News Media Coverage of Refugees in 2016:
A GDELT Case Study

Emina Boudemagh, Izabela Moise
Computational Social Science
ETH Zurich, Switerland

Abstract

For many citizens, the main source of information about world events is the news media. Undeniably, media coverage of topics influences the perceived importance of these topics. Recent media discourse has given extensive attention to the topic of refugees and migration. In this context, we conduct a study on news media coverage of refugees for the time period of January 1st 2016 to December 31st 2016. We rely on the Global Database of Events, Language, and Tone (GDELT) project that monitors and analyses news articles around the world. In a first phase, we explore the global media conversation on refugees along two important dimensions: article quantity and sentiment. We identify and characterize events that generated an extensive media coverage and also extreme sentiment in the news reports. In a second phase, we refine the analysis by focusing on the news media coverage related to refugees in Europe. GDELT features allow us to identify the key countries and time evolution of the media coverage. Lastly, we determine the main actors linked to refugees in the news and study their interaction through a network analysis.

Introduction

News media plays a central role in linking individuals to global events and governmental activities (Robinson 2001). Media coverage of a topic increases the importance of that topic among members of the audience (Wood and Peake 1998). However, the news conveys not only facts but also biases via article sentiment and word selection, becoming instrumental in individual attitude formation. In other words, media influences what we think about and how.

In the last two decades, the number of people displaced worldwide has dramatically increased, rising from 37.3 million in 1996 to 65.3 million in 2015, as reported by official UNHCR studies (UNHCR 2016). To put numbers into perspective, one in every 122 humans is now either a refugee, internally displaced or seeking asylum (UNHCR 2016). In Europe in particular, more than a million migrants and refugees arrived in 2015 (UN 2016) creating division in the European Union over how best to deal with accommodating people. Undeniably, the refugee crisis is a humanitarian, political and societal challenge of great importance that has been extensively covered by the news.

Previous literature has investigated the media influences on public opinion on immigration and immigrants. In (Boomgaarden and Vliegenhart 2009), the authors study the effects of news coverage about immigration in Germany and found that frequency and tone of coverage affected anti-immigration attitudes. (Marcel Lubbers 2000) established that exposure to certain Dutch newspapers (i.e., those characterised by negative reporting on immigrants) significantly increased ethnic threat perceptions. Similar studies (Gilliam and Iyengar 2000; Domke 2001; Schemer 2012) revealed that negative media environments produce negative immigrant stereotyping. The UNHCR has expressed its concerns on refugee media representation in the news and its consequences on public opinion (Berry, Garcia-Blanco, and Moore 2016).

In this context, we conduct a study on news media coverage of refugees for the time period from January 1st 2016 to December 31st 2016. The goal of this study is to characterize the refugee media discourse by identifying the global features that define it. For this purpose, we rely on the Global Database of Events, Language, and Tone (GDELT) project that monitors and analyses news articles around the world. The main goal of the GDELT project is to compile a massive inventory of political events that is updated on a daily basis. We utilize three main GDELT algorithms for our purposes: theme extraction determines if "refugees" is a main topic in the article; fulltext geocoding identifies locations and central actors; sentiment analysis provides the tone for each article.

Our study proceeds in three main phases. In a first phase, we explore the global media discourse on refugees along two important dimensions: article quantity and the sentiment of news reports. By combining different features of GDELT, we are able to identify several events that were linked with refugees during 2016. Some of these events received a lot of attention in the media, characterized by a high number of articles, and others a shift in article sentiment. GDELT allows us to characterize the nature of these events, providing initial insights into why sentiment and quantity are influenced differently by refugee-related events. In a second phase, we explore the European perspective of the refugee media conversation. We pinpoint the European countries highly associated with refugees in the news and argue it stems for geographical, political and specific event-related reasons. Time evolution of the number of articles provides

the detailed visibility of each country throughout the year and uncovers similarities and differences in media coverage. Some countries are continuously linked to the subject of refugees while others are cited sporadically but with high intensity. We highlight primary hints for country-based "media response". We search for further evidence of similarities and differences in article quantity and sentiment by considering four of the countries with the most extensive media coverage on refugees in 2016: Turkey, Greece, Germany and the United Kingdom. Lastly, we identify and characterize the influential actors (persons and organizations), as extracted from refugee media reports, by conducting a network analysis. Our findings indicate that both person and organisation networks exhibit the small world property (Watts and Strogatz 1998). A potential consequence is that a spreading process, such as information spreading, happens much faster than in networks that do not display this property, by reaching all the nodes in the network rapidly. We uncover mainly political actors such as decision-makers and electoral candidates as well as intergovernmental institutions. We also find that news media organizations and social media platforms are important nodes.

GDELT Database

The Global Database of Events, Language, and Tone (GDELT) monitors and analyses web, print and broadcast news from around the world with the purpose of compiling a massive inventory of political events that is updated on a daily basis. Introduced in 2013, the GDELT dataset had recorded up to now nearly 250 million events going back to the 1979. Each news report monitored by GDELT is first machine-translated into English then passed as input into a complex pipeline of algorithms. These algorithms extract metadata which codify 58 fields: from the location where an incident took place to what sort of event it was (from protest to peace appeals), to ethnic and religious affiliations, among other categories. This codified metadata (but not the actual text of the articles) is then released as an open data stream, providing a multilingual annotated index of the world's news. In the following, we provide a short description of the key features of GDELT and focus on the analytical tools relevant for our purposes.

News Sources GDELT includes all international and national news coverage from global news agencies (Agence France-Presse, The Associated Press and Xinhua - the official and largest Chinese news agency) and other major news sources (AfricaNews, BBC Monitoring, Christian Science Monitor, Facts on File, Foreign Broadcast Information Service, United Press International, Washington Post, New York Times, and Google News).

Actor coding Actor coding In GDELT implements the identification of key individuals, countries, political figures that are discussed in the article. This is achieved by cross-referencing actors with an expanded *agents* dictionary containing around 60,000 entries. In case an actor's name is not found in the dictionary, a new entry is created in another dictionary that compiles a list of 'generic agents' (listing approximately 1,500 entries). Hence, the GDELT algo-

rithm does not discard articles with new actors. Next, articles go through several processing iterations in order to improve the accuracy of associating actors with locations: (i) raw text, (ii) actors replaced with known countries, (iii) cities replaced with countries, and (iv) both actors and cities are replaced. Actors are assigned to countries if such contextual information is available within the text (see Location subsection for more information). For example (provided by GDELT documentation), in the actor-replacement iteration, the second sentence of "Egyptian Minister of Foreign Affairs Mohamed Orabi attended the summit yesterday. While he was there, Orabi pledged support for..." is replaced with "While he was there, Egypt pledged support for..". This process improves the matching process of diplomatic events. To explain in more detail, consider the following example from an actual excerpt analyzed by the GDELT algorithm:

Example (Schrodt, 2012): "The United Nations will provide nearly 25,000 tons of emergency food aid to refugees fleeing the civil war in Liberia , the World Food Program (WFP) said on Monday."

According to the GDELT algorithm, (red) indicates the main actor within the sentence, in this case the UN, (green) indicates the action, and (blue) indicates the recipient, which is categorized as refugees.

Location Articles run through a full-text geocoding that identifies the geo-location of every actor (Leetaru 2012). The algorithm first identifies all geo-references in the article; then, each identified actor is assigned to the nearest geo-reference. While this may seem simple, the reported accuracy of the algorithm is quite high (Leetaru 2012), (Leetaru and Schrodt 2013). One example provided in the documentation is if the President of Russia attends a meeting in Washington D.C., then the algorithm assigns actor 1's country as Russia and not Washington D.C. If no country reference is specified in the article and/or if no cross-reference is available, then the geo-reference to the actor is left unspecified.

Sentiment algorithm Another important aspect of the GDELT algorithm is the sentiment algorithm. For each article, the sentiment is measured via a sentiment mining algorithm (Hu and Liu 2004) identifying negative and positive words in the text, based on a positive-negative lexicon dictionary. The sentiment value is obtained by subtracting the number of negative words to the number of positive words. By averaging the sentiment over all articles referring to the same event, we obtain the sentiment of the event. Scores range from -100 to +100, where 0 is neutrality.

Additional Tonal Characteristics It is important to note that the GDELT dataset consists of two versions: the GDELT Knowledge Graph and the GDELT Event Database. The first consists of daily files with more complex algorithms such as article polarity (i.e. in-article sentiment variance). The GDELT Event Database, however, spans across all 2002-2015 but only performs the standard sentiment algorithm. For the purpose of our study, we use the GKG database and the analytical tools it provides. Note that we refer to the "tone" metadata computed by GKG, as *sentiment* through-

out the rest of the paper.

GDELT filtering Considering the time range of this study and its focus on the news media, we rely on the GKG database in order to benefit from the theme extraction feature and the stronger location detection and sentiment analysis algorithms. The database was imported using GDELT's GKG exporter, filtering all events presenting the keyword "refugees" in any of the fields. The temporal filter was set to range from January 1st to December 31st 2016. A total dataset of 1.1 million entries was obtained and later aggregated by day for analysis and visualization purposes. The method used to build the actor and organization graphs was different: network files were directly produced from the GKG by GDELT analytical service, maintaining the same filtering criteria.

Using GDELT for research Several research studies make use of the extensive database and ambitious coverage that GDELT has to offer. GDELT has been compared to other databases such as ICEWS (Ward et al. 2013) and EventRegistry (Kwak and An 2016). Both study conclude that GDELT data volume is higher and usually convey the same information on major events than its rivals. It was also stated that using GDELT for geo-spatial analysis in research lead to uncertainties at the sub-national level (Hammond and Weidmann 2014). In close connection with our work, GDELT was used to investigate biases in news reporting of disasters (Kwak and An 2014) unveiling strong regionalism in news geography. Main use of GDELT in research focuses on conflicts analysis and detection.

Global Media coverage and event detection

Extensive media coverage supplies media consumers with salience cues regarding the relative importance of the issues that are being discussed. Along with the number of articles, the sentiment carried in news reports, whether neutral, negative or positive, has a potential impact on opinion formation. Based on these insights, we explore refugee media reports published from January 1st 2016 to December 31st 2016 to detect events that caused a high number of articles or a shift in the sentiment of the articles. We utilize three main GDELT analyses: theme extraction determines if "refugees" is a main topic in the article; fulltext geocoding identifies locations, central actors and their corresponding country; sentiment analysis provides the average sentiment for each article. Figure 2 shows the average article quantity and sentiment, aggregated across all countries, throughout the year of 2016 for articles referring to refugees.

Events with high media salience

According to Figure 2, the average quantity of articles remains relatively constant throughout 2016, with the exception of three sudden increases. In order to investigate what events triggered these peaks, we make use of an important information provided by GDELT: the most mentioned countries and actors at those particular dates. An example of this valuable data is presented in the form of word clouds in Figure 1 for the event identified on *February 4 2016* (Event 1.).

(a)   (b)

(c)

Figure 1: Word cloud of most mentioned organizations (a), persons (b) and locations (c) in articles linked to refugees on *February 4 2016*, as extracted from GDELT. Among the most cited are Julian Assange, United Nations, United Kingdom, the UN working group on Arbitrary Detention and the Ecuadorian Embassy allowing for an event detection on that date: the UN Working Group on Arbitrary Detention ruled in favour of diplomatic refugee Julian Assange, residing at the Ecuadorian Embassy in London since 2012.

Peaks that are close in time were identified as a unique peak but holding several events.

This approach allowed us to pinpoint the following events:

1. *February 4*, the UN Working Group on Arbitrary Detention ruled in favor of Wikileaks founder Julian Assange, who claimed that he was unlawfully detained at the Ecuadorian Embassy in London since 2012, after been granted diplomatic asylum. Among the most mentioned locations and actors are the United Kingdom, Julian Assange, the United Nations, the UN Working Group on Arbitrary Detention and the Ecuadorian Embassy.

2. *September 17-19*, three bombs exploded and several explosive devices were found in New York and New Jersey. The suspected perpetrator was arrested two days later.

   September 20, Barack Obama addressed his final speech to the United Nations, mentioning refugees several times and encouraging nations to welcome them.

   September 21, Donald Trump linked the New York attacks to the admission of refugees.

   Most cited entities in this time period are the United States, Barack Obama, the United Nations, Donald Trump and the name of the arrested terrorist.
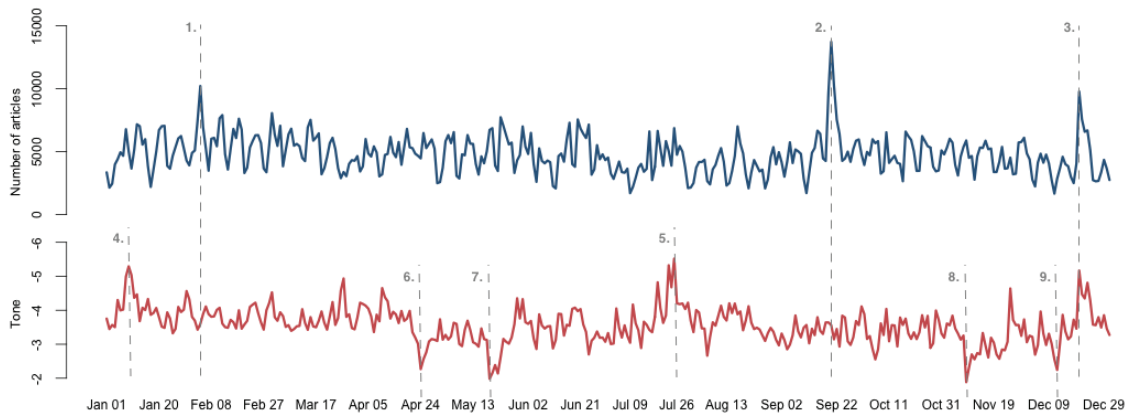
Figure 2: Global average article quantity (top) and sentiment (bottom) on the theme of refugees per day in 2016. Nine events related to refugees in the media and causing high media salience or sentiment shifts are identified by investigating the peaks.

3. *December 19*, the Berlin Christmas market attacks and Angela Merkel's speech on the following day. Germany, Angela Merkel, Twitter and the Kaiser Wilhelm memorial church, the name of a relative of the murdered truck driver were extensively cited by the news on that day.

Events shifting media sentiment

The average sentiment of refugee media reports is mostly negative and exhibits seven noticeable peaks: three negative drops and four positive jumps.

Negative drops   A similar analysis of the sentiment peaks identifies two events, in addition to the previous ones (the third negative peak corresponds to December 19-20):

4. Beginning of *January*, molests on women on New Year's Eve in Cologne, Germany, triggered an increase in the negative sentiment carried by refugee news reports.
   January 7, a suspected terrorist is arrested in California. He had come to the United States as a refugee.
   During this period news mentions mostly the United States, the attorney office in Sacramento, Germany, the name of the suspect and Angela Merkel.

5. *July 22*, a shooting occurred in the vicinity of a shopping mall in Munich, Germany. This terrorist act was wrongly attributed to refugees by the news media, causing a significant drop in news sentiment.
   July 24, a suicide bombing was carried out by a refugee close to Nuremberg. During the following days, Germany, the name of the Munich Police President and the Bavarian Interior Minister as well as Facebook, McDonald's and the name of the Shopping mall were highly cited by the press.

Positive jumps   A similar approach is employed to identify the positive jumps in 2016 news media coverage. They correspond to the following events:

6. *April 23*, European Union leaders praise Turkey's efforts in welcoming refugees.
   April 24, Barack Obama opens the Hannover Messe in Germany by praising Angela Merkel for her welcoming

attitude towards refugees and method employed to handle the escalating migrant situation in the European Union.

7. *May 19*, TV host Karl Stefanovic criticizes Australian Immigration Minister Peter Dutton for the comments he made on the dangers of accepting "illiterate" refugees. *May 22*, Obama visits Vietnam prior to his last G-7 meeting Japan where the refugee crisis was discussed.

8. *November 9*, Donald Trump is elected president of the United States. A disengagement towards refugees was at the heart of his electoral campaign.

9. *December 12*, Former UNHCR president Antonio Guterres is elected the UN General Secretary.

Qualitative characterization of events

By combining different features of GDELT, we were able to identify several events that were linked with refugees during 2016. In the group of events receiving a high media attention (a high number of articles), we find events directly related to refugees, such as Julian Assange's case and Obama's speech mentioning refugee policies. Events reporting on attacks were linked in the media with refugees. When analysing media sentiment shifts, we observe that negative drops are associated with terrorism events and deadly attacks. The case of the Berlin attack in December 2016 generated a high number of articles as well as a sentiment drop, due to the severity of the event. On the other hand, positive jumps characterize mostly political speeches and issues. The sentiment expressed by the actors themselves is strongly reflected in the sentiment of the article: Obama praising Merkel, international community applauding Turkey's role in welcoming refugees.

Media coverage of refugees in Europe

The overall analysis of the global refugee media coverage described in the previous section, was performed at the aggregate level across all countries. In the following, we zoom into the European perspective of refugee media conversation. In a first step, we decompose the refugee media signal

at the country level, in terms of article quantity and sentiment. We further proceed with a fine-grain country-specific analysis with a focus on four of the countries with the most extensive media coverage on refugees in 2016.

Throughout the rest of the article, every mention of the word "countries" should be interpreted as "countries mentioned by the media" and not necessarily countries hosting the media itself.
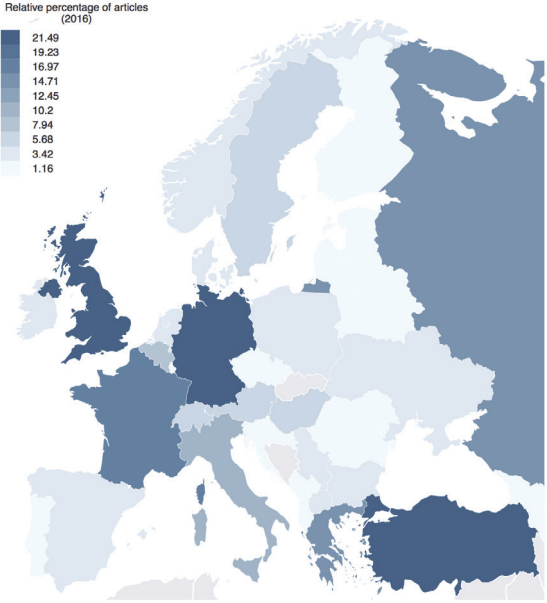
Geographical coverage

Figure 3: Spatial distribution of the average relative percentage of articles mentioning a European country and refugees in 2016. The most visible countries are Germany, Turkey, the United Kingdom, France, Greece, Russia and Italy.

Figure 3 illustrates the spatial distribution of the average number of articles mentioning European countries and refugees, for the year of 2016. The most visible countries are Germany, Turkey, the United Kingdom, France, Greece, Russia and Italy. On average refugees represent between 22% (Germany) and 11% (Italy) of their yearly news media coverage. The figure highlights the heterogeneous visibility of European countries in the news media when it comes to refugees. We explain this heterogeneity by relying on political and geographical insights: we identify European border countries such as Greece, Italy and Turkey. In addition, countries such as Turkey, Germany, Sweden and Russia, were the top refugee hosting countries in 2015, highly involved in debates around refugee policy making. Based on the salient event identification presented in Section , the high refugee media coverage in United Kingdom and Germany may also be explained by the fact that these countries represent the locations of the previously identified events.
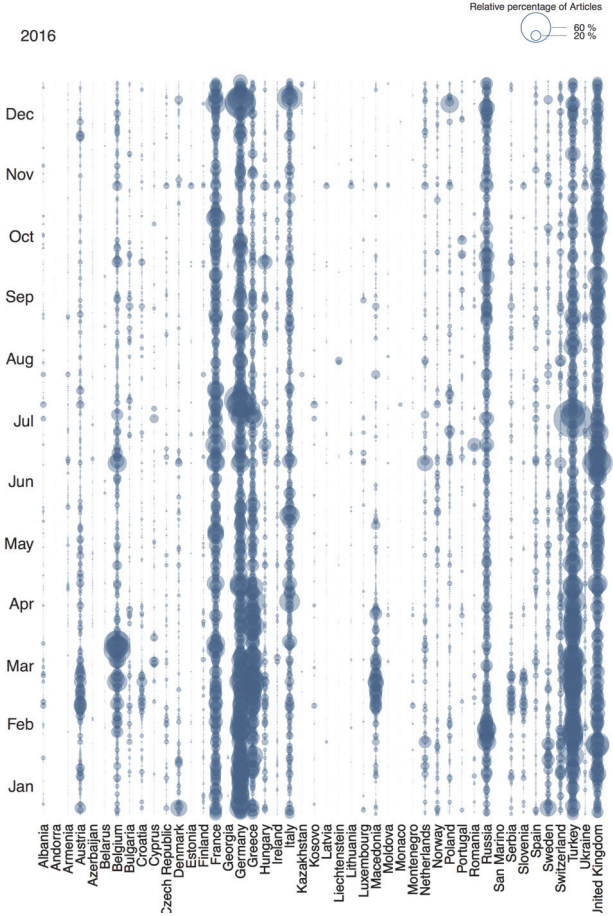

Figure 4: Time evolution of the percentage of articles mentioning a European country and refugees each day of 2016. France, Russia and Italy are continuously highly cited during the year while Belgium displays sudden increase in visibility during attacks in Brussels end March. Germany, Turkey and the United Kingdom undergo both high coverage all year long and additional sudden increases.

Time evolution

A more detailed plot is described in Figure 4, displaying the temporal evolution of the number of articles, decomposed at the country-level over the year. The figure reveals the heterogeneity in visibility for European countries over time. Countries such as France, Russia and Italy are continuously mentioned during the year, while others such as Belgium, Austria and Macedonia are cited sporadically but in a high number of articles. In the latter case, shifts are triggered by real-world events that brought the refugee topic in the media spotlight: the Brussels attacks at the end of March, the Turkey attacks in July. Moreover, Germany, Greece, Belgium, Turkey and the United Kingdom exhibit both a continuously high media coverage and sudden shifts in intensity. Most of them were uncovered by the event detection. Countries like Denmark, Croatia, Norway, Poland exhibit
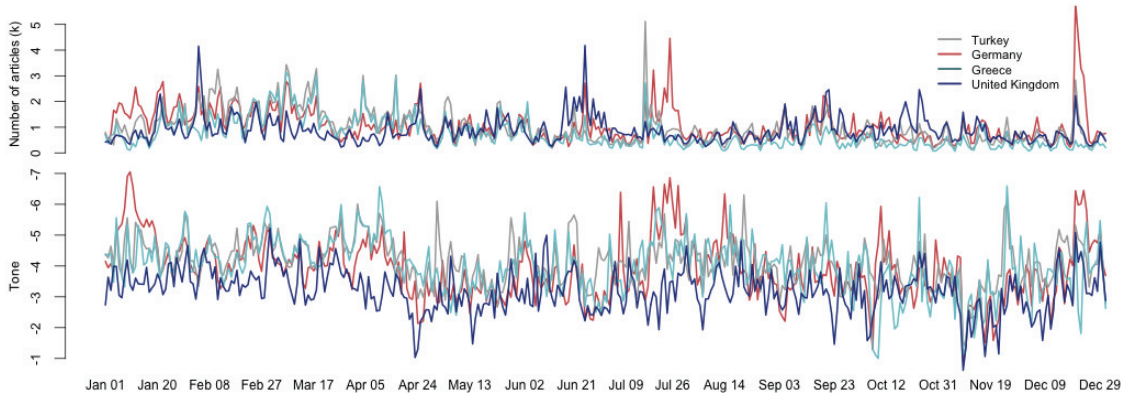
Figure 5: Global country-specific article quantity (top) and sentiment (bottom) per day on the theme of refugees in 2016 for Turkey, Germany, Greece and the United Kingdom. Turkey and Greece have a similar coverage and sentiment in the United Kingdom is generally more positive. It appears article quantity accounts for event location.

a relative constant evolution in time, with low association to events related to refugees. Sweden, among the European countries hosting most refugees in Europe, after Turkey, shows relatively low association with refugees in the media. This figure also allows for detection of simultaneous occurrence of a subset of countries for a certain event. This simultaneity calls for further fine-grain investigation. In particular, it might constitute an argument in saying that the location of the article also represents the hosting country of the news media itself. It would then be possible to study the "media response" to an event at the country-level.

Focus on four salient countries

Based on these insights, we zoom into country-specific media coverage for countries with the most extensive refugee media reports: Germany, United Kingdom, Turkey and Greece. They represent between 25% (Greece) and 39% (Turkey) of the total number of articles covering the refugee topic.

We present the timeseries for article sentiment and article quantity for the year 2016 in Figure 5. A similar approach to the one adopted in the previous section could uncover more events from timeseries peaks. However, we choose to investigate correlations between countries. The goal of this analysis is to find possible similarity or heterogeneity in media visibility across countries. Table 1 summarizes the results for the correlation analysis in tone and number of articles between countries. A first observation is that Turkey and Greece indeed share the highest correlations (84% in number of articles and 65% in sentiment). A possible explanation for the similar behavior adopted by media in both countries, stems from their geographical location: both are European border countries, thus directly concerned by incoming refugees matters. We also observe a generally more positive tone for the United Kingdom. These findings seem to indicate that GDELT "number of articles" feature, used at the country level, can be an interesting proxy for detecting events happening in a specific country. However, sentiment seems to be a characteristic of the country-specific media itself. These observations strengthen the similar findings of the previous section.

	Turkey	Germany	Greece	U.K.
Turkey		0.59	0.84	0.35
Germany	0.50		0.46	0.42
Greece	0.65	0.49		0.25
United Kingdom	0.43	0.48	0.34	

Table 1: Correlation between countries with respect to number of articles (shaded cells) and sentiment (regular cells). Significance for all values: $p \leq 0.001$ (***)

Influential Actors

In the following, we direct our focus on discovering what are the important actors in the global media refugee conversation and how they are interconnected. To address these questions, we leverage the GKG tool for building the co-occurrence network of persons and organizations, as extracted from refugee media reports. A co-occurrence network is defined as the collective interconnection of actors (persons or organizations) based on their paired presence within the same news article.

Identifying and characterizing influential Actors

The co-occurrence networks of persons and organizations are shown in Figure 6. It comes as no surprise that the most visible persons are political figures and also religious figures (Pope Francis). Among them we distinguish political leaders that are also policy makers (such as Angela Merkel, Barack Obama, Tayyip Erdogan, Francois Hollande, Vladimir Putin etc.) and U.S. presidential candidates that addressed immigration and refugee concerns in their electoral campaign (Donald Trump and Hilary Clinton), forming a strong closed triad with Barack Obama. Among important organizations, we find the European Union, the United States and the United Nations are highly interconnected. Interestingly, major news media actors such as Reuters and the Associated
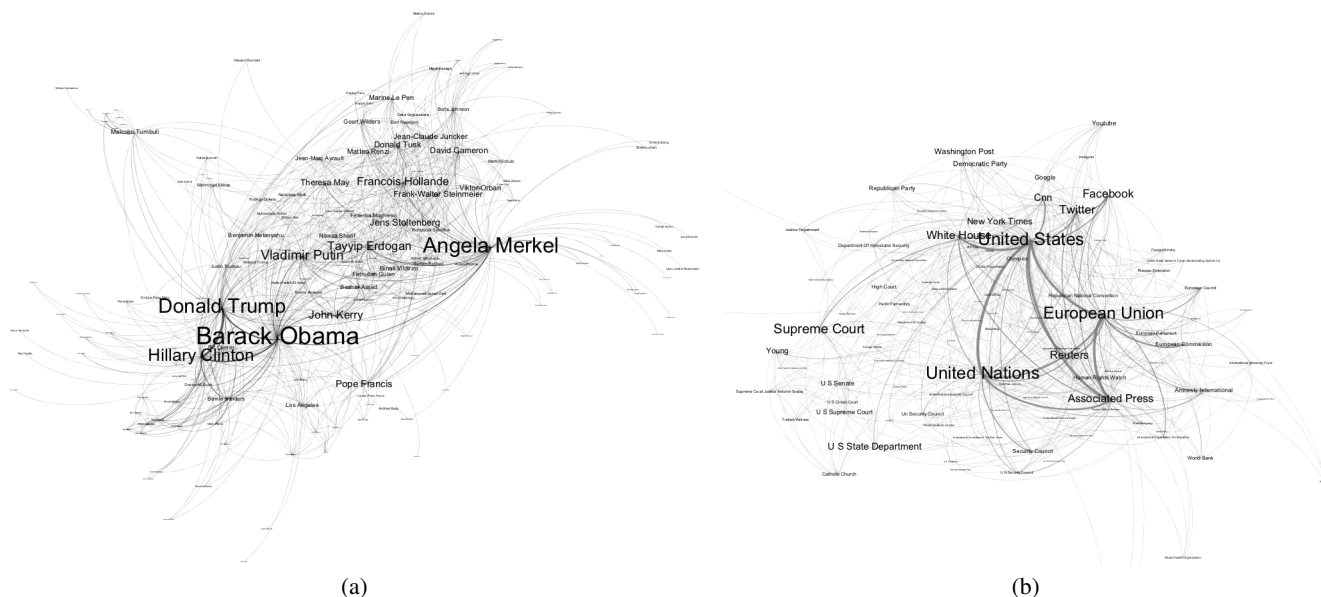
(a)   (b)

Figure 6: Networks of co-occuring actors. Node label size represents the degree of the node. Edge width reflects the weight of the link between the connected nodes. Persons (a) are mostly political actors and Organizations (b) governmental institutions. Social and news media are tighlty connected to important institutions.

Press seem to play a very influential role when it comes to refugees, strengthening the motivation of our study. Social media platforms (Twitter, Facebook) are also present.

Network analysis

To further characterize the structure of the co-occurrence networks, we compute several network metrics, which are summarized in Table 2.

Metric	Persons	Organizations
number of nodes	206	214
number of edges	1008	1131
diameter	3	3
average path length	1.96	1.95
modularity	0.145	0.046
connected components	101	30
average clustering coefficient	0.833	0.854
number of triangles	7596	4506

Table 2: Network metrics for the co-occurrence networks of persons and organizations

Network metrics for the co-occurrence graph of persons indicate a network diameter (defined as the longest of the shortest distances between all pairs of nodes in the network) equal to 3 and an average path length of 1.96. Both values are relatively small compared to the size of the network. Another noticeable remark is the high value (0.83) of the average clustering coefficient (a measure of the degree to which nodes in a graph tend to cluster together), along with a high number of triangles (7596). These observations indicate that the network of persons exhibits the *small world* property. Many real work networks exhibit what is called the *small*

world property, in which most vertices can be reached from the others through a small number of edges. In other words, despite their large size, most real world networks have a relatively short distance between any two nodes. This concept originated from the famous experiment made by Milgram in (Milgram 1967), who found that two US citizens chosen at random were connected by an average of six acquaintances.

A similar analysis can be applied to the network of organizations, leading to the same conclusion: the network also exhibits the small world property. As the two networks have similar number of nodes and edges and both exhibit the small world property, it is important to also evidentiate the differences between them. Most metrics have similar values for the two networks, with the exception of modularity and connected components. Both metrics have higher values for the actor network, indicating that the persons are better connected than organizations.

Conclusion

This study investigates the news media coverage of refugees in 2016 by relying on the GDELT project. GDELT compiles an enormous inventory of political events, providing valuable metadata for each event. For the purpose of our analyses, we use features such as location, actors, theme, number of articles and sentiment.

At the global level, our findings show that events causing negative drops in sentiment are deadly events and terrorists attacks and events causing positive jumps are political issues. In terms of number of articles, it is still to be investigated what triggers these peaks as media salience is an important feature of the agenda-setting literature. Few of the events identified had refugees as main actors. A closer

look into Europe uncovered heterogeneity in the association of countries with refugees. The countries that were mostly linked to refugees were Turkey, Germany, the United Kingdom, Greece, France, Italy and Russia. We argue it accounts for geographic, political and event-related reasons. Among other examples, Russia and France display a constant and high association with refugees in the media over time while Belgium shows low association but high increase in article quantity during attacks in Brussels end of March 2016. Germany and Turkey exhibit both characteristics. We also found that we can consider article quantity as a primary proxy for identifying the event location, while article sentiment represents a characteristic of the country's media itself. This constitutes a good starting point for studying country-specific "media response" in future research. Networks of main actors of the refugee media coverage exhibit the small world property. We identify mostly political actors: current decision-makers such as Angela Merkel but also the American 2016 election candidates connected though a strong triad with Barack Obama. Institutions show few strongly connected central actors : the United States, The United Nations and the European Union. Social media platforms such as Twitter and Youtube and news media institutions such as Reuters and the Associated Press play an important role and are tightly connected to principal institutions.

In a report investigating the news media coverage of refugees in five European countries (Berry, Garcia-Blanco, and Moore 2016), the UNHCR states that : "It is impossible to ignore the role of the mass media in influencing public and elite political attitudes towards asylum and migration. The mass media can set agendas and frame debates." Moreover, the report comes to the conclusion that media coverage varies from one country to another, in terms of word selection, general tone and precedence. The study presented in this paper therefore sets the basis for addressing an important question: is there an interaction between media response and policy change? Given the potential influence of media on opinion formation, what role does media play in the act of policy making?

Acknowledgments

References

Berry, M.; Garcia-Blanco, I.; and Moore, K. 2016. Press coverage of the refugee and migrant crisis in the eu: a content analysis of five european countries. Technical report, The UN Refugee Agency, Geneva.

Boomgaarden, H. G., and Vliegenhart, R. 2009. How news content influences anti-immigration attitudes: Germany, 1993-2005. *European Journal of Political Research* 48(4):516–542.

Domke, D. 2001. Racial cues and political ideology. *Communication Research* 28(6):772–801.

Gilliam, F. D., and Iyengar, S. 2000. Prime suspects: The influence of local television news on the viewing public. *American Journal of Political Science* 44(3):560–573.

Hammond, J., and Weidmann, N. B. 2014. Using machine-coded event data for the micro-level study of political violence. *Research & Politics* 1(2):2053168014539924.

Hu, M., and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, 168–177. New York, NY, USA: ACM.

Kwak, H., and An, J. 2014. *A first look at global news coverage of disasters by using the GDELT dataset*, volume 8851 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag. 300–308.

Kwak, H., and An, J. 2016. Two tales of the world: Comparison of widely used world news datasets GDELT and eventregistry. *CoRR* abs/1603.01979.

Leetaru, K., and Schrodt, P. A. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. *ISA Annual Convention* 2(4).

Leetaru, K. H. 2012. Fulltext geocoding versus spatial metadata for large text archives: Towards a geographically enriched wikipedia. *D-Lib Magazine* 18(9):5.

Marcel Lubbers, Peer Scheepers, M. V. 2000. Exposure to newspapers and attitudes toward ethnic minorities: A longitudinal analysis. *Howard Journal of Communications* 11(2):127–143.

Milgram, S. 1967. The Small World Problem. *Psychology Today* 2:60–67.

Robinson, P. 2001. Theorizing the influence of media on world politics models of media influence on foreign policy. *European Journal of Communication* 16(4):523–544.

Schemer, C. 2012. The influence of news media on stereotypic attitudes toward immigrants in a political campaign. *Journal of Communication* 62(5):739–757.

UN. 2016. International migration report, 2015. Technical report, United Nations, Geneva.

UNHCR. 2016. Global trends: Forced displacement in 2015. Technical report, The UN Refugee Agency, Geneva.

Ward, M. D.; Beger, A.; Cutler, J.; Dickenson, M.; Dorff, C.; and Radford, B. 2013. Comparing gdelt and icews event data. *Analysis* 21:267–297.

Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of'small-world'networks. *Nature* 393(6684):409–10.

Wood, B. D., and Peake, J. S. 1998. The dynamics of foreign policy agenda setting. *American Political Science Review* 92(1):173–184.