruption Perception Index, when corruption consists of acts committed to a high level of government that distort policies or the central functioning of the state, allowing leaders to benefit at the expense of the public good, UK is the less corrupted country (in 2016 is in 10th place) while the countries with more corruption are Greece (in 2016 is in 69th place) and Bulgaria (75th place in 2016).

Furthermore, there are two curious aspects:

- The online hate speech and also what people think of the migratory phenomenon have certainly fed the way the newspapers describe the events and news, and consequently the word-of-mouth that derives from it. In fact, although the media talk about a constant and big migratory flows, looking at the Table 43 it is possible to see how the legal immigrants who then settle in the countries are really few, if compared to the total population. The legal aspect becomes a lens through analysed the OHS.

- On the other hand, Greece and UK seemed to be the countries most affected by terrorism, which is a phenomenon felt very deeply. This “felt” has been confirmed for UK, both by the peak of words registered after the terrorist attack in France on the 14th of July 2016, and by the attention that has been expressed for the religion (Muslims, Islamists).

On the other hand analysing the Greek scenario, the results demonstrate how strong is the complaint against criminal acts. In fact, there are words like hatred, prosecutor, dead, terrorist, criminals and words like refugees and also a great attention to Islam. Therefore a climate of intolerance emerges: Greek people are aware of the continuous incoming of refugees and the matured level of hate. It is also shown a fear context in terms of death, persecution, criminality and terrorism.
References

(Web Sites)

- Avvenire on Facebook: https://www.FB.com/avvenire.it/
- Avvenire on Twitter: https://twitter.com/avvenire_nei
- BLITZ.BG on Facebook: FB.com/blitz.bg
- BLITZ.BG on Twitter: twitter.com/blitz.bg
- Digital economy and society statistics - households and individuals Source processing on We are social and Hootsuite 2016 data. Available at:
- Donald Trump's policies, not Mexicans, are the real immigration problem in America. Available at: http://www.nydailynews.com/news/politics/trump-policies-real-immigration-problem-u-s-article-1.2971123
- Source processing on We are social and Hootsuite 2016 data. Available at: ec.europa.eu/eurostat/statisticsexplained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals
- European Journalism Centre. Available at: http://ejc.net/
- Global Terrorism Database. Available at: http://www.start.umd.edu/gtd/using-gtd/
- Il Giornale on Facebook: https://www.FB.com/ilGiornale/
- Il Giornale on Twitter: https://twitter.com/ilgiornale
- Install Watson/PythonAPI. Available at: https://github.com/CognitiveBuilder/HelloCognitiveWorld/blob/master/prerequisites/3-watson_api.md
- Internet access and use statistics - households and individuals. Available at: http://ec.europa.eu/eurostat/statisticsexplained/index.php/Internet_access_and_use_statistics_-_households_and_individuals
- La Repubblica on Facebook: https://www.FB.com/Repubblica/
- La Repubblica on Twitter: https://twitter.com/repubblica
- Market share held by mobile operating systems in the United Kingdom (UK) from December 2011 to November 2017. Available at: https://www.statista.com/statistics/262179/market-share-held-by-mobile-operating-systems-in-the-united-kingdom/
• Najviše govora mržnje na portalu Dnevno.hr. Available at: http://hnd.hr/najvise-govora-mrznje-na-portalu-dnevno-hr
• Nova on Facebook: FB.com/Nova.bg/
• Nova on Twitter: twitter.com/NoviniteNaNova
• Novini on Facebook: FB.com/Novini.bg
• Novini on Twitter: twitter.com/Novini.bg

• Number of Twitter users in the United Kingdom (UK) from 2012 to 2018 (in million users). Available at: https://www.statista.com/statistics/271350/twitter-users-in-the-united-kingdom-UK/
• Reporters Without Borders. Available at: https://rsf.org/en

(Reports, articles and books)

• Alexa/SimilarWeb. License. CC-BY-NC
• Centre for the Analysis of Social Media, Demos, Islamophobia on Twitter: March to July 2016 Carl Miller, Josh Smith, Jack Dale. Available at: https://www.demos.co.uk/wp-content/uploads/2016/08/Islamophobia-on-Twitter-March-to-July-2016-.pdf
• Freedom of the Press 2016 – Bulgaria. Available at: http://www.refworld.org/docid/582ac6dbf.html
• Joao Augusto Sobreiro Sigora, Corina Lovison Nassif Avellar, Cristal Augustus Carneiro Ribeiro- “Individual empowerment in the international system: Towards development, through freedom”, 2012, Brasilia. Available at: https://books.google.it/Individual empowerment in the international system: Towards development, through freedom.
• Legal framework, societal responses and good practices to counter online hate speech against migrants and refugees. Available at: https://positivemessengers.net/images/library/pdfs/comparative_report-1.1.4.09.pdf
• Mapping Out the National Context of Online Hate Speech in Greece. Available at: https://www.positivemessengers.net/images/library/pdfs/OHS-report_Greece-fn-eng__form.pdf
• Mapping Out the National Context of Online Hate Speech in United Kingdom. Available at: http://www.adl.org/Internet/Binder_final.pdf
• Media Content Analysis on Online Hate Speech. National Report- Bulgaria. Available at:
• Media Content Analysis on Online Hate Speech. National Report- Italy. Available at: https://positivemessengers.net/images/library/pdfs/MCA1.2_IT_eng.pdf
• Media Content Analysis on Online Hate Speech. National Report- Romania. Available at: https://positivemessengers.net/images/library/pdfs/MCA-1.2-RO-eng.pdf
• Media Content Analysis on Online Hate Speech. National Report- Romania. Available at: https://positivemessengers.net/images/library/pdfs/MCA_1.2_CzR-eng.pdf
• Package ‘rtwitter’. Available at: https://cran.r-project.org/web/packages/rtweet/rtweet.pdf
• Package ‘twitteR’. Available at: https://cran.r-project.org/web/packages/twitteR/twitteR.pdf
• Responding to Cyberhate, Toolkit for Action (ADL). Available at: http://dc.adl.org/resources-for-responding-to-hate-in-the-community/
• World Economic Forum. Available at: http://reports.weforum.org/
SECTION 2 - Neural Networks analysis of the on line Hate speech in 2016

Authors:
Massimo Buscema*
Giulia Massini*

*Semeion Research Center of Sciences of Communication.
Via Sersale 117, Rome, Italy, 00128
Introduction

Preliminary definition, Data and Methods

This section of the report is the result of the study of the hate speech on line phenomenon in the 7 partner countries, over a period of two months in 2016, chosen by each project partner. The choice of the time span is related to a particular event occurred during the year.

The purpose of the analysis is to determine whether and how a certain event is linked to the spread of hate speech online in each country, and also in terms of comparison between the analyzed countries. The data collected refers to the social media Facebook and Twitter used by three national media selected by a scientific committee, formed in each country by the project partners.

In each media the posts and tweets were collected in the chosen period, mapped with 10 words, which express the hate speech. There are some discrepancies in the data, due to the lower number of words found from the various media, compared to those chosen, or the choice of a greater number of words, as in the case of the UK partner.

The methodology of the data collection is explained in detail in the Section one of this report. Even if was choices a limited number of words and social media for the creation of a data base, the size of the data set can be considered big wide to reach the results of the analyzes, which offer useful food for thought.

Methods

The methodology of analysis is based on the use of a battery of innovative artificial neural networks tools for the analysis of the data in each country (part 1 of the present section), and a comparative analysis between the various countries (part 2 of the present section).

The systems used for the analysis are the following:

- Auto Contractive Maps (AutoCM) is an Artificial Neural Networks (ANN) designed to highlight the similarities of the verbal behaviour shown by each country;
Maximally Regular Graph (MRG) and Graph In Out ((G IN-OUT) are complex algorithms, able to depict the fundamental relationships among the data by mean a weighted graph of similarity.

Dynamic Networks Block (DNB) is a complex evolutionary algorithm able to detect the temporal cause-effect relations, between the words used by each country;

Adaptive Fusion System (from: Theory of Impossible Worlds (TIW)) to merge the databases of different countries into a single structure of behaviour.

AutoCM neural network identifies the value of existing relationships between variables or records, and the MRG graph displays only the fundamental relationships, the most relevant that each variable (or record) maintains at least with another. For this reason, the MRG graph is a weighted graph, since the branches, in addition to connecting 2 nodes (variables), display the value.

The Maximally Regular Graph (MRG) is an undirected weighted graph, which shows the fundamental relationships between the variables (in this case, the words selected by each country), or the records (in this case, the days of the study period), through the elaborations that the AutoCM neural network operates on the HateSpeech dataset.

Of particular interest, in the analysis of an MRG graph, are the indices that characterize each network graph (Fig. 1):

- **the clique**: group of nodes (greater than 2) that are highly correlated with each other;
- **the leaves (leaf or external vertex)**: terminal nodes placed in the periphery of the graph (outlier);
- **Betweenness nodes**: nodes that are obliged passage for the communication of other nodes that are not directly connected to each other.
- **Hub nodes**: nodes that act as mediators, in relationships with different nodes, even if they are not necessarily more relevant than the others.

In other words, the MRG represents the framework of the "bone and nervous system" of the analysed dataset.
The DNB system is an Evolutionary Algorithm that attempts to reconstruct the predominant cause-effect relationships among variables using their Probability Density Function distribution. An oriented weighted graph shows the cause-effect relationships identified by the DNB system. Such a graph can be spread, that is, may be formed from multiple subgraphs not connected to each other.

Adaptive Fusion System (from: Theory of Impossible Worlds (TIW)) is a new hybrid adaptive system that, by using self-associated neural networks, is able to project different datasets by type of variables and records on the same spatial domain (Hidden Unit of networks). The result of this fusion can be processed through an AutoCM network and represented by different graphs such as MRG or G-IN-OUT.

The Graph (G-IN-OUT) is an algorithm suitable for filtering a square matrix of weights, with the main diagonal null, generated by any other algorithm.

The G IN-OUT algorithm is the matrix of weights assigned in the form of weighted graph with specific characteristics:

- The graph obtained is an oriented graph when the original matrix is not symmetric \((v_i, v_k) \neq (v_k, v_i)\);
The resulting graph could be composed of sub-graphs not necessarily connected to each other (scattered graph, forest).

For an in-depth knowledge of the tools of analysis used and their references, see the Section 3 at the end of this paper.

0.3 Definition

To proceed with the presentation of the results it is necessary to introduce two concepts that will be used during the presentation and interpretation of the results.

- **Ideologhema**, an important association between very connected and very frequent concepts.
- **Causal Chain**, a linked series of significant cause - effect relationships that when is reciprocal, it is said that is reciprocal or bidirectional.

1. Data analysis of each country

In this part of the report we present the results of the data analysis of each country, divided into four components:

1. Analysis of the frequency and distribution of the words in the chosen period: it is a graphical representation of the frequency of the words that appear in the media in each country, during the period covered by the study;
2. Analysis of the similarity of behaviour of the words through the graph MRG, produced with the processing of the data with AutoCM;
3. Analysis of the graph that represents the cause-effect relationships of the use of words, obtained through the data processing system with DNB;
4. With the MRG graph, obtained with the AutoCM system, analysis of the similarity of the verbal behaviour between the observed days and the day chosen, based on an event that was considered potentially significant in terms of hate speech.
1.1. Bulgaria

1.1.1. Analysis of the frequency

With the analysis of the frequency (Tab.1.1) and the distribution of the words during the days observed (Figure 1.1), it can be easily identified three scenarios that come out from the routine:

- The period that begin on 12-04-2016 and ends on 14-04-2016 (peak B). In this period, which begin from the day chosen, the word migrant is activated with emphasis and then the words refugee and illegal in succession.

- A short period after the day chosen, that begin on 19.04.2016 (peak C) in which, again, it’s highlight the word migrant. This time in connection with the words threat and out.

- A period before the chosen day that starts on 23-04-2016 (peak A), in which the word terrorist has a peak of frequency, followed by the frequency of the words refugees and migrants.

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>refugee</td>
<td>3193</td>
</tr>
<tr>
<td>terrorist</td>
<td>3117</td>
</tr>
<tr>
<td>migrant</td>
<td>2887</td>
</tr>
<tr>
<td>out</td>
<td>1285</td>
</tr>
<tr>
<td>illegal</td>
<td>581</td>
</tr>
<tr>
<td>threat</td>
<td>451</td>
</tr>
<tr>
<td>Islamist</td>
<td>410</td>
</tr>
<tr>
<td>turn_soap</td>
<td>136</td>
</tr>
<tr>
<td>Taliban</td>
<td>61</td>
</tr>
<tr>
<td>nigger</td>
<td>9</td>
</tr>
</tbody>
</table>

*Table 1.1 Table of frequencies*
1.1.2. Similarities in behaviour

The basic concepts that occur most frequently are four: refugee, terrorist, migrant and out. Each of them with a different relevance that emerges in the analysis of the MRG graph of similarities (fig.1.2).
The word migrant is at the centre of the graph MRG, the one from which branch out various possibilities. It therefore seems to be the most neutral word that lends itself to be associated with other concepts that provide the content.

- **migrant** comes strictly (0.73) connected to the word **Taliban** (infrequent), which is connected with the very frequent concept out;
- **migrant** is closely connected (0.69) also to the word **refugee**, which in turn is divided into terrorist, Islamist and nigger;
- **migrant**, finally, is strongly linked even if separately, to the words threat (0.74), turn_soap (0.64) and illegal (0.63).
- The most relevant ideologhema is identifies in the set of three words: Migrantrefugee-terrorist.

### 1.1.3. Cause-Effect Relationships

Analyzing the table 1.2 and the graph (fig.1.3) that connects the words in terms of cause and effect, we find a very close association (probability = 0.86) between migrant and illegal, that reinforce each other, with a distance of 1 day and 5 days: when a word is used, the effect is to use also the other, and vice versa.

A case-effect feedback loop is also identified: if treat then out (probability = 0.86), if out then terrorist (probability = 0.53), if terrorist then refugee (probability = 0.69), if refugee then nigger (probability = 0.86), if nigger then Taliban (probability = 0.83), if Taliban then treat (probability = 0.47). It can be said that the chain ends with **islamist**.

<table>
<thead>
<tr>
<th>cause</th>
<th>effect</th>
<th>reactivity</th>
<th>strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>BU_migrant</td>
<td>(\rightarrow) BU_illegal</td>
<td>1</td>
<td>0.705499</td>
</tr>
<tr>
<td>BU_refugee</td>
<td>(\rightarrow) BU_nigger</td>
<td>4</td>
<td>0.464974</td>
</tr>
<tr>
<td>BU_threat</td>
<td>(\rightarrow) BU_out</td>
<td>1</td>
<td>0.857095</td>
</tr>
<tr>
<td>BU_terrorist</td>
<td>(\rightarrow) BU_refugee</td>
<td>2</td>
<td>0.687364</td>
</tr>
<tr>
<td>BU_illegal</td>
<td>(\rightarrow) BU_migrant</td>
<td>5</td>
<td>0.776638</td>
</tr>
<tr>
<td>BU_Taliban</td>
<td>(\rightarrow) BU_threat</td>
<td>6</td>
<td>0.467631</td>
</tr>
<tr>
<td></td>
<td>(\rightarrow) BU_tur_soap</td>
<td>6</td>
<td>0.322665</td>
</tr>
</tbody>
</table>
Table 1.2. Table obtained from the processing of the DNB system in which there are shown every word in the role of cause and/or in the role of effect. The "reactivity" column shows the number of days after which the effect word is used and in the "strength" column the probability of this happening.

<table>
<thead>
<tr>
<th>Word1</th>
<th>Word2</th>
<th>Reactivity</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>BU_tur_oap</td>
<td>BU_Islamist</td>
<td>5</td>
<td>0.529491</td>
</tr>
<tr>
<td>BU_nigger</td>
<td>BU_Taliban</td>
<td>4</td>
<td>0.826869</td>
</tr>
<tr>
<td>BU_out</td>
<td>BU_terrorist</td>
<td>1</td>
<td>0.532424</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>3.5</td>
<td>0.617065</td>
</tr>
</tbody>
</table>

Fig. 1.3. Graph obtained with the elaboration of the DNB system that highlights the causal chain of the use of words.

1.1.4. Relations between the analyzed days (13.03.2016 - 11.05.2016)

With regard to the dynamics of the relationships of observed days, from the analysis of the graph MRG (fig.1.4), it can be noted that there are periods of maximum uniformity of behaviour, represented by days with a dense network of interconnections (click1 and click2). As can be seen in the graph, the focus day 11/04/2016 (red circle) and the following day 12/04/2016 (blue circle) are branches of the graph and therefore to be considered atypical days (which comes out with the routing expressed by the other days). However, the day 12-04-2016 is associated with the other subsequent days of 13 and 14-04-2016, as shown in the frequency graph (peak B). It means that the logic of the change of use of the words has had 2 days of effect. In the MRG graph the days referred to both the peak A and the peak B of the graph of frequency, always constitute the branches of the graph. These are therefore atypical situations.
Fig. 1.4. Graph MRG obtained from the processing of AutoCM system that shows the verbal behaviour of similarity connections between the days analyzed. The red circle shows the focused day while the blue circle highlights the following day to day focused. The dashed circles show the clicks, days of uniformity of behaviour.

1.2. Czech Republic

1.2.1. Analysis of the frequency

With the analysis of the frequency (Tab. 2.1) and the distribution of the words during the days observed (Figure 2.1), it can be easily identify isolated frequency peaks that concern one word at a time:

- The word with the highest peaks is **migrant**.
- **Refugee** has only one peak even if it is the most relevant,
- Finally, there are contained frequencies of peaks for the word **Islam**.

The day focus 21_08_2016 and the following day 22_08_2016 do not fall within a period of particular turbulence.
1.2.2. Similarities in behaviour

In the MRG graph (fig.2.2) the word Islam is at the centre and it is connected, albeit weakly, with all the words, mostly with the words: Islamists, terrorism and slunickari.

The latter, Islamists, terrorism and slunickari, form a triad ideologhema with strong connections.

The words migrants and refuges, although very present in terms of frequency, do not create a system with other words and are rather isolated. The word Prague-cafe is totally isolated.
Fig. 2.2. Graph MRG obtained with the elaboration of the AutoCM system, that shows the connections of similarity between the concept-words used. The size of the circles refers to the frequency of words. The values refer to the strength of similarity between words. In red the word (or words) in the centre of the graph.

1.2.3. Cause-Effect Relationships

Analyzing the graph (fig. 2.3) that connects the words in terms of cause and effect, we find a very close association (probability = 0.92), between *refuges* and *Islam*, that reinforce each other in 3 days: "refuges then Islam" and at the same time "Islam then refuges". *Terrorism*, on the other hand, is isolated.

We find a chain of strong causal relationship between the words: *slunickari* (probability = 0.97) => *Islamists* (probability = 1.0) => *migrants*.

<table>
<thead>
<tr>
<th>cause</th>
<th>effect</th>
<th>reactivity</th>
<th>strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH_refuges</td>
<td>CH_Islam</td>
<td>3</td>
<td>0.905741</td>
</tr>
<tr>
<td>CH_Islam</td>
<td>CH_refuges</td>
<td>3</td>
<td>0.919729</td>
</tr>
<tr>
<td>CH_Islamists</td>
<td>CH_migrants</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2.2. Table obtained from the processing of the DNB system in which there are shown every word in the role of cause and / or in the role of effect. The "reactivity" column shows the number of days after which the effect word is used and in the "strength" column the probability of this happening.

<table>
<thead>
<tr>
<th>Word 1</th>
<th>Word 2</th>
<th>Reactivity</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH_terrorism</td>
<td>CH_terrorism</td>
<td>1</td>
<td>0.898158</td>
</tr>
<tr>
<td>CH_Prague_cafe</td>
<td>CH_slunickari</td>
<td>6</td>
<td>0.904031</td>
</tr>
<tr>
<td>CH_slunickari</td>
<td>CH_Islamists</td>
<td>2</td>
<td>0.973459</td>
</tr>
<tr>
<td><strong>average</strong></td>
<td></td>
<td><strong>3.33</strong></td>
<td><strong>0.933519667</strong></td>
</tr>
</tbody>
</table>

Fig. 2.3. Graph obtained with the elaboration of the DNB system that highlights the causal chain of the use of words.

1.2.4. Relations between the analyzed days (23.07.2016 - 22.09.2016)

The MRG graph (Figure 2.4), on the dynamics of relationship between the days observed are related, highlights periods of maximum uniformity of behaviour, represented by the days with a dense network in interconnections (click1 and click2). As can be seen, the days of the focus 06/21/2016 and the day that follow, 06.22.2016 are in the graph areas less typical. The graph presents the days in which none of the selected words has appeared, with connections marked by dashes.
Fig. 2.4. Graph MRG obtained from the processing of AutoCM system that shows the verbal behaviour of similarity connections between the days analyzed. The red circle shows the focused day while the blue circle highlights the following day to day focused. The dashed circles show the clicks, days of uniformity of behaviour.

1.3. Croatia

1.3.1. Analysis of the frequency

With the analysis of the frequency (Tab. 3.1) and the distribution of the words during the days observed (Figure 3.1), various situations are noted

- A turbulent period that starts from 11-02-2016 and ends on 16-02-2016, where the frequency of the words ustashe and etnic is very high, and the word terrorists joins them in the a peak.
- Then we follow isolated peaks of the words ustashe and etnici.
- The day chosen 09-03-2016 and the following day are not related to peaks of frequency of particular words.
### Table 3.1. Table of frequencies

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>etnici</td>
<td>758</td>
</tr>
<tr>
<td>ustasho</td>
<td>579</td>
</tr>
<tr>
<td>terrorists</td>
<td>179</td>
</tr>
<tr>
<td>refugee_crisis</td>
<td>87</td>
</tr>
<tr>
<td>refugees</td>
<td>84</td>
</tr>
<tr>
<td>pederi</td>
<td>41</td>
</tr>
<tr>
<td>antifa</td>
<td>16</td>
</tr>
<tr>
<td>goat_fuckers</td>
<td>7</td>
</tr>
<tr>
<td>balije</td>
<td>7</td>
</tr>
</tbody>
</table>

**Fig. 3.1. Graph of the frequency of words in the observation period.**

### 1.3.2. Similarities in behaviour

In the graph MRG (fig.3.2) the word Terrorists occupies the central position. However, the graph presents mostly very low connections, with even negative values. Some ideologhemi are identified:

- **A significant relationship (0.75) between** the word **refugees** and **refugee_crisis**;
- followed by the relation (0.57) between the words **Terrorists** and **Ustashe**;
- and between **Terrorists** and **Goat-Fuckers** (0.32).
1.3.3. Cause-Effect Relationships

In the graph (Fig. 3.3) connecting the words in terms of cause and effect, we find a triangular structure in which ustache cause Terrorists (probability = 0.27), which refers to Etnici (probability = 0.61), which refers again to ustache (probability = 0.45).

On the other hand, the system identifies a chain-cause-effect structure among the remaining words, starting with a strong double relation (probability = 0.93) between the words "pederi" -> balije.

<table>
<thead>
<tr>
<th>cause</th>
<th>effect</th>
<th>reactivity</th>
<th>strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRO_terrorists</td>
<td>&gt; CRO_etnici</td>
<td>6</td>
<td>0.608278</td>
</tr>
<tr>
<td>CRO_goat_fuckers&gt; CRO_refugee_crisis</td>
<td>6</td>
<td>0.967383</td>
<td></td>
</tr>
<tr>
<td>CRO_ustashes</td>
<td>&gt; CRO_terrorists</td>
<td>2</td>
<td>0.26533</td>
</tr>
<tr>
<td>CRO_pederi</td>
<td>&gt; CRO_balije</td>
<td>2</td>
<td>0.930197</td>
</tr>
<tr>
<td>CRO_balije</td>
<td>&gt; CRO_pederi</td>
<td>1</td>
<td>0.387897</td>
</tr>
</tbody>
</table>

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
Table.3.2. Table obtained from the processing of the DNB system in which there are shown every word in the role of cause and / or in the role of effect. In the "reactivity" column you are shown the number of days after which the effect word is used and in the "strength" column the probability with which this happens.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Cause</th>
<th>Effect</th>
<th>Reactivity</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRO_refugees</td>
<td>&gt; CRO_ustashe</td>
<td>CRO_etnici</td>
<td>5</td>
<td>0.239584</td>
</tr>
<tr>
<td>CRO_ustashe</td>
<td>&gt; CRO_refugees</td>
<td>CRO_antifa</td>
<td>5</td>
<td>0.454825</td>
</tr>
<tr>
<td>CRO_antifa</td>
<td>&gt; CRO_goat_fuckers</td>
<td>CRO_refugees</td>
<td>10</td>
<td>0.788335</td>
</tr>
<tr>
<td>CRO_refugees</td>
<td></td>
<td>average</td>
<td></td>
<td>4.33</td>
</tr>
</tbody>
</table>

1.3.4. Relations between the analyzed days (07.02.2016 - 12.04.2016)

With regard to the dynamics of the relationships of observed days, in the MRG graph (fig.3.4), few days of maximum behavioural uniformity are represented at the centre, represented by reciprocal interconnections (click 1, 2, 3). However, if it’s observed the days that are part of the clicks, it’s noticed that they belong to days not close temporally. As can be seen, the day of the focus 09/03/2016 falls between these days (click 1). The following day 10-03-2016 is instead recognized...
as atypical, and placed in a more marginal area of the graph. However, even the latter is not associated with other days of the same period.

**Fig. 3.4.** Graph MRG obtained from the processing of AutoCM system that shows the verbal behaviour of similarity connections between the days analyzed. The red circle shows the focused day while the blue circle highlights the following day to day focused. The dashed circles show the clicks, days of uniformity of behaviour.

1.4. Greece

1.4.1. Analysis of the frequency

With the analysis of the frequency (*Tab. 4.1*) and the distribution of the words during the days observed (*Figure 4.1*), there are several peaks of frequencies, especially for the words **Refugees** and **Immigrants**. and less frequently for the words **Jihadists** and **Invasion**.
<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refugees</td>
<td>1293</td>
</tr>
<tr>
<td>Immigrants</td>
<td>745</td>
</tr>
<tr>
<td>Jihadists</td>
<td>104</td>
</tr>
<tr>
<td>Invasion</td>
<td>78</td>
</tr>
<tr>
<td>Rape</td>
<td>35</td>
</tr>
<tr>
<td>Criminals</td>
<td>33</td>
</tr>
<tr>
<td>Muslim</td>
<td>20</td>
</tr>
<tr>
<td>Terrorist</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 4.1. Table of frequencies

1.4.2. Similarities in behaviour

The MRG graph (fig.4.2), obtained with the weights of the AutoCM, shows a destructured situation. The only significant relation (0.62) is between Refugees and Rape, forming a idealohema. Refugees is connected, although weakly, with the word Criminals that, albeit at low frequency, however, is at the centre of the graph. Also, the word Immigrants is at high frequency, however, it is isolated and weakly linked with the word Rape.
The other words **Jihadists, Invasion, Muslim, and Terrorist** do not constitute a system. Each one is probably used in a diversified way.

![Graph MRG obtained with the elaboration of the AutoCM system, that shows the connections of similarity between the concept-words used. The size of the circles refers to the frequency of words. The values refer to the strength of similarity between words. In red the word (or words) in the centre of the graph.](image)

**Fig.4.2.**

### 1.4.3. Cause-Effect Relationships

Analyzing the graph (fig.4.3) that connects the words in terms of cause and effect, we find an articulated structure, in which the word **Terrorist**, directly activated (probability = 0.9) from the word **Muslim**, in turn activates the word **Rape**, or **Immigrants** and therefore **Criminals**.

However, the temporal distance of activation between a cause word and effect suggests a situation where either one or the other word comes into play.

<table>
<thead>
<tr>
<th>cause</th>
<th>effect</th>
<th>reactivity</th>
<th>strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRE_Terrorist</td>
<td>GRE_Rape</td>
<td>7</td>
<td>0.489065</td>
</tr>
<tr>
<td></td>
<td>GRE_Immigrants</td>
<td>8</td>
<td>0.440289</td>
</tr>
<tr>
<td>GRE_Muslim</td>
<td>GRE_Terrorist</td>
<td>5</td>
<td>0.903386</td>
</tr>
<tr>
<td>GRE_Jihadists</td>
<td>GRE_Jihadists</td>
<td>2</td>
<td>0.096752</td>
</tr>
</tbody>
</table>
Table 4.2. Table obtained from the processing of the DNB system in which there are shown every word in the role of cause and/or in the role of effect. In the "reactivity" column you are shown the number of days after which the effect word is used and in the "strength" column the probability with which this happens.

<table>
<thead>
<tr>
<th>Cause</th>
<th>Effect</th>
<th>Reactivity</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRE_Refugees</td>
<td>GRE_Invasion</td>
<td>5</td>
<td>0.5223</td>
</tr>
<tr>
<td>GRE_Invasion</td>
<td>GRE_Muslim</td>
<td>3</td>
<td>0.63594</td>
</tr>
<tr>
<td>GRE_Refugees</td>
<td>GRE_Invasion</td>
<td>5</td>
<td>0.331867</td>
</tr>
<tr>
<td>GRE_Immigrants</td>
<td>GRE_Criminals</td>
<td>10</td>
<td>0.821058</td>
</tr>
</tbody>
</table>

**average**       **5.625**   **0.53**

Fig. 4.3 Graph obtained with the elaboration of the DNB system that highlights the causal chain of the use of words.

1.4.4. Relations between the analyzed days (19.09.2016 - 17.11.2016)

At the centre of the graph MRG (Figure 4.4), which represents the dynamics of the relationships of the observed days, there is a core of days with maximum uniformity of behaviour, represented by reciprocal interconnections. As can be seen, the focus day 16/10/2016 falls between these days.
(click 1). Even the following day of 17/10/2016 is recognized as typical and placed in an area connected with the centre of the graph (click 2).

![Graph MRG obtained from the processing of AutoCM system that shows the verbal behaviour of similarity connections between the days analyzed. The red circle shows the focused day while the blue circle highlights the following day to day focused. The dashed circles show the clicks, days of uniformity of behaviour.](image)

**Fig.4.4.** Graph MRG obtained from the processing of AutoCM system that shows the verbal behaviour of similarity connections between the days analyzed. The red circle shows the focused day while the blue circle highlights the following day to day focused. The dashed circles show the clicks, days of uniformity of behaviour.

1.5. Italy

1.5.1. Analysis of the frequency

The table of frequencies (tab.5.1) records 6 words with high frequency, and 3 absolutely sporadic: **NOT_taking_our_jobs** (2), **Italians_first** (1) and **fake_refugees** (1). The structure of the graph showing the frequencies of the words during the observed days is very articulated (fig.5.1). The focus day 23-08-2016 and the following day 24-08-2016 are at the centre of a zone of turbulence, followed by a peak of the words **Immigrants**, **migrants** and **clandestine**.
Table 5.1. Table of frequencies

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>clandestine</td>
<td>8311</td>
</tr>
<tr>
<td>immigrants</td>
<td>6316</td>
</tr>
<tr>
<td>Islamics</td>
<td>6139</td>
</tr>
<tr>
<td>migrants</td>
<td>4599</td>
</tr>
<tr>
<td>refugees</td>
<td>3903</td>
</tr>
<tr>
<td>barges</td>
<td>1161</td>
</tr>
<tr>
<td>NOT_taking_our_jobs</td>
<td>2</td>
</tr>
<tr>
<td>Italians_first</td>
<td>1</td>
</tr>
<tr>
<td>fake_refugees</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 5.1. Graph of the frequency of words in the observation period.

1.5.2. Similarities in behaviour

The MRG graph (Fig.5.2) obtained with the weights of AutoCM, shows a structured and compact situation, where the 6 higher frequency words are closely related to each other, with high-value relationships. It then identifies an ideologhema extended to six words: *Migrants, immigrant, refugees, illegal, barges, Islamics.*
Fig. 5.2. Graph MRG obtained with the elaboration of the AutoCM system, that shows the connections of similarity between the concept-words used. The size of the circles refers to the frequency of words. The values refer to the strength of similarity between words. In red the word (or words) in the centre of the graph.

1.5.3. Cause-Effect Relationships

In graph (fig. 5.3) that connects the words in terms of cause and effect, it identifies an articulated structure, in which the word migrants is at the centre of a junction, which is activated directly by the word Immigrants. Migrants in turn, activates the high-frequency words, but according to an order with low probability effect:

Migrants -> (probability = 0.19) Refugees-> (probability = 0.36) Clandestine -> (probability = 0.44) barges

This structure is therefore not very significant.

<table>
<thead>
<tr>
<th>cause</th>
<th>effect</th>
<th>reactivity</th>
<th>strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT_refugees</td>
<td>IT_clandestine</td>
<td>4</td>
<td>0.364675</td>
</tr>
<tr>
<td>IT_migrants</td>
<td>IT_refugees</td>
<td>1</td>
<td>0.180322</td>
</tr>
<tr>
<td>IT_migrants</td>
<td>IT_fake_refuge</td>
<td>9</td>
<td>0.185184</td>
</tr>
<tr>
<td>IT_immigrants</td>
<td>IT_migrants</td>
<td>1</td>
<td>0.292303</td>
</tr>
<tr>
<td>IT_immigrants</td>
<td>IT_immigrants</td>
<td>1</td>
<td>0.319621</td>
</tr>
</tbody>
</table>
Table 5.2: Table obtained from the processing of the DNB system in which there are shown every word in the role of cause and / or in the role of effect. The "reactivity" column shows the number of days after which the effect word is used and in the "strength" column the probability of this happening.

Fig. 5.3: Graph obtained with the elaboration of the DNB system that highlights the causal chain of the use of words.

1.5.4. Relations between the analyzed days (25.07.2016 - 23.09.2016)

The MRG graph (fig.5.4), on the dynamics of the relationships of observed days, highlights at the centre of the graph some days of maximum uniform behaviour, represented by reciprocal interconnections (click 1, 2). As can be seen, the day of the focus 24/08/2016 falls between these days (click 1). The next day of 25-08-2016 is strictly connected with the following two days on 26-08-2016 and 27-08-2016 (click 2).
1.6. Romania

1.6.1. Analysis of the frequency

With the analysis of the frequency (Tab.6.1) and the distribution of the words during the days observed (Figure 6.1), it can be easily identify five scenarios that come out from the routine:

- The period on the graph that is marked with A, in which there is a peak frequency of the words Migrants and refugees.
- The period on the graph that is marked with B, which includes the focus day 24_03_2016 and the subsequent 25_03_2016, with successive peaks of the words Muslims, terrorists and refugees.
- The periods that are reported on the graph with C - D - E, with peaks of the word Muslims.
1.6.2. Similarities in behaviour

The graph MRG (fig.6.2), obtained on the weights of the AutoCM, shows a structured situation with high value relationships between the various words. At the centre there is the word refugees, which creates a trio of high relations (idealoghema), with the word fag (0.74) and the word migrants (0.75).

Refugees is articulated in a tree of relationships (idealoghema), maintained in sequence between the words Islamists, Muslims and Terrorists.
On a separate tree, Refugees is also related to the word Jew.

The word Fag besides constituting (as already seen) an idealoghema with refugees and migrants is a sort of crossroads. Fag is also related to the word gypsy and then to crows. Again the word Fag is related (0.75) with the word bozgor.

1.6.3. Cause-Effect Relationships

Analyzing the graph (fig.6.3) that connects the words in terms of cause and effect, we find a circular structure that involves all the selected words in the sequence: Islamists-> migrants-> bozgor-> crows-> gypsy-> refugees-> Jew-> fag-> Muslims-> Terrorists-> Islamists.

<table>
<thead>
<tr>
<th>cause</th>
<th>effect</th>
<th>reactivity</th>
<th>strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO_Islamists</td>
<td>-&gt; RO_migrants</td>
<td>5</td>
<td>0.905689</td>
</tr>
<tr>
<td>RO_Muslims</td>
<td>-&gt; RO_terrorists</td>
<td>1</td>
<td>0.69003</td>
</tr>
<tr>
<td>RO_migrants</td>
<td>-&gt; RO_bozgor</td>
<td>8</td>
<td>0.757389</td>
</tr>
</tbody>
</table>

Fig.6.2. Graph MRG obtained with the elaboration of the AutoCM system, that shows the connections of similarity between the concept-words used. The size of the circles refers to the frequency of words. The values refer to the strength of similarity between words. In red the word (or words) in the centre of the graph.
Table 6.2. Table obtained from the processing of the DNB system in which there are shown every word in the role of cause and/or in the role of effect. In the "reactivity" column you are shown the number of days after which the effect word is used and in the "strength" column the probability with which this happens.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Reactivity</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO_refugees -&gt; RO_Jew</td>
<td>1</td>
<td>0.731957</td>
</tr>
<tr>
<td>RO_terrorists -&gt; RO_Islamists</td>
<td>3</td>
<td>0.471572</td>
</tr>
<tr>
<td>RO_crows -&gt; RO_gypsy</td>
<td>9</td>
<td>0.816352</td>
</tr>
<tr>
<td>RO_gypsy -&gt; RO_refugees</td>
<td>8</td>
<td>0.571744</td>
</tr>
<tr>
<td>RO_fag -&gt; RO_Muslims</td>
<td>5</td>
<td>0.238086</td>
</tr>
<tr>
<td>RO_Jew -&gt; RO_fag</td>
<td>9</td>
<td>0.588007</td>
</tr>
<tr>
<td>RO_bozgor -&gt; RO_crows</td>
<td>3</td>
<td>0.55689</td>
</tr>
<tr>
<td><strong>average</strong></td>
<td><strong>5.2</strong></td>
<td><strong>0.63</strong></td>
</tr>
</tbody>
</table>

1.6.4. Relations between the analyzed days (23.02.2016 - 23.04.2016)

With regard to the dynamics of the relationships of observed days, the MRG graph (fig.6.4), few day shows a rather branched structure, where only at the centre of the branches there are few days of behavioural uniformity (click 1). The day of the focus 24-03-2016 is atypical, placed on the edge of a branch, however it is close to other similar days of the same period. The following day of 03/25/2016 is disposed on a leaf of another extreme branching, indicating also an atypical.
Fig. 6.4. Graph MRG obtained from the elaboration of the AutoCM system, which shows the similarity connections of verbal behaviour between the analyzed days. The red circle highlights the focused day while the blue circle highlights the day following the focused day. The dashed circles highlight the clicks, days of uniform behaviour.

1. 7. United Kingdom

1.7.1. Analysis of the frequency

The table of frequencies (tab.7.1) shows various very high frequency words, such as migrants, immigrants, deport and foreigners. The structure of the graph of words frequency, during the observed days, is very articulated (fig.7.1).

Very high peaks for the words: migrants, immigrants and deport. Only one very high peak in the day 27_06_2016 for the word vermin. The focus day of 23_06_2016 and the subsequent 24_06_2016 fall within the period of greatest turbulence.
<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>migrants</td>
<td>62636</td>
</tr>
<tr>
<td>immigrants</td>
<td>60117</td>
</tr>
<tr>
<td>deport</td>
<td>29620</td>
</tr>
<tr>
<td>foreigners</td>
<td>12303</td>
</tr>
<tr>
<td>vermin</td>
<td>9720</td>
</tr>
<tr>
<td>invaders</td>
<td>3287</td>
</tr>
<tr>
<td>romanians</td>
<td>2186</td>
</tr>
<tr>
<td>rapefugees</td>
<td>236</td>
</tr>
<tr>
<td>norefugees</td>
<td>124</td>
</tr>
<tr>
<td>gimmigrants</td>
<td>38</td>
</tr>
<tr>
<td>fakefugees</td>
<td>24</td>
</tr>
</tbody>
</table>

*Table 6.1. Table of frequencies*

**Fig.7.1. Graph of the frequency of words in the observation period.**

### 1.7.2. Similarities in behaviour

The graph MRG (fig.7.2), obtained with the AutoCM weights, shows a structured situation with high value relationships between some of the selected words. At the centre the word immigrants, that create two diversified structures with the word norefugees.

In one side of the graph there is a first ideologhema in which the words immigrants and norefugees are associated with the words gimmigrants, invaders and migrants.
In another side of the graph there is a second ideologhema, where the words **immigrants** and **norefugees** are associated with the words **rapefugees** and **foreigners**.

The other words are weakly associated or not associable (negative value).

![Graph MRG](image)

**Fig. 7.2.** Graph MRG obtained with the elaboration of the AutoCM system, that shows the connections of similarity between the concept-words used. The size of the circles refers to the frequency of words. The values refer to the strength of similarity between words. In red the word (or words) in the centre of the graph.

### 1.7.3. Cause-Effect Relationships

<table>
<thead>
<tr>
<th>cause</th>
<th>effect</th>
<th>reactivity</th>
<th>strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK_invaders</td>
<td>UK_fakefuguees</td>
<td>1</td>
<td>0.052591</td>
</tr>
<tr>
<td>UK_fakefuguees</td>
<td>UK_invaders</td>
<td>1</td>
<td>0.800274</td>
</tr>
<tr>
<td>UK_norefugees</td>
<td>UK_gimmigrants</td>
<td>3</td>
<td>0.827413</td>
</tr>
<tr>
<td>UK_rapefugees</td>
<td>UK_norefugees</td>
<td>7</td>
<td>0.227719</td>
</tr>
<tr>
<td>UK_vermin</td>
<td>UK_rapefugees</td>
<td>4</td>
<td>0.522684</td>
</tr>
<tr>
<td>UK_migrants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK_foreigners</td>
<td>UK_romanians</td>
<td>2</td>
<td>0.326449</td>
</tr>
</tbody>
</table>

*Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931*
Table 7.2. Table obtained from the processing of the DNB system in which there are shown every word in the role of cause and/or in the role of effect. In the “reactivity” column you are shown the number of days after which the effect word is used and in the “strength” column the probability with which this happens.

<table>
<thead>
<tr>
<th>Word</th>
<th>Next Word</th>
<th>Days</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK_romanians</td>
<td>UK_deport</td>
<td>3</td>
<td>0.545728</td>
</tr>
<tr>
<td>UK_deport</td>
<td>UK_immigrants</td>
<td>6</td>
<td>0.601633</td>
</tr>
<tr>
<td>UK_immigrants</td>
<td>UK_vermin</td>
<td>2</td>
<td>0.223538</td>
</tr>
<tr>
<td></td>
<td>UK_foreigners</td>
<td>10</td>
<td>0.252079</td>
</tr>
</tbody>
</table>

average                     3.636  0.419

In the graph (fig.7.3) that connects the words in terms of cause and effect, a circular structure is identified, that involves the words in ordered sequence:

**immigrants->foreigners->deport->romanians->immigrants**

Immigrants also determines the word vermin, which has as its effect on the one hand the generic word migrant, and on the other hand the word vermin opens to a sequence:

vermin->rapefugees->noreferrer->gimmigrants.

The words fakefugees and invaders constitute a mutual reference structure: fakefugees<->invaders.

---

Fig. 7.3. Graph obtained with the elaboration of the DNB system that highlights the causal chain of the use of words.
1.7.4. Relations between the analyzed days (24.05.2016 - 24.07.2016)

The MRG graph (Figure 7.4), on the dynamics of relationship between the days observed, shows a rather branched structure, where are identified typical days of behavioural uniformity (click 1,2,3,4). The focus day 23-06-2016 is positioned close to the following day of 24-06-2016 and is arranged in a structure (click3) near the centre of the graph.

Fig. 7.4. Graph MRG obtained from the elaboration of the AutoCM system, which shows the similarity connections of verbal behaviour between the analyzed days. The red circle highlights the focused day while the blue circle highlights the day following the focused day. The dashed circles highlight the clicks, days of uniform behaviour.

2. Comparative analysis

The comparative analysis among all the countries required the use of a technique of "fusion" of the data that in this project finds its first empirical application. For more information of this technique, please refer to the documents in the annex.
2.1. Fusion G-IN-OUT

With the analysis of the words selected by the various countries, at least 3 shared semantic areas are identified:

- **Semantic Area A**, attributable to the words: migrant / migrants / immigrants of which at least one is present in all countries except in CRO.
- **Semantic Area B**, attributable to the words: terrorism / terrorists / terrorist of which at least one is present in the countries BU, CH, CRO, GRE, RO, while it has not been selected by IT and the UK.
- **Semantic Area C**, referable to the words: Islam (CH) / Islamist (BU) / Islamists (RO) / Islamics (IT) / Muslim (GRE) / Muslims (RO) / Jihadists (CH-GRE) / Taliban (BU) / balije (CRO) (rude word for Muslims) present in all countries.

As for the semantic area A, attributable to the words: migrant / migrants / immigrants, the graph G-IN-OUT (fig.2.2) seems to indicate that these terms have often a generic use, without necessarily referring to a negative connotation, linked to the other selected words. For the countries BU, RO, and UK the term is linked to 2 terms of IT (in fig.2.2 with brown dashed circle) at very low...
This relationship is therefore not significant. Also the term of **GRE immigrants** is connected with the term of the **CH** (in fig.2.2 with red dashed circle) at very low frequency (frequency = 2) and therefore with non-significant relation. In all these cases, moreover, these words are a leaf of the graph and therefore they are isolated from the other terms. The term **migrants** of **CH** is linked to the term **CRO ethnici**, but the latter is also connected to one of the very low-frequency **IT** terms. Instead, for **IT** the two terms immigrants and migrant are significantly linked to the term **RO terrorists**. **IT immigrants** also constitutes an ideologhema with the word **IT refugees** while **ITmigrants** constitutes an ideologhema with the word **IT_clandestine**.

For what concerns the semantic area B, it is attributable to the words: **terrorism / terrorists / terrorist** present in the countries **BU, CH, CRO, GRE, RO**. These terms for the countries **BU, CH** and **GRE** are connected on the graph G-IN-OUT (fig.2.3) directly to the term **CH Islamists**. While the term **CROterrorists** is in turn connected directly to the term **IT Islamics**.

---

**Fig.2.2.** Graph G-IN-OUT highlights (orange dashed circle) words referring to the same semantic area: migrant / migrants / immigrants. In the table on the left, in the first column there is the symbol of each country, followed by the frequency and therefore the percentage of frequency of the word with respect to the total frequency of words in the same country.
The only ones who do not have the word **terrorism** / **terrorists** / **terrorist** are IT and the UK. Note that in the graph most of the terms chosen by IT have direct connections with the term **terrorists** of the RO, and the term **IT Islamics** is directly connected with the term CRO **Terrorist**.

For the UK this happens for the terms **vermin**, **deport** and **norefugees** directly connected with the term **GRE Terrorist**.

For what concerns the semantic area C, it is attributable to the words: **Islam (CH)** / **Islamist (BU)** / **Islamists (RO)** / **Islamics (IT)** / **Muslim (GRE)** / **Muslims (RO)** / **Jihadists (CH-GRE)** / **Taliban (BU)** / **balije (CRO)**. All that words are originated in three countries positioned on the same tree of the graph G-IN-OUT (fig.2.4): CH with **Islam**, BU with **Islamist**, CRO with **balije** (rough word for Muslims), which mean that these words in the three countries are used in a similar way.

The word **RO Muslims** re-presents the word **RO terrorists**, already identified in the single country's MRG graph.

---

*Fig.2.3. Graph G-IN-OUT highlights (orange dashed circle) words referring to the same semantic area: terrorism / terrorists / terrorist. In the table on the left, in the first column there is the symbol of each country, followed by the frequency and therefore the percentage of frequency of the word with respect to the total frequency of words in the same country.*
The term **IT Islamics** is linked to the term **CRO terrorists**. With the term **CH Islamists** are connected with the concept of terrorism both the **CH** itself, the **GRE** and the **BU**. The connections with the other countries are less relevant.

![Diagram](image)

**Fig.2.4.** Graph G-IN-OUT highlights (orange dashed circle) words referring to the same semantic area: Islam (CH) / Islamist (BU) / Islamists (RO) / Islamics (IT) / Muslim (GRE) / Muslims (RO) / Jihadists (CH-GRE) / Taliban (BU) / balije (CRO). In the table on the left, in the first column there is the symbol of each country, followed by the frequency and therefore the percentage of frequency of the word with respect to the total frequency of words in the same country.

### 2.1.1. Fusion G-IN-OUT - Data analysis of each country in the global graph

In this section we analyze how the words of each country are distributed in the graph G-IN-OUT, compared to the terms of the other countries. The ideologues in the graphs MRG of the each country tend to remain stable even in the European context. The words of different countries connected to each other in the graph have the meaning to highlight the similar use that the two countries make of those words.
Let's begin with Bulgaria (BU) (fig.2.5). The three words **BU illegal**, **migrant** and **refugee** are also connected in the European context. The connection with other numerous words from other countries in the central star of the graph is not significant, because the word IT at the centre has very low frequency (frequency = 1).

The word **BU Islamist**, little related to the context of the other words of the BU, is connected to the words **CH Islam** and **CRO balije** (rude word for Muslims), meaning a similar use of these words by the 3 countries. Also the word **BU nigger**, is not tied to the context of the other words of the BU, and it is connected to the words **CRO antifa** (derisive term for liberal or left-wing members of community) and **UKfakefugees**.

On the other side of the graph there are the words **BU threat**, **turn_soap** and **Taliban** through the word **GRE Terrorist** that creates an HUB, also with other words, especially from the UK.

---

**Fig.2.5.** Graph G-IN-OUT highlights (orange dashed circle) words referring to the same semantic area, selected by Bulgaria (BU). In the table on the left, in the first column there is the symbol of each country, followed by the frequency and therefore the percentage of frequency of the word with respect to the total frequency of words in the same country.
Let's move to the **Czech Republic (CH)** (fig. 2.6). A part of the terms related to CH remain linked to each other, in particular the terms: **terrorism**, **Islamists** and sluninckari and the word **refuges**, albeit connected to the GRE Terrorist. The word CH Islam is connected to the word BU Islamist. The word **migrants** to the CRO etnici. The word CH Prague_cafe to the UK rapefigees.

![Diagram](image)

Fig. 2.6. Graph G-IN-OUT highlights (orange dashed circle) words referring to the same semantic area, selected by Czech Republic (CH). In the table on the left, in the first column there is the symbol of each country, followed by the frequency and therefore the percentage of frequency of the word with respect to the total frequency of words in the same country.

Let's move on to **Croatia (CRO)** (fig. 2.7). The CRO terms **refugees** and **refugee_crisis** remain linked to each other (as they were strongly linked in the MRG graph on the individual country) but are strongly disconnected from the context of other European terms and constitute itself a graph. The CRO terms **ustashe** and **terrorist** also remain connected. To the latter term terrorist is connected the IT term **Islamics**. The CRO word **ethnici** is connected to the CH word **migrants**. The term CRObalije (rude word for Muslims) is connected on one side to the words BU Islamist and CH Islam, and on the other side to the words RO Jew and then UK invaders.
Let's move on to Italy (IT) (fig.2.8). Five of the six IT most frequently words (refugees, migrants, immigrants, clandestines and barges) appear in a branch of the graph that converges to the word RO terrorists. Just like the word IT_Islamics, it is associated with the word CROterrorists. In the set of IT words there is not the word Terrorist and in the graph we see how it is "borrowed" from other countries (RO and CRO). The incorporation of the IT words into a specific branch of the graph suggests that there is no sharing in the way these terms are used with those used by other European countries. The other very low-frequency IT terms are at the centre of the graph, but their relations in the European context are not very significant.
Let's move to Romania (RO) (fig.2.9). The ideologhema ROTerrorist - Muslim, presents in the graph MRG of the single country RO, remains stable also in this G-IN-OUT graph of the European context. To this ideologhema are connected the most relevant words of IT and also the term GRE Refugees.

Even the ideologhema RO migrants - refugees remains firm, but disconnected from the European context, as it is associated in the graph with an IT word with very low frequency. The RO word crows is associated with other words in the European context: UK foreigners, CROpederi, BU threat. The word RO Jew not very well inserted with the other RO words, finds in this European context a position between the words UK invaders and CRO balije.

The connections between the other RO words and the European context are not relevant.
Let's move on to the **United Kingdom (UK)** (fig.2.10). In general, the words **UK** are placed at the edge of the G-IN-OUT graph, and do not constitute a system between them or with other European words. The only relevant relationship is between the word **UK gimmigrants** and **BU out**. Also worthy of mention is the relationship between the **UK invaders** and **RO jew**.

All the 3 words **UK vermin, deport, norefugees** flow into the low-frequency **GRE terrorist** word and therefore this relationship is not significant. The other words **UK fakefugees, romanians, migrants, rapefugees** converge equally in low-frequency words.
Fig. 2.10. Graph G-IN-OUT highlights (orange dashed circle) words referring to the same semantic area, selected by United Kingdom (UK). In the table on the left, in the first column there is the symbol of each country, followed by the frequency and therefore the percentage of frequency of the word with respect to the total frequency of words in the same country.

2.2. Fusion MRG
The graph emphasizes the most significant relationships between the words used by all the various EU countries. At the centre of the graph we can find all those words that, although chosen by each project partner, were found with low frequency, for **IT**: `NOT_taking_our_jobs` (2), `Italians_first` (1) and `fake_refugees` (1), for **CH**: `Islamists`, for **GRE**: `Terrorist`, for the **UK**: `fakefugees`.

The **IT** words with high frequency are all arranged in a dedicated area in the graph, symbolizing an Italian style use of these words, that deviate from the use made by other European countries. The set of these words constitute an ideologhema in a system of combinations.

For the other countries, however, some ideologhemas are remaining, mostly composed of two words that are connected with other ideologhemas or single words of other countries.

### 3. Machine Learning used for this analysis

#### 3.1. Auto Contractive Map (Auto-CM)

The Auto Contractive Map (Auto-CM) neural network is characterized by a three-layer architecture (Figure 1.1.): an Input layer, where the signal is captured from the environment, a Hidden layer, where the signal is modulated inside the CM, and an Output layer, by means of which the CM influences the environment on the basis of previously received stimuli. Each layer contains an equal number of N units. The connections between the Input and the Hidden layers are mono-dedicated, whereas those between the Hidden and the Output layers are at maximum gradient.

![Fig.1.1. Architecture of the Auto-CM.](image_url)
The learning algorithm of CM may be summarized in a sequence of four orderly steps:

1. Signal Transfer from the Input into the Hidden layer
2. Adaptation of the connection values between the Input and the Hidden layers
3. Signal Transfer from the Hidden into the Output layer
4. Adaptation of connection value between the Hidden and the Output layers.

During the training phase, they develop only positive values for each connection. Therefore, they do not present inhibitory relations among the nodes, but only different strengths of excitatory connections.

After the learning process, each Input vector belonging to the learning set will generate a null Output vector. Therefore, the energy minimization of the training vectors is represented by a function through which the trained connections completely absorb the Input formation vectors.

The matrix $w_{ij}$ (connection weights between the Hidden and the Output layers) represents the Auto-CM knowledge about all the dataset. It is possible to convert the matrix $w_{ij}$ into a squared matrix $d_{ij}$ of distances among nodes.

The new matrix $d$ is again a squared symmetric matrix, where the main diagonal entries are null (i.e., they represent the zero distance of each variable from itself) and where the off-diagonal entries represent “non Euclidean distances” between each couple of variables.

Each distance $d_{ij}$ between a pair of nodes may therefore be regarded as the weighted edge between these pair of nodes in a suitable graph-theoretic representation, so that the matrix $d$ itself may be analyzed through the graph theory toolbox.

A graph is a mathematical abstraction that is useful for solving many kinds of problems. Fundamentally, a graph consists of a set of vertices and a set of edges, where an edge is an object that connects two vertices in the graph. More precisely, a graph is a pair $(V, E)$, where $V$ is a finite set and $E$ is a binary relation on $V$, to which it is possible to associate scalar values (in this case, the distances $d_{ij}$). $V$ is called a vertex set whose elements are called vertices. $E$ is a collection of edges, where an edge is a pair $(u, v)$ with $u, v$ belonging to $V$. In a directed graph, edges are ordered pairs, connecting a source vertex to a target vertex. In an undirected graph, edges are unordered pairs and connect the two vertices in both directions, hence in an undirected graph $(u, v)$ and $(v, u)$ are two ways of writing the same edge.

The Graph Minimum Spanning Tree (MST) solves a problem defined as follows: find an acyclic subset $T$ of $E$ that connects all of the vertices in the graph and whose total weight (viz., the total distance) is minimized.
From a conceptual point of view, the MST represents the energy minimization state of a structure. In fact, if we consider the atomic elements of a structure as vertices of a graph and the strength among them as the weight of each edge, linking a pair of vertices, the MST represents the minimum of energy needed so that all the elements of the structure preserve their mutual coherence. In a closed system, all the components tend to minimize the overall energy. So the MST, in specific situations, can represent the most probable state for the system to tend. To determine the MST of an undirected graph, each edge of the graph has to be weighted.

3.2. Maximally Regular Graph (MRG)

The Graph Minimum Spanning Tree (MST) represents the nervous system of any dataset. In fact, the summation of the strength of the connection among all the variables represents the total energy of that system. The MST selects only the connections that minimize this energy. Consequently, all the links shown by MST are fundamental, but not every fundamental link of the dataset is shown by MST.

Such limit is intrinsic to the nature of MST itself: every link able to generate a cycle into the graph is eliminated, however its strength. To avoid this limit and to explain better the intrinsic complexity of a dataset, it is necessary to add more links to the graph according to two criteria:
- the new links have to be relevant from a quantitative point of view;
- the new links have to be able to generate new cyclic regular microstructures, from a qualitative point of view.

Consequently, the MST Tree-graph is transformed into an undirected graph with cycles. Because of the cycles, the new graph is a dynamic system, involving in its structure the time dimension. This is the reason why this new graph should provide information not only about the structure but also about the functions of the variables of the dataset.

To build this new graph we need to proceed in this way:
- assume the MST structure as a starting point of the new graph;
- consider the sorted list of the connections skipped during the MST generation;
- estimate the H Function of the new graph each time we add a new connection to the MST structure, to monitor the variation of the complexity of the new graph at every step.
So, we have named **Maximally Regular Graph (M.R.G.)** the graph whose **H Function** is the highest, among all the graphs generated adding to the original MST the new connections skipped before to complete the MST itself.

Consequently, the MRG is given by the following equations:

\[ H_i = f(G(A_p, N)); \quad /* \text{Generic Function on a graph with } A_p \text{ arcs e } N \text{ Nodes */} \]

\[ H_i = \frac{\mu_p \cdot \varphi_p - 1}{A_p}; \quad /* \text{Calculation of H Function, where } H_o \text{ represents MST complexity */} \]

\[ \text{MRG} = \text{Max}\{H_i\}. \quad /* \text{Graph with highest H */} \]

\[ i \in [0, 1, 2, \ldots, R]; \quad /* \text{Index of H Function*/} \]

\[ p \in [N - 1, N, N + 1, \ldots, N - 1 + R]. /* \text{index for the number of graph arcs */} \]

\[ R \in \left[0, 1, \ldots, \frac{(N - 1) \cdot (N - 2)}{2}\right]; /* \text{Number of the skipped arcs during the M.S.T. generation */} \]

**Equations to calculate MRG**

The “R” variable is a key variable during the MRG generation. “R” variable, in fact, could be also null, when the generation of MST implies no connections to be skipped. In this case, there is no MRG for that dataset. The “R” variable, further, makes sure that the last and consequently the weakest connection added to generate MRG is always more relevant that the weakest connection of MST.

The MRG, finally, generates, starting from the MST, the graph presenting the **highest number of regular microstructures using the most important connections** of the dataset. More the H Function selected to generate the MRG is high, more meaningful the microstructures of the MRG.

### 3.3. Dynamic Networks Block (DNB)

**Dynamic Networks Block (DNB)** is a processing system able to detect and cluster the complex cause-effect relationships in discrete dataset representing a multivariate temporal process. An evolutionary algorithm (GenD) select the optimum step of influence between each variable and all (or each one of) the others during a temporal process. The fitness is controlled by a set of equations measuring the temporal correlation and the probability density function among the variables, in order to define the proper likelihood of cause-effect relationship among the variables themselves.
DNB, at the end of its search, generates a direct graph of the variables able to explain and predict the global dynamic of the analyzed process.

DNB algorithm is composed of three parts:

a. Equations to detect the local cause-effect relations between the variables in the temporal flow;

b. Equations to establish the global and prevalent cause-effect relations between the variables in the whole temporal multivariate process;

c. Equations to select the most likelihood cause-effect relations between variables at different scale of time in order to cluster them and to project them in direct graph, sparse or connected.

DNB system can be implement with three different algorithms, in order to respond to three different questions:

a. **DNB Scalar**: how the variables are linked in a cause-effect relationship at step 1, using a Multi Scale Entropy (MSE) algorithm to sample the whole process at different time scale;

b. **DNB Vector**: how the variables are linked in a cause-effect relationship when each variable may be the cause of the others with specific temporal step. An evolutionary algorithm (GenD) be used to execute this optimization;

c. **DNB Matrix**: how the variables are linked in a cause-effect relationship when each variable may act on each one of the others with a different temporal step. A complex evolutionary algorithm (Complex GenD) will be used in order to execute this optimization.

### 3.4. Adaptive Fusion System (from: Theory of Impossible Worlds (TIW))

In relational database theory entities can be fused by joining operations. Without filtering (no where condition), the join can be a cartesian product, for each observation of the first entity, so many combined observations are generated, how many observations are in the second entity, with each resulting combined observation having all the attributes of both, unless specifically removed. If, however, there is reason to put match criteria between some attributes during the join (where condition), only the filtered observations are generated which respect the match criterion. The other pairs of observations, for which there has been no match, may or may not be returned, depending on the type of join. In the case of inner join, no observation is returned that does not match the defined criteria. In the case of left join, all the first entity observations that does not match the the second ones according to the criteria are also returned; in these cases attributes coming from the first entity will be valued, while those coming from the second are defined as null values. Conversely, in the
case of right join, all the second entity observations which does not match the first ones are also returned, nulling the attributes coming from the first entity and repeating the ones coming from the second. Finally, with the full outer join, both types of left and right match types of observations are returned.

Combining data in relational databases is usually done through inner join on keys between tables that represent different aspects of the same phenomenon. The relational keys establish in these cases precise connections between observations in the various entities. Any related observations constitute an enrichment of the information content of the others. When relations become low, inner joins return a few combined observations and, to have results, match criteria must be made less restrictive; every observation of an entity refers to usually more than one observation on the others. Informative content is enriched in a weaker way. When then the criteria are completely absent, the cartesian product represents a fusion, but of all with all, without substantial enrichment of the information content of the observations between them. When known relationships do not allow you to get more than one cartesian product between two entities, to enrich the information content of both, you need a way to make them otherwise interact.

In relation to this issue, one of the most challenging targets in Machine Learning research is then the data fusion between datasets in terms of attributes (variables) and observations (records), whether the two or more entities are related to the same phenomenon, or when they are only partially related, that even when they are not related at all. It is important to emphasize that the in latter case, the coupling between sources relating to different phenomena extends the concept of fusion to the connection of different worlds. In the classical theory of possible worlds and its updating an individual can be transferred from a source world (dataset A) to a destination world (dataset B) if at least one of its attributes is shared in both worlds, not in case of void intersection. In this paper a way to make such impossible transfer actually possible is presented. We have named this approach new Theory of Impossible Worlds (TIW for short). It can be applied to both data sets on the same domain and heterogeneous dataset. As in human reasoning, where we often use mental models drawn from the analysis of a certain phenomenon to metaphorically infer properties of a completely different phenomenon, we would like to be able to perform the same kind of inferences in the machine learning domain. TIW does exactly this, helping us to empower machine learning environments with the capacity to fuse different databases and variables belonging to different phenomena into a same cognitive space, so that it is possible to examine each phenomenon from the structural viewpoint of the other. If this leads to a significant improvement of our capacity to carry out pattern recognition in the specific domains of the source phenomena, then this way of
proceeding makes sense and is conceptually useful. We will show that this is actually the case for very diverse examples. This approach proves to be particularly useful in social sciences where most interesting phenomena are not amenable to empirical analysis for lack of data on joint occurrences, but where partial data for sub-phenomena are often available.

References

- P.M. Buscema, *Dynamic Networks Block Algorithm (DNB)*, Rome 2017, Semeion Pre - Print.

Coalition of Positive Messengers to Counter Online Hate Speech - JUST/2015/PRAC/AG/BEST/8931
RECOMMENDATIONS

- The EU policies developed so far are aimed at protecting minorities, which are socially exposed to the phenomenon of hate, which is expressed in various ways. The adoption of directives in each country should be devised with attention towards the emerging phenomena of easily identifiable intolerance, thus proceeding to an institutionalized listening, rather than the simple ex post regulation adopted against the irreversible effects;

- The analyses conducted over the phenomenon of Hate Speech have highlighted that it is a very polychrome phenomenon and that above all it tends to manifest itself with a wide range of terms, which can give some interpretations, if and only if, analysed simultaneously and for a long period of time, facilitating the drafting of policies effective and permanent in the various regional contexts;

- The sustainability of the principles' policy that are implemented is an important aspect that could be confirmed if there is a permanent structure for the monitoring of the various expressions related to Hate Speech, with techniques and methodologies for the comparison of data over time;

- Social media are dominated by few players, a context whose boundary with hate speech is very vulnerable. The EU has established guidelines, but there is a lack of action for a strong communication and monitoring among social media players;

- Each social media player should report its control policies towards adverse dialogues and, if possible, should cooperate for the creation of a single decalogue. The publication of this decalogue would also act as an indirect driver for the promotion of a global citizenship;

- Monitoring the use of adverse dialogues on the internet leads to the creation of new competences: the social auditor, experts in IT and social and legal themes;

- Hate Speech can be countered through a conscious use of social platforms, but it is necessary to invest in dedicated trainings on social media and informatics, especially for all those people who do not belong to the category of digital natives;
The countering of Hate Speech can have some general positive effects, as the protection of minorities could foster an environment of greater trust in public institutions;

A possible lever for contrasting Hate Speech is related to the deconstruction of collective imagery, created on the population exposed to the phenomenon of hate. The deconstruction, through the identification of the components of diversity that lead to the generation of hate, could instead be the foundation of social integration and therefore the creation of a global citizenship.