

# My EU = Your EU? Differences in the Perception of European Issues Across Geographic Region

Koustav Rudra, Gerhard Backfried, Miroslav Shaltev, Claudia Niederée, and Erick Elejalde

**Abstract**—Our perception of the situation in a country or a region is strongly influenced by the reflection of this situation in mass and social media channels. This effect is even more pronounced for geographically and culturally distant regions, for which no first hand experience is available. To avoid information overload, news outlets typically filter the available news from foreign countries based on the expected interest of the target audiences. Such filtering imposes an inherent bias in the reporting and can create a distorted perception of a region among the consumers of news of other regions. This might lead to misunderstandings between countries and unsubstantiated political and individual decisions (e.g., in the context of migration). In this paper, we systematically analyze the bias created in news reports. We consider Europe, or more precisely the European Union (EU) as our *zone of concern*, and examine its image in the media (news outlets) of other regions, EUROPE(NON-EU), AFRICA, ASIA, MIDDLE-EAST, AMERICA, and OCEANIA. An analysis of the year 2018 (January-December 2018) of news published in those regions reveals marked differences in the editorial policies and presented narrative when dealing with EU-related news. We observe a significant variation in the sentiment polarity of the reported EU-related stories between European and other regional news outlets. We further analyze the polarity variation among different sub-regions of large geographical areas such as AFRICA, ASIA, and AMERICA. We observe a contrasting difference in their editorial policies. This trend also holds for news related to different topics such as politics, business, economy, health, international relation etc.

**Index Terms**—News article; media bias; perception; EU

## 1 INTRODUCTION

With globalization, the world has to be increasingly regarded as a complex system of interacting and interdependent national entities and supranational organizations such as the European Union (EU). A substantiated and unbiased mutual understanding of the situation in the individual countries is crucial for decision- and policy-makers (DPMs) as well as for citizens, to ensure sound decisions and actions, which contribute to the aspired continued positive development of society.

However, DPMs often do not (only) respond to objective facts, but are influenced by an image of a situation [1]. Although there is some evidence of the prominent role that media coverage plays in national-level policy practices [2], our primary concern is on the formation of a “popular image”. Entities affected by implemented policies (i.e., citizens) react according to their perception of the world. Among other factors, the mass media play a central role in shaping the perceived reality [3]. Moreover, this induced image can also be influenced by power structures controlling the news outlets [4]. In turn, the concept of universal franchise, and with this, the possibility for citizens to influence DPMs through, for example, votes, protests and social media activity (i.e., react on incumbents actions), stresses the importance for the media to depict the reality in a clear, unbiased

way. Furthermore, it emphasizes the importance of carefully analyzing the bias introduced by the media.

In the international sphere, the problem of divergent internal and external popular images is of more direct concern for the DPMs. For example, in cases where it endangers their influence, bilateral economic and cooperative treaties, or the effectiveness of public diplomacy towards conflicted neighbors [5], [6]. At a global level, incomplete presentation of facts or their filtering to advance a narrative may transform a natural phenomenon such as migration into a problem for countries of origin, migrants, as well as for target countries. An overly-positive image of a region might motivate individuals to migrate only to face a very different reality in the receiving country. Clashes between perception and reality can prevent integration in the hosting society or even foster polarization or radicalization. The media contribute to mold the opinion of the local audience. A conflict between the internal and external views of an issue can lead to precarious developments. For example, according to a report on media coverage of the “refugee crisis” in Europe [7], the press played a central role in shaping the public opinion and declaring the “refugee crisis”. As Europe still faces multiple challenges from this crisis [8], it is of great interest to continue studying its image.

According to [9], three main elements should be considered in the study of EU external perception. The first one is the study of EU imagery in the national (and regional) news media (e.g., press, radio and television). The other two elements identified by the authors are the gaining of insights into the public’s opinion and making an accurate assessment of the views of the national (and regional) DPMs. In this paper, we aim to address the first element by applying an exhaustive and scalable computational methodology. Our

- Koustav Rudra, Miroslav Shaltev, Claudia Niederée, and Erick Elejalde are with the L3S Research Center, Leibniz University Hannover, Germany and Gerhard Backfried is with HENSOLDT Analytics, Austria.  
E-mails: {rudra, shaltev, niederee, elejalde}@l3s.de, gerhard.backfried@hensoldt-analytics.com

approach leverages the automatic processing of large collections of documents, which also allows for spatial scaling.

We perform a detailed analysis of the representation of the EU in the news across multiple geographical regions (EUROPE(NON-EU), ASIA, AFRICA, MIDDLE-EAST, AMERICA, and OCEANIA) over one year (January-December 2018). For a deeper understanding of coverage, we also analyze the news, which are directly adopted from the EU news sources (i.e., overlapping coverage with EU-internal media). Furthermore, we compare the coverage of EU-related news across eleven broad topics in three dimensions, namely, volume, sentiment polarity, and editorial strategies. We carry out our evaluation by exploring the differences and possible bias not only in the general news context but also by focusing on specific topics. In particular, we are interested in indicators that contribute to creating a perception of the Quality of Life in Europe [10] (e.g., economy, health, education, leisure, security, etc.). However, we also include other general topics that may act as pull and push factors according to migration-related literature [11], [12]. As European countries are a frequent target for migration, we expect these topics to be of interest to the international press.

Our main contributions in this work are as follow:

- 1) We propose a large scale quantitative approach to the investigation of the image of Europe in the mass media of other geographic regions.
- 2) We demonstrate to what extent news outlets of other regions report news also published by European media.
- 3) We identify differences in the sentiment polarity of the EU-related external coverage and how it compares to the internal EU reporting. These differences are studied in the combined corpus of news, but also for specific events that were reported both in the EU media and elsewhere.
- 4) We perform a topic level(‘business and economics’, ‘politics and economics’, ‘health and medicine’ etc.) analysis over different regional media.
- 5) We explore the differences in the news polarity in different sub-regions of AFRICA, ASIA, and AMERICA. This also reveals the sub-regional bias which exists within large geographical regions.

An in-depth analysis reveals that there is significant variation in the polarity of the reported news between Europe and other regions. Different regions paint a different image of the EU for their audiences and we observe how this effect also varies across the topics<sup>1</sup> reported upon.

## 2 BACKGROUND AND RELATED WORK

The perception of Europe, both from inside and from abroad, has been studied in a qualitative manner in several studies[13], [9], [5], [6], [14]. Typical for the qualitative approach, the focus is on interviews and the analysis of small data samples. In contrast, quantitative research, as it is used in our work, offers the opportunity of large-scale analysis. This is facilitated by automated analysis methods, e.g., relying on machine learning techniques, for inspecting higher

volumes of data. Despite the current availability of massive amounts of digital records of news reporting, quantitative analyses of the reflection of Europe in the mass media are scarce. In this work, we introduce a methodology for the quantitative investigation of the reflection of European topics in the news as a foundation for a better understanding of the perception of this region. Notwithstanding, we try to leverage some of the valuable insights produced by the qualitative literature on the subject.

### 2.1 Qualitative studies

#### 2.1.1 *The impact of essential events*

An influential set of qualitative studies dealing with the image of Europe and of the European Union (EU) is presented in [15]. When combined, the studies in this book cover data and events recorded in more than a decade of EU transformation and evolution (i.e., 2000’s). How essential events, such as the Lisbon Treaty and Eurozone debt crisis, impacted the role and image of Europe as a global actor, is analyzed in detail. Similar to these studies, in our analysis we understand ‘*image*’ as ‘the total cognitive, effective, and evaluative structures of the behavior unit, or its internal view of itself and its universe’ [1].

#### 2.1.2 *Analysis of topic selection and reporting*

Several other studies approach the perception of the EU by comparing the selection and reporting of topics in the news. For example, a meta-analysis of several qualitative studies focusing on the coverage of European topics in the media suggests that prior to the enlargement of the EU, these subjects were underrepresented in favor of national topics [13]. All of the studies examined by the authors compare European topics being reported in the media in different European countries. In [16], the author recognizes the emergence of the news media as actors capable of influencing the opinions of citizens. Furthermore, a topic-wise analysis of the Europeanization trend in different national newspapers has been pursued. As found in a study of the picture of one specific migrant community in three major British newspapers [17], the choice of topics and the context in which facts (such as police findings) are presented and framed, can indeed shape the public opinion. Further, it has been found that the arguments of politicians, institutions and pressure groups are presented on a regular basis, in disproportion to the sporadic presentation of opinions from ordinary British citizens or migrants. In this study, we also take a contrast analysis approach for the coverage of specific EU-related topics originating from multiple sources in different countries.

All these findings show the need for further analysis of the media agenda in general.

### 2.2 Quantitative studies

#### 2.2.1 *The bias perspective*

In recent years, a large number of quantitative studies addressed the behavior of the media from a ‘bias’ perspective. As discussed in [18], researchers are facing two main challenges in the empirical study of biased news, namely the problem of the unobserved population and the subjectivity

1. Topics and categories are interchangeably used in this paper.

problem. Recent studies [19], [20], [21], [22] tried to estimate the ideological score for several major news outlets. Keywords present in the news titles have been used in [23] to measure the relative bias between four leading newspapers in Germany. In [24], the authors tried to identify sentiment expressed across various topics to measure the leaning of outlets with respect to political, social, and economic issues. In this study, we also measure polarity of EU-related news but on a different level of aggregation. Exposing narrative discrepancies in EU-related news over different regions could help in a later analysis of the reported EU image subjectivity [?]. Our analysis of the perception is also limited by the multiple unobservable factors external to the news media that influence the final image as sensed by each news consumer. However, previous work on local media systems suggests that the average reader’s view of the social environment (e.g., social and economic issues) aligns with the discussion range in its own media horizon [24]. From this perspective, the problem we are addressing is a required prior step that aims at identifying when the internal and external perceptions of the EU differ in a significant way, which could lead to conflicting images.

Recently, different forms of media bias have been reported [25], for example, ‘event selection [26], [27], [28]’, ‘source selection [29], [30]’, and ‘labeling and word choice [31], [32]’. All such models aim at identifying and predicting different forms of media bias in the news production process. However, our task focuses on the analysis of the post-selection phase of news outlets. In this paper, we measure the coverage and polarity variation of EU-related news in different geographic zones. Our work mainly differentiates from the above-mentioned bias analysis in that we are concerned with checking the possible resulting impact rather than the behavior of the media.

### 2.2.2 Setting the agenda through mass media

A complementary line of research to bias analysis is the study of opinion-shaping or agenda-setting [3] through mass media. In [33] a model capable of imposing the view of a news outlet to a large number of consumers is analysed. Recently, the difference between objective and perceived bias has been pointed out [34]. Here, we focus only on the objective analysis of the media (i.e., “coverage bias”). However, the role of perceived bias in the communication of the EU image is an important aspect that we will pursue in future work.

### 2.2.3 Computational approaches

In recent times, a few computational approaches have been proposed to analyze the selection of political speeches, conventions, media highlights, propagation pattern of different news, effect of bias in opinion shaping and other important subjects [35], [36], [37], [38], [39]. Notably, most of the quantitative analysis for EU-related news focus on one or two topics at a time, and only consider European media. To the best of our knowledge, none of the prior studies focused on a detailed large-scale analysis of the external image of the EU in regional media across the globe. This paper performs a detailed computational analysis of EU-related news and reports some of its potential implications.

## 3 DATASET

In this section, we take a closer look at the dataset used in our study. A comprehensive and diverse dataset covering a large number of countries is required for our analysis. For this purpose, we start from an extensive collection of almost 200.000 news outlets compiled by GDELT Technology<sup>2</sup>. This collection contains outlets from all continents and provides one of the broadest samples of the global news media landscape available.

To map news sources to their host countries, GDELT relies on the strong geographic bias ingrained in most news institutions’ editorial policies. News outlets work on an economy of scale with a substantial first copy cost. Thus, outlets will give priority to stories where their reporters can get quick and easy access (again, to minimize the cost of the piece of news). According to Zipf’s Gravity Model [40], [41], the interest of a piece of news decreases as we move further away from the source of the event. This behavior has also been observed for online media [42]. Hence, outlets will be most probably located in their “primary” country of focus.

As mentioned in Section 1, the primary objective of this work is to capture the variations in the representation of the EU in different regions of the world. Hence, we extract information about all the events which happened in any of the 27 constituent EU countries during the year 2018. The following steps are executed to collect our EU related dataset from GDELT.

- 1) To collect information about EU-centric events, we focus on three key fields of the **Event** database of GDELT: *Actor1Geo\_CountryCode*, *Actor2Geo\_CountryCode*, *ActionGeo\_CountryCode*. If any of these fields contains one of the 27 EU countries, the news event is marked as an EU related event. In this way, we are able to collect news about all the EU countries from articles from all over the globe.
- 2) The **Event** database contains only the first mention of an event. All the follow-up mentions of the events are present in the **Mention** database. Hence, we extract all the mentions of an event within a lookahead period of three months by querying this **Mention** database based on the *GlobalEventID* field. We chose a three-months window because the number of mentions of an event typically almost drops to zero after this period. From the retrieved entries, we get the article identifiers of the documents containing the mentions of the desired event.
- 3) Finally, using the document identifiers collected in the previous step, we query the **Knowledge Graph** database (GKG). For each document, we extract all its metadata, such as themes, organizations, locations, and other content analysis measures. After this step, we have obtained per EU-country events related information and their coverage in the media.

As mentioned earlier, this dataset also contains information about host countries of different news channels. For example, ‘zznews.cn’ and ‘edgehospitality.ca’ are hosted in China and Canada respectively. The GDELT repository contains host country mappings for around 190k source URLs.

2. <https://www.gdeltproject.org/>

Zone	# Sub-regions	# Countries	# EN sources	# non-EN sources
EU	1	27	3822	7333
EUROPE(NON-EU)	5	26	346	1503
ASIA(AS)	5	35	1848	798
AFRICA(AF)	5	60	955	635
MIDDLE-EAST(ME)	1	14	558	892
AMERICA(AM)	4	57	14223	2640
OCEANIA(OC)	4	29	892	36

TABLE 1

Data statistics of news sources over different geographic regions.

# Distinct Events	# News	Zones						
		EU	EUROPE(NON-EU)	AS	AF	ME	AM	OC
13.24	30.18	11.59	2.19	1.69	0.89	1.45	11.7	0.62

TABLE 2

Data statistics(x1M) of news collection of EU events over different geographic regions from Jan-Dec 2018.

In general, GDEL T has defined an affinity-first approach that maps news outlets to their host country based first on their top-level domain in the Domain Name System of the Internet (e.g., news sources with `.at` domain are assigned to **Austria**), then to their primary country of focus, and finally to the country where they are incorporated, or the entity that controls their domain is registered. The motivation behind this country-based mapping of news channels is two-fold. First, we perform the whole study under the assumption that news channels hosted in a country influence the perspectives in that country towards local and global (worldwide) issues. In other words, European issues presented by a news channel in a country will probably influence the perspectives of that country’s audience about Europe. Second, this strategy allows to aggregate news channels at different geographic levels based on their hosting countries (e.g., German or European). We get the host country information of all the news articles collected in the previous steps.

These host countries are mapped to one of the following six geographic regions: (i). EU, (ii). EUROPE(NON-EU), (iii). ASIA(AS), (iv). AFRICA(AF), (v). MIDDLE-EAST(ME), (vi). AMERICA(AM), and (vii). OCEANIA(OC). Some regions are further divided into sub-regions (e.g., AFRICA is divided into north, east, and south sub-regions). The region-wise country distribution is obtained from the United Nations database<sup>3</sup>. Antarctica is dropped from this list due to unavailability of any news outlets. In Table 1 we show the distribution of sub-regions, countries, English and non-English news sources for each region present in our dataset<sup>4</sup>.

Overall, we were able to gather 30M news documents over the period Jan-Dec 2018. This final dataset of news constitutes our corpus for all further analysis presented in this paper. Table 2 presents the detailed statistics of the dataset. For each document, GDEL T provides further annotations of sentiment-related attributes such as *tone*, *positive score*, *negative score*, *polarity*. These scores are identified by the Global Content Analysis Measures (GCAM) system. Each document is also associated with categorical themes (e.g., TAX\_FNCACT, HUMAN\_TRAFFICKING, HEALTH\_VACCINATION). The system recognizes 284 general themes. Apart from these general themes, the

current GDEL T system also identifies several specific themes (e.g., TAX\_FNCACT\_CARTEL is a special case of TAX\_FNCACT). Altogether the system identifies a total of 56,840 themes combining general and specific ones<sup>5</sup>.

## 4 METHOD

In this section we elaborate on the method employed in our analysis. Since we are interested in studying how the mass media portray Europe (EU) from outside of its borders, our global strategy is to compare the coverage of European issues across different geographic zones. We use the internal European coverage as a base-line, assuming it will constitute the origin for most of the reporting. To some degree, this should also show how the image that Europe tries to present of itself morphs based on different geopolitical interests.

It is worth noticing that GDEL T provides sentiment annotation for documents in multiple languages. This is very important as it allows us to consider the EU representation on both international and local media. According to GDEL T, Non-English language documents are automatically translated into English and subsequently processed. Previous research shows that sentiment analysis of English translations of texts in a resource-poor source language produces competitive results, w.r.t. native sentiment analysis [43]. Also, previous studies have effectively conducted content-analysis of cross-linguistic mass media supported by an automatic translation into English [44]. Here, we make the practical assumption that the translated news analysis does not significantly alter the corpus’ sentiment distribution.

We collected event-centric data for each of the EU countries for the period of Jan-Dec 2018 (for details see Section 3). Although the individual countries’ image of Europe is of high interest, we will further restrict our analysis to bigger geographic areas. Previous works have shown that neighboring countries and those sharing strong cultural and economic ties will cover issues more similarly [44]. Here, we check the image of the EU across seven different geographic zones: (i). EU, (ii). EUROPE(NON-EU), (iii). AFRICA, (iv). ASIA, (v). MIDDLE-EAST, (vi). AMERICA, and (vii). OCEANIA (see Table 1 for more details). However, AFRICA, ASIA, and AMERICA cover large geographical areas, consisting of several sub-regions. Each of these sub-regions might follow different editorial policies for their regional audiences. Hence, we also explore the representation of the EU in these smaller zones. Side by side, differences might exist in the representation of individual EU-countries across the globe; hence, we also extend our analysis to individual EU members as targets of the media coverage. In this case, the events are considered separately for each of the EU countries.

As mentioned in Section 3, the documents are categorized into 284 general themes by the GDEL T system. We further annotate these themes into 11 broad wiki topics as proposed in [45]. For this mapping, two independent coders (two of the authors) annotated each of the GDEL T themes as belonging exclusively to one of the 11 topics. In a second stage, coding disagreements were solved through a negotiation among coders in order to improve inter-rater reliability.

3. <https://unstats.un.org/unsd/methodology/m49/>

4. Note that this is not the distribution of the entire collection of GDEL T. This distribution ranges over the sources present in our dataset.

5. <http://data.gdel tproject.org/api/v2/guides/LOOKUP-GKGTHEMES.TXT>

Topic	#Themes	Sample GDELT themes
business and economy	49	'Agriculture', 'Econ_Bankruptcy', 'Econ_Debt'
health and medicine	11	'Disability', 'Health_Pandemonic', 'Health_Vaccination'
international relations	23	'Borders', 'Refugees', 'Seize'
arts and culture	5	'Media_MSM', 'Media_Social', 'Religion'
armed conflicts	11	'Armedconflict', 'Blockade', 'Ceasefire'
law and crime	51	'Arrest', 'Corruption', 'Confiscation'
disasters and accidents	17	'Aviation_Incident', 'Manmade_Disaster', 'Evacuation'
politics and elections	56	'Constitutional', 'Democracy', 'Alliance'
science and technology	3	'Education', 'Internet_Blackout', 'Science'
environment	19	'Env_Biofuel', 'Env_Hydro', 'Env_Mining'
others	38	'Appointment', 'Checkpoint', 'Delay'

TABLE 3  
Topical distribution of GDELT themes.

The final mapping was established with high agreement, Cohen’s Kappa  $\kappa=.74$ . These high level wiki topics and their mapping to the GDELT ground level categorical themes are presented in Table 3.

A news item is mapped to the most relevant topic. For example, news about ‘*Imprisonment order of a French woman issued by Iraq court due to IS membership*’ is part of the ‘**armed conflicts**’ topic. Since we only map categorical GDELT themes to their corresponding topic, we map specific sub-themes annotations in the documents to the matching high-level categorical theme (contained as a substring) and then to the topic. Some of the newly introduced themes such as ‘WB\_2167\_PANDEMICS’, ‘UNGP\_FORESTS\_RIVERS\_OCEANS’ do not belong to any categorical theme; hence, such themes are not mapped to any topic. We assign such themes and corresponding news events to the ‘missing’ category.

The algorithm for news themes to topic mapping is presented below:

- 1) A news item is associated with several themes marked by GDELT. For example, the news ‘Austrian cafe brings authentic treats’ contains themes such as ‘TAX\_ETHNICITY\_AUSTRIAN’, ‘IMMIGRATION’, ‘TAX\_FNCACT\_CHEF’, etc. The same theme may present more than once because different parts of a news item may be associated with that theme.
- 2) We count the frequency of occurrence of different themes and consider the top theme associated with the news.
- 3) Finally, the most frequent theme is mapped to the broad topic. Some of the themes are not present within the 284 annotated themes because they may represent specialized versions of the annotated ones (generic in nature). For example, ‘TAX\_ETHNICITY\_AUSTRIAN’ is a specialized case of ‘TAX\_ETHNICITY’. If we don’t find a direct match of the top theme, we check whether it contains any of the 284 annotated themes as its substring. If we still don’t find a match, it is marked as ‘missing’. For example, the story <http://www.dailydemocrat.com/lifestyle/20180102/austrian-cafe-brings-authentic-treats> is associated with topic ‘*arts\_and\_culture*’.

Another important aspect of our analysis is the overlap in the selection of events by different regions. As mentioned in Section 1, the same event may be reported in different news outlets of multiple regions. For example, the news about ‘**resignation of Brexit secretary David Davis**’ was

published by *India TV News* and *vnewsbd.com* articles of ASIA. We will define these news stories that cover the same event  $i$  as  $E^i = \{e_1^i, \dots, e_n^i\}$ . In general, not all the news reported in other zones are also published in the EU and vice-versa. Interestingly, other regions tend to cover many EU-related events that are absent in the media outlets of the EU. That is why we define the following two categories of EU-related news based on their presence on EU-media:

- 1) **EU-TOTAL**: All the EU-related news (i.e., events linked to any of the 27 EU countries) across all the media outlets in our dataset regardless of the origin.
- 2) **EU-COVERED**: This is a subset of EU-TOTAL where *the events must be covered by at least one of the media outlets of the EU region, and the source country of the news document must be in another zone.*

$$\text{EU-COVERED} = \left\{ e_j^i \mid \exists k \neq j, e_k^i \in E^i, e_j^i \in E^i, \text{host}(e_k^i) \in \text{EU}, \text{host}(e_j^i) \notin \text{EU} \right\} \quad (1)$$

For EU-COVERED, we consider events covered by media outlets of the EU-zone and select only those event-related news from media outlets of other zones. Table 4 (columns (T) and (C)) provides the detailed topic-wise statistics about EU-TOTAL and EU-COVERED news respectively. As expected, a majority of the European news gets filtered and does not appear in the news outlets of other zones (see ‘(C)overed’ columns in Table 4). In their role as gatekeepers, journalists, editors, and other involved parties have to decide and pick what to cover from a massive pool of stories. Their selection is constrained by a combination of organizational factors, news norms, and audience interests [46], [47]. On the other side, what they do select to report about could be very telling of their editorial strategy and underlying forces shaping their framing of real-world events [47]. From the **Total** row in Table 4, we see that the proportion of EU-COVERED to EU-TOTAL for each region varies from 17% (EUROPE(NON-EU) and MIDDLE-EAST) to 29% (AFRICA and AMERICA). However, the ratios remain quite consistent, even across the topics. This might be due to multiple factors such as the different editorial policies followed by the news outlets of different areas based on their target audiences.

In the following, we perform a detailed analysis of EU-TOTAL and EU-COVERED news across different regions and sub-regions to get an initial understanding about the variation in the representation of the EU in different parts of the globe. We discuss our observations and some of the important straightforward implications of this study in the next section.

## 5 ANALYZING THE NEWS

In this section, we perform a detailed category-wise analysis of the European news portrayed in different regions and point out the differences in their representation.

### 5.1 Polarity differences in EU-TOTAL and EU-COVERED news coverage

As mentioned before, we intend to examine the way the EU is represented in other geographic areas. To this end, we

Topic	EU		EUROPE(NON-EU)		ASIA		AFRICA		MIDDLE-EAST		AMERICA		OCEANIA	
	T	C	T	C	T	C	T	C	T	C	T	C	T	C
business and economy	376	-	140	139	575	121	27.5	6.57	29.1	6.07	243	72.3	11.1	2.69
health and medicine	277	-	14.9	3.44	16.7	4.07	12.7	2.87	20.5	2.33	157	46.7	10.1	2.86
international relations	199	-	24.1	4.98	43.5	13.1	11.6	4.55	35.1	7.07	243	85	9.14	2.51
arts and culture	2019	-	543	99.7	411	99.4	218	72.5	543	73.3	2473	808	128	30.6
armed conflicts	169	-	23.2	4.83	29.9	8.49	11.9	4.10	25.1	6.05	319	97.4	29.3	4.41
law and crime	971	-	147	31.1	116	36.7	63.7	21.4	91.2	22.8	1014	384	56	17.9
disasters and accidents	375	-	39.8	10.5	49.6	16.1	30.3	13.3	51.4	11.3	505	148	21.4	5.71
politics and elections	2079	-	529	96.9	296	74	151	44	210	43.7	1947	657	77.3	23.5
science and technology	218	-	13	25.8	18.1	2.98	11.1	1.99	16.3	2.01	208	49.5	10.4	1.99
environment	88.5	-	9.53	2.58	18.5	4.45	8.15	2.98	7.08	1.48	128	46.3	7.83	2.67
other	2599	-	438	80.8	297	76.9	168	48.2	223	42.9	2255	701	141	37.7
<b>Total</b>	<b>9374</b>	<b>-</b>	<b>1924</b>	<b>351</b>	<b>1357</b>	<b>348</b>	<b>715</b>	<b>223</b>	<b>1253</b>	<b>219</b>	<b>9495</b>	<b>3097</b>	<b>502</b>	<b>133</b>

TABLE 4

Data statistics(x1K) of documents per topic over different geographic regions. For each topic we list (T)total values (EU-TOTAL), and number of news from other zones that were also (C)overed in the media of EU(EU-COVERED).

consider the whole set of EU news (Table 4) from different regions, and we inspect the variation in their sentiment polarity distribution. We conduct a comparative analysis using the reported view of the EU in the (EU-)internal media as the baseline (see Section 4).

We observe that the overall characteristic tone (not only EU-related but all news) of different regions are quite different as this might depend on the culture and language of that region. For example, the average reporting tone of MIDDLE-EAST(-1.21) is more negative than AMERICA(-0.80), which is still more negative than ASIA(-0.59). To account for a possible bias introduced by regional idiosyncrasies, we deduct this average sentiment of a region from the sentiment of each of the news articles [2]. This will yield a relative sentiment that represents how each EU-related piece of news deviates from the average tone of the source country. With this, we obtain a **normalized sentiment score** for each of the articles of the different regions. Throughout this paper, we work on this normalized sentiment score.

First, we compare the average sentiment (tone) of the reported news from different regions towards each of the EU member states aggregated over the year 2018. Figure 1 and Figure 2 represent the difference in the sentiment score of other regions compared to the EU for the EU-TOTAL and EU-COVERED datasets respectively. By using the average sentiment of the coverage from the EU as our baseline, we can evaluate to what extent the European view of different members of the EU might change when reported in other geographic regions. For example, we observe that the ‘image’ that is presented in the media from outside the EU about Italy (IT) or Slovakia (LO) is, on average, more unfavorable when compared to the EU’s coverage of the same country.

To identify significant bias toward some EU countries within a region and to have a fair comparison among scores from different regions for the same EU country, we compute the standard score ( $z - score$ ). We calculate the z-scores using the mean and standard deviation of the relative sentiment within each region (e.g., in Figure 1, for the brown series, only average opinions towards the EU-countries in American media are considered). This should help to further remove any inter-regional differences in baseline attention towards the EU. Figures 1 and 2 show the z-score representing the average sentiment towards each EU-member on

the corresponding region’s mean for EU-TOTAL and EU-COVERED respectively. To illustrate, let’s examine the case of Finland (FI) for EU-TOTAL news. Although this country receives a more positive (or less negative) coverage on average from every other zone compared to the EU’s, the graph shows that only ASIA gives significantly favorable coverage to Finland even when taking into account this regions’ standards. In other words, Finland’s relative coverage in ASIA deviates significantly (in favor) from the normal coverage that this region gives to EU countries. Indeed, these results reflect the good reputation of Finland in Asia, where the country is perceived among the world’s happiest countries, with a world-leading education system, trend setting in the technology sector (e.g., Nokia), and one of the most popular travel destinations for Asians. Another interesting example is the over-positive coverage from MIDDLE-EAST of Portugal (PO), which could be a reflection of the welcoming attitude in Portugal (at all levels) towards middle-eastern refugees. For the EU-COVERED analysis (Figure 2), if we go back to the Finland example, we can see that although Asia’s coverage deviates significantly in the positive direction from the internal EU coverage, so does the coverage of several other regions. This “agreement” in the external representation might indicate that the internal EU coverage was maybe too harsh in its reporting about Finland.

For clarity in the figures, we are only showing in the graph EU countries for which at least one region is more than one standard deviation above or below the region’s mean (i.e.,  $|z - score| > 1$ ).

To complement our analysis, we also measure the distribution of polarity values for all EU countries from each of the regions. This will provide an intuition about the viewpoint of the different regions towards the EU. Figure 3 represents the cumulative distribution frequency (CDF) of polarity values of the EU countries from each of the six geographic regions. Trends shown in Figure 3 can be analyzed in detail by also observing the region-wise series in Figures 1 and 2. Regions like AMERICA or OCEANIA( $\geq 98\%$ ) present most of the EU countries in a relatively negative way (intercept of regions’ series with  $X=0$ ). Meanwhile, AFRICA and MIDDLE-EAST represent more than 50% and 40% of the EU countries respectively in a more positive way (including countries such as Bulgaria (BU), Finland (FI),

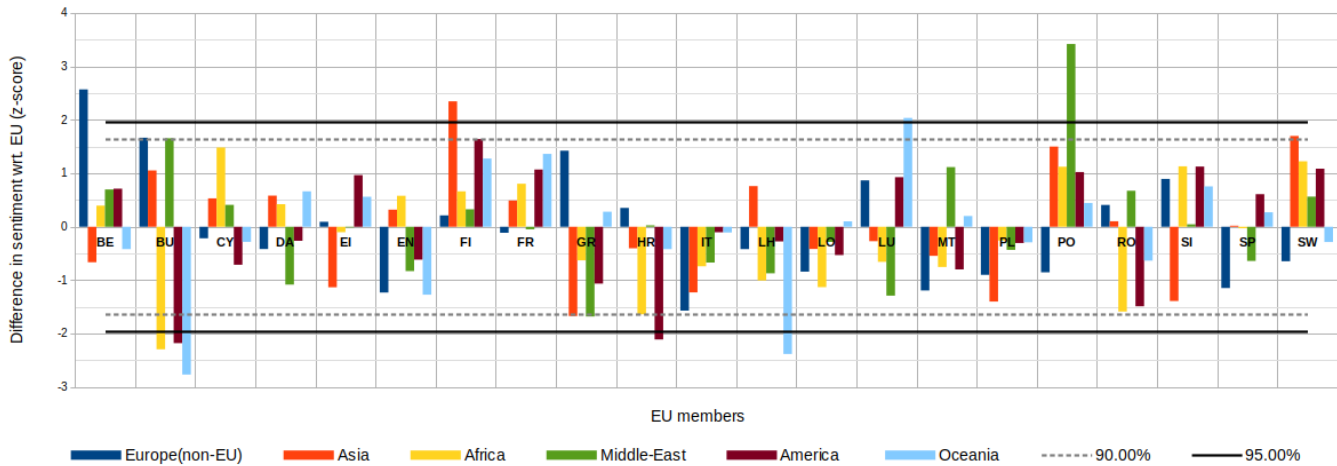


Fig. 1. Divergence of each region's average tone over news in EU-TOTAL w.r.t. the EU coverage. News are aggregated for EU member states.

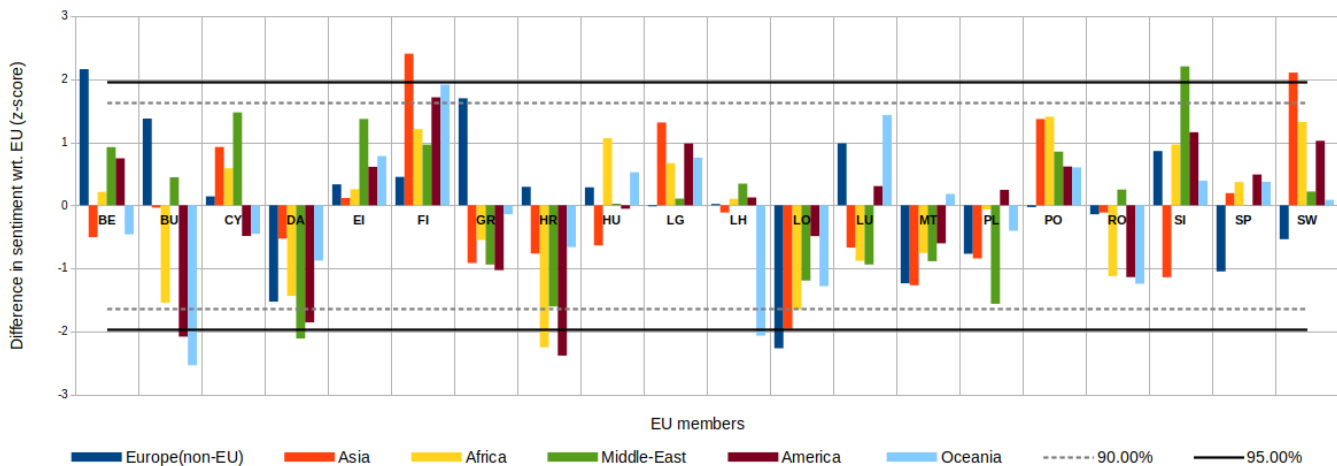


Fig. 2. Divergence of each region's average tone over news in EU-COVERED w.r.t. the EU coverage. News are aggregated for EU member states.

Malta (MT), Portugal (PO), and Sweden (SW)). Moreover, those countries covered with a comparatively more negative press in Middle-East are closer to the average sentiment used in the EU than in any other region.

It is interesting to note that European countries (non-EU members) tend to present EU members from a more negative point of view compared to the EU. Although, it should be noted that this region tends to stay relatively close to the baseline (EU coverage) in both the negative and positive direction. In an attempt to explain this behavior, let us first divide the non-EU Countries into former western and eastern block states. From a social and economic point of view, the former western block countries are closer in nature and standard to the EU (e.g., Norway and Switzerland). These countries are typically in a tight relationship with the EU. Without clear benefits of possible EU membership, it is reasonable to prefer and defend the status quo also by depicting a more negative image of the EU. The former eastern block countries, on the other hand, still encounter tremendous economic difficulties and geopolitical interests and influence of third parties, most notably of Russia (e.g., in Armenia, Georgia, Moldova, and Serbia) and the US (in

countries such as Albania, Bosnia, Macedonia, Montenegro, and Kosovo)[48], [49], [50]. Country-Country conflicts also cannot be excluded, especially when it comes to states established after the break-up of Yugoslavia and the war on the West Balkan (e.g., Serbia and its tensions with Croatia, or Macedonia with its dispute with Greece about its name, or the claims about history against Bulgaria). Last but not least, the negative opinion of some EU members on non-EU European countries, may lead to the reciprocal negative presentation of the whole European Union in the local media (e.g., Turkey or Ukraine).

For EU-COVERED, EUROPE(NON-EU), MIDDLE-EAST, AMERICA, and OCEANIA follow a quite similar trend to EU-TOTAL news. On the other hand, there is a difference in the pattern for ASIA and AFRICA. This regions drift away from MIDDLE-EAST and EUROPE(NON-EU), and get closer to the more negative tone of AMERICA. For some countries (e.g., 'Denmark (DA)') they even took a reverse standpoint from EU-TOTAL.

Finally, Figure 3 reveals the differences in the representation of EU countries over different regions for EU-TOTAL and EU-COVERED news. Although, most regions present to

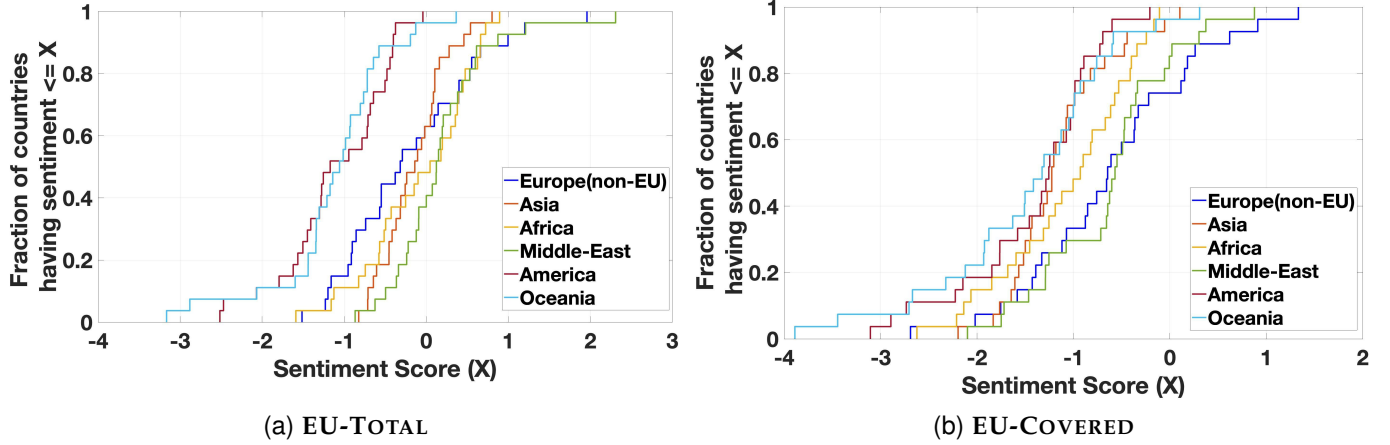


Fig. 3. CDF of normalized sentiments (relative to EU) of EU countries in regional media for EU-TOTAL and EU-COVERED news events.

their audience a relatively negative view of EU-TOTAL for a large percentage of the EU, EU-COVERED gives an even more negative picture compared to their EU-TOTAL versions. The differences are quite clear for ASIA and AFRICA. AFRICA presents 50% of the EU countries with sentiment score  $< 0$  for the EU-TOTAL news whereas all the countries are covered under negative scores for EU-COVERED.

This leads to another interesting question: “Do different regions favor EU countries in the same way?” We rank the EU countries based on their sentiment scores for each of the regions and measure the Spearman rank correlation between each pair of regions. The pairs  $\langle \text{EU-ASIA} \rangle$ ,  $\langle \text{EU-MIDDLE-EAST} \rangle$ ,  $\langle \text{MIDDLE-EAST-ASIA} \rangle$ ,  $\langle \text{AFRICA-AMERICA} \rangle$ ,  $\langle \text{AFRICA-OCEANIA} \rangle$  and  $\langle \text{AMERICA-OCEANIA} \rangle$  show strong positive correlation ( $\rho(27) > .60, p < .001$ ) [51]. This indicates that there is affinity in how these regions promote individual EU-countries. For EU-COVERED, we have observed some new pairs that reveal significant correlation such as  $\langle \text{EU-EUROPE(NON-EU)} \rangle$ ,  $\langle \text{EUROPE(NON-EU)-MIDDLE-EAST} \rangle$ ,  $\langle \text{ASIA-AFRICA} \rangle$ ,  $\langle \text{ASIA-AMERICA} \rangle$ ,  $\langle \text{AFRICA-MIDDLE-EAST} \rangle$ ,  $\langle \text{AMERICA-MIDDLE-EAST} \rangle$ . Overall, EU-COVERED reveals more correlation between different regions than EU-TOTAL, showing a stronger agreement in the newsworthiness and coverage of this set of events. The ranked internal coverage of the EU shows a relatively weaker correlation with all other regions. We hypothesize that each region prioritizes the coverage of some countries, or portrays them in a more or less positive light, based on their political and economic objectives [52]. Our topic-wise analysis in Section 5.3 further supports this idea.

So far, we have analyzed the average sentiment values towards each of the EU countries on other regions. This region-wise analysis is performed over all the news reported in Table 2. We perform a two-sample Welch’s t-test [53] to check the significance in the difference between regions in their calculated sentiment for the EU news coverage. For the EU-TOTAL and EU-COVERED dataset, most other regions’ coverage of EU is significantly more negative ( $\alpha = .001$ ) than the internal EU news.

Regions	Sub-regions	EU-TOTAL	EU-COVERED
AFRICA	EAF	-1.23	-2.19
	MAF	-1.06	-1.73
	NAF	-0.05	-0.72
	SAF	-1.40	-2.25
	WAF	-0.99	-2.12
ASIA	EAS	-0.42	-1.44
	CAS	0.06	-0.70
	WAS	0.04	-0.94
	SAS	-0.95	-1.70
	SEAS	-2.23	-2.86
AMERICA	CBM	-1.10	-1.73
	CAM	-0.91	-1.06
	SAM	-0.73	-0.84
	NAM	-1.69	-2.06

TABLE 5  
Average normalized sentiment score in different sub-regions of AFRICA, ASIA, and AMERICA for EU-TOTAL and EU-COVERED news.

## 5.2 Polarity differences in sub-regions

Figure. 1 and 2 demonstrate the differences in the tone of EU-TOTAL and EU-COVERED news in different regions of the world. However, we acknowledge that some of the regions are quite large and their cultures and news consumption habits also quite diverse. Hence, a single aggregated view might not clearly represent the editorial policy of that region. This especially holds true for AFRICA, ASIA, and AMERICA that constitute large geographical areas. These regions are further divided into smaller sub-regions and news representation policies of each of these sub-regions are inspected separately in this section. AFRICA is divided into five sub-regions: (i). EAST-AFRICA, (ii). MIDDLE-AFRICA, (iii). NORTH-AFRICA, (iv). SOUTH-AFRICA, and (v). WEST-AFRICA. ASIA also consists of the following five regions: (i). SOUTH-ASIA, (ii). WEST-ASIA, (iii). EAST-ASIA, (iv). CENTRAL-ASIA, (v). SOUTH-EAST-ASIA. Similarly, AMERICA contains the following four sub-regions: (i). CARIBBEAN-AMERICA, (ii). CENTRAL-AMERICA, (iii). SOUTH-AMERICA, (iv). NORTH-AMERICA.

Table 5 shows the average relative sentiment value of different sub-regions of AMERICA, AFRICA and ASIA towards the EU for both EU-TOTAL and EU-COVERED news. The



Topic	ER		AS		AF		ME		AM		OC	
	T	C	T	C	T	C	T	C	T	C	T	C
<b>business and economy</b>	1.31	0.53	-0.62	-1.30	0.15	-0.98	0.14	-0.54	-0.81	-1.11	-0.96	-1.19
<b>health and medicine</b>	-1.08	-1.27	-0.69	-1.31	0.54	-0.60	-0.18	-0.29	-0.54	-0.80	-0.84	-0.51
<b>international relations</b>	-1.28	-1.46	-2.06	-2.34	-1.45	-1.79	-1.35	-1.70	-1.72	-2.03	-2.12	-2.08
<b>arts and culture</b>	0.34	-0.19	-0.30	-1.36	-0.43	-1.31	-0.07	-0.78	-1.42	-1.82	-1.03	-1.52
<b>armed conflicts</b>	-0.71	-0.79	-0.50	-0.59	-1.07	-1.46	-0.16	-0.38	-0.69	-0.89	-0.14	0.01
<b>law and crime</b>	0.11	-0.52	-0.87	-1.60	-0.30	-1.13	-0.23	-1.02	-0.92	-1.47	-0.97	-1.43
<b>disasters and accidents</b>	-0.78	-0.80	-0.79	-1.22	-0.08	-0.47	0.43	-0.39	-0.68	-0.76	-0.76	-1.27
<b>politics and elections</b>	0.93	0.59	-0.17	-0.97	-0.05	-1.03	-0.24	-0.55	-0.79	-0.88	-1.19	-1.11
<b>science and technology</b>	-1.09	-1.06	0.44	-0.39	0.79	-0.05	0.85	0.80	-0.76	-0.92	-0.92	-1.63
<b>environment</b>	-1.57	-1.75	-0.16	-1.26	-0.30	-1.49	-0.33	-1.05	-0.99	-1.50	-1.12	-2.14
<b>others</b>	0.48	-0.10	-0.13	-1.02	0.25	-0.77	0.44	-0.42	-0.78	-1.21	-0.72	-1.14
<b>Average</b>	-0.30	-0.62	-0.53	-1.21	-0.18	-1.01	-0.06	-0.57	-0.92	-1.22	-0.98	-1.27

TABLE 6

Topic-wise normalized sentiment polarity representation of EU-TOTAL(T) and EU-COVERED(C) events among different zones relative to EU.

behavior of some sub-regions differs among themselves, with both EU-TOTAL and EU-COVERED news having a similar pattern. For example, NORTH-AFRICA tends to provide a less negative view of EU to their audiences compared to other sub-regions of AFRICA. Similarly, CENTRAL-ASIA and WEST-ASIA provide a more positive reference of EU than the south-east part. EAST-ASIA and SOUTH-ASIA hold a less negative view than SOUTH-EAST-ASIA. Interestingly, these regions (NORTH-AFRICA, CENTRAL-ASIA, WEST-ASIA) gradually appear to be a major source of migrants to the EU according to UNHCR statistics [8]. However, a more detailed analysis is required to understand the reason behind this phenomenon. In the case of AMERICA, NORTH-AMERICA presents EU-related news on average in a more negative tone than the other sub-regions of AMERICA. These two regions (NORTH-AMERICA and EU) are important global actors in different aspects such as trade, health-services, education, etc. A possible reason for advancing a rather negative image of the EU zone might be due to such competing interests in the political and economic fields.

Evidence suggests that different regions follow different policies about their EU coverage. As we move to a finer granularity in the geographic aggregation such as states or cities, we also expect to find different patterns. Still, some similarities exist among them to justify their joint analysis (e.g., language, culture, and values). A more disaggregated analysis could be interesting and beneficial. However, we leave this for a future analysis.

### 5.3 Differences in news polarity across topics

In Section 5.1, we observed that the presentation of the EU related news varies significantly between different geographical regions. On average, all the regions have a tendency to portray EU-related events in a negative direction. This trend is reflected more precisely for the EU-COVERED news than the EU-TOTAL. Section 5.2 shows the variation among different sub-regions of AMERICA, AFRICA and ASIA. We observed that some of the sub-regions portray the EU in a more positive (or less negative) direction than others. All these analyses were carried out over the whole set of news without considering their themes/topics. However, sentiment of a piece of news is highly related to the

topic and this has great influence in the representation of a news article [17], [52], [54], [55]. Hence, in this section, we inspect the impact of topics on the news sentiment across different regions and sub-regions.

#### 5.3.1 Topic-wise sentiment distribution of EU-TOTAL and EU-COVERED news

Table 6 reports the topic-wise difference in the EU-TOTAL (T) and EU-COVERED (C) news sentiment polarity between each region and the EU. Similar to the previous analysis, the average sentiment of a region is deducted from the corresponding region's topic-wise sentiments. Different regions follow a quite different trend in terms of EU-news coverage. EUROPE(NON-EU) presents topics such as 'business and economy', 'arts and culture', 'politics and elections', and others in a more positive way than the EU. For example, 'Discussion about web tax' is covered with more positive sentiment in EUROPE(NON-EU) than inside the EU. ASIA only presents 'science and technology' with a positive light but 'international relations' appears with a strongly negative bias. AFRICA covers economic, health, and scientific aspects in a positive way to their audiences. On the other hand, coverage of AMERICA and OCEANIA is mostly negative. International relations news is here again strongly negative.

We perform a statistical Welch's t-test between the sentiment distribution of EU and each of the other regions for each of the topics. The distributions turn out to be significantly different from EU, which suggests that each of these regions follows different editorial and word selection strategies to present the EU news. It is also shown in Table 6 that the negative representation is highest for the topic 'international relations'.

Another interesting finding in Table 6, similar to Figure 3, is that the more negative sentiment for EU-COVERED compared to EU-TOTAL news is also reflected along all the topics. This reveals that this effect is not driven by a difference in the coverage of some topics, but rather by the EU-related events that are skipped by EU-media (EU-MISSING henceforth). These are covered by most of the other zones in a positive (or at least neutral) way, shifting the total average. A detailed topic-wise investigation of such news shows several interesting patterns:

EUROPE(NON-EU)	ASIA	AFRICA	MIDDLE-EAST	AMERICA	OCEANIA
fax.al	xinhuanet.com	fasopresse.net	english.wafa.ps	reuters.com	nzherald.co.nz
time.mk	business-standard.com	allafrica.com	wafa.ps	msn.com	dailytelegraph.com.au
daily.mk	indiatimes.com	ahram.org.eg	timesofisrael.com	yahoo.com	couriermail.com.au
sputniknews.com	thestar.com.my	unitaangola.org	palestinemonitor.org	iheart.com	theaustralian.com.au
swissinfo.ch	china.org.cn	alufuqnews.com	dailysabah.com	trust.org	heraldsun.com.au

TABLE 7  
Top-5 news channel promoting EU-MISSING news in different regions.

- Most of the missing news events (i.e., EU-MISSING) are from the topics ‘business and economy’, ‘science and technology’. Apart from these, other regions follow quite different trends. For example, in EUROPE(NON-EU), most of the missing events also come from politics. The Spearman correlation between different regions based on the topic-wise coverage (number of hits) of EU-MISSING news is significantly high ( $> 0.80$ ).
- Distribution of sentiment values of EU-MISSING and EU-COVERED news for each region is significantly different as per Welch t-test. As mentioned before, the events not covered by EU-media are represented in a more positive way than their counterparts. Table 8 shows examples of ‘business and economy’ related EU-MISSING news. We manually checked 100 news items from the ‘business’ and ‘culture’ categories. For business, we observe that news are predominantly related to the agenda of the reporting zone (e.g., trade agreements with countries in the EU). Also, some regions (e.g., non-EU) make them promotional news following their potential propaganda interests. Regarding the ‘Arts and culture’ topic, this covers information about themes like main stream media, social media, etc; hence, it also captures information about journalists, media, democracy, interviews with celebrities, etc. In the GDELT dataset, a single news may be composed of multiple events and it depends on the presenters/editors point of view to highlight one of them. For example, the interview of French first lady Brigitte Macron is composed of several event-ids such as ‘her view about Melania Trump’, ‘her viewpoint about French activities’, etc. According to the data from GDELT, her view about Melania Trump is not covered in the EU media. However, it is reported in EUROPE(NON-EU) media.
- Figure 4 presents the distribution of news polarity for the EU-MISSING and EU-COVERED news over different regions for topics ‘Business and Economy’, and ‘Science and Technology’. It confirms that self-reported EU-related news of different regions get more positive scores than the EU-COVERED news. AFRICA and MIDDLE-EAST present these topics in a more positive way than their average representative tone. For EUROPE(NON-EU), the tone of business-related issues is more positive than its baseline, but they are apparently more critical of the EU regarding science and technology. On the other hand, AMERICA follows the reverse trend for these two topics. They criticize business and economy but cover sci-

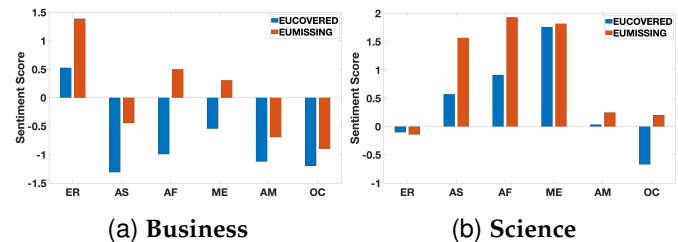


Fig. 4. Mean normalized polarity value of EU-COVERED and EU-MISSING news for business and science across different zone. EU-MISSING news are presented in a positive than the EU-COVERED news.

- entific aspects in a slightly positive direction.
- The distribution of news channels involved in broadcasting EU-MISSING news follows a power-law distribution i.e., some specific news channels from each region are primarily involved in circulating such EU-related news that are not covered in EU-media. This could be an indicator of the subjectivity of the newsworthiness value of the news in EU-MISSING [56]. Table 7 shows the top five such promoting news channels from each of the regions. Some of these outlets (e.g., RT and Sputnik) have been identified in previous studies as sources of misinformation/propaganda campaigns [57], [58]

### 5.3.2 Topic-wise sentiment distribution of EU-TOTAL and EU-COVERED news over sub-regions of AFRICA

Results discussed in Section 5.2 indicate that different sub-regions of AFRICA follow different selection processes for their audiences. The polarity distribution is significantly different across all the sub-regions. Hence, it is interesting to explore the topic-level variance in the polarity representation of different sub-regions of AFRICA. Figure 5 shows the topic-wise variation in polarity over AFRICA sub-regions.

According to EU-TOTAL, NORTH-AFRICA and MIDDLE-AFRICA tend to represent news related to topics such as ‘business and economy’, ‘science and technology’, or ‘law and crime’ in a similar or more positive tone than EU. Such topics usually represent the socioeconomic status of a country and they are frequently used to define its position in comparison to other countries (e.g., Global Social Mobility Index<sup>6</sup>). Particularly, for MIDDLE-AFRICA, topics like ‘law and crime’, ‘disasters and accidents’, and ‘health and medicine’ stand out for their positive bias.

6. [http://www3.weforum.org/docs/Global\\_Social\\_Mobility\\_Report.pdf](http://www3.weforum.org/docs/Global_Social_Mobility_Report.pdf)

Region	News
	Trump pushes deadline for aluminum and steel tariffs
EUROPE(NON-EU)	European commission launches battle for post successful EU multiannual budget
	Workers activists mark May day with defiant rallies
AMERICA	Delay in Trump's decision making on steel tariffs for UK and other allies

TABLE 8

Sample EU-MISSING news from 'business and economy' in regions EUROPE(NON-EU) and AMERICA.

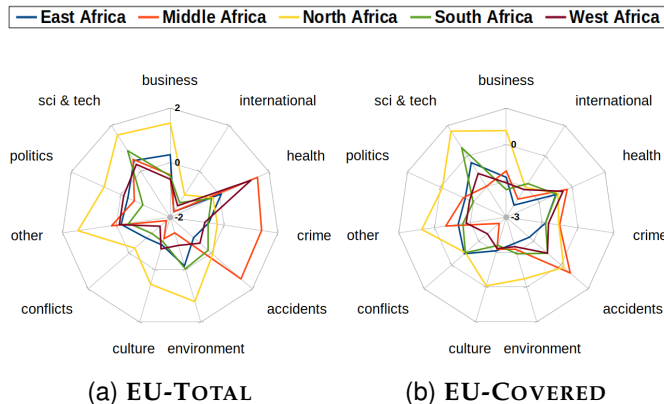


Fig. 5. Topic-wise sentiment polarity representation of (a) EU-TOTAL and (b) EU-COVERED events in sub-regions of AFRICA relative to EU.

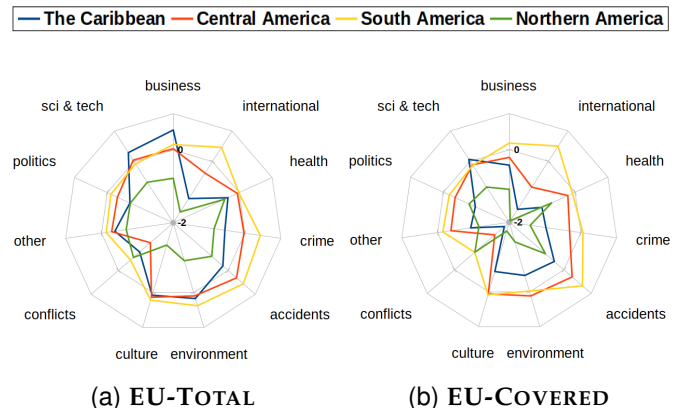


Fig. 7. Topic-wise sentiment polarity representation of (a) EU-TOTAL and (b) EU-COVERED events in sub-regions of AMERICA relative to EU.

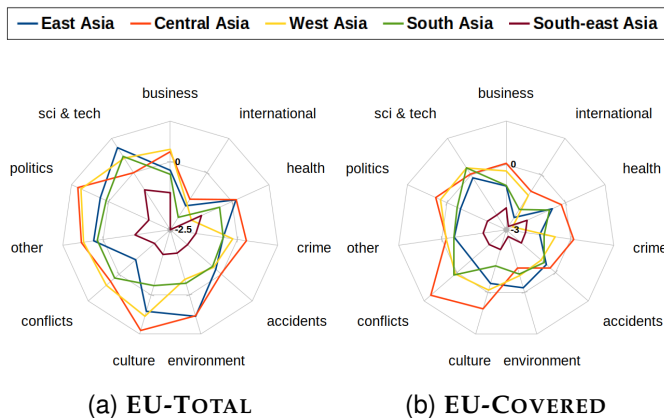


Fig. 6. Topic-wise sentiment polarity representation of (a) EU-TOTAL and (b) EU-COVERED events in sub-regions of ASIA relative to EU.

We also observe that topics like 'environment' and 'health and medicine' are conflicting, with wide differences between sub-regions. For topics such as 'armed conflicts' and 'international relations' seems to be a consensus in the more negative representation when compared to the EU.

Similar to previous findings, here also, tones are relatively negative compared to the baseline zone for the EU-COVERED news. However, audiences of a region will be fed EU-TOTAL news i.e., the news broadcast over the media in the respective regions. Hence, overall, end users of the corresponding regions are expected to get a similar or mildly positive view of the EU region compared to EU-media. For example **Morocco** and **Nigeria** report business agreements with the EU such as 'fisheries agreement' and 'agricultural deals' with sentiment rating of 3 and 4 respectively.

### 5.3.3 Topic-wise sentiment distribution of EU-TOTAL and EU-COVERED news over sub-regions of ASIA

Fig. 6 shows the topic-wise sentiment values for different sub-regions of ASIA. ASIA also reveals similar trend as AFRICA. CENTRAL-ASIA and WEST-ASIA present EU-related news in a relatively positive manner to their audiences. On the other hand, SOUTH-EAST-ASIA seems to have a rather negative view of the EU. Unlike AFRICA, here the pattern for different topics is more homogeneous across sub-regions, following similar trends (except for 'health and medicine'). In general, EU-COVERED news are presented in a relatively negative tone in all the sub-regions except CENTRAL-ASIA. CENTRAL-ASIA (Kazakhstan, Uzbekistan etc.) and WEST-ASIA (Azerbaijan, Armenia etc.) lag other sub-regions in terms of social and economic factors. Also, these two sub-regions report a large number of migration cases over previous couple of years according to UN-HCR statistics [8]. For example, **Kazakhstan** and **Kyrgyzstan** present the cultural and business related news about 'London-based fashion show' and 'information about a new industrial site' respectively with a sentiment score of 2.

### 5.3.4 Topic-wise sentiment distribution of EU-TOTAL and EU-COVERED news over sub-regions of AMERICA

Different sub-regions of AMERICA reveal slightly different patterns than sub-regions of AFRICA and ASIA. In the former cases, all the sub-regions except the ones that are at war, distress, or conflicting state represent several economy-driven topics ('business and economy', 'politics and elections', 'law and crime', etc.) portraying EU in a pessimistic way to their audiences. Here, SOUTH-AMERICA, CENTRAL-AMERICA, and to a lesser extend CARIBBEAN-AMERICA,

report a similar or more positive view for most of the topics in their corresponding media. However, NORTH-AMERICA follows a consistent trend as observed in general cases (Section 5.1 and 5.2). It presents the EU in a less optimistic way. For example, NORTH-AMERICA tried to highlight the deficits of EU in ‘international relations’ that has a direct impact on business, export, and political decisions.

As shown in Fig. 7, relative sentiment of EU-TOTAL news are in general more positive than EU-COVERED ones, except for a couple of cases. In both EU-TOTAL and EU-COVERED, we observe a significant different sentiment distribution between EU and each of the sub-regions as per Welch’s t-test ( $p \leq 0.001$ ). For example, NORTH-AMERICA favors news such as ‘EU’s decision about ‘Iran’s nuclear deal’ and ‘Vienna’s rejection about exemption of senior citizen’s police fee’ which are covered with a negative tone of -3. On the opposite side, CARIBBEAN-AMERICA (such as Barbados and Jamaica) gives more coverage to news like ‘reopening of Sunbury plant’ and ‘Jamaica’s deal with Ireland for potatoes’ with positive polarity of 2.

#### 5.4 Implications of the polarity differences

In this paper, we carried out a thorough analysis of different types of EU-related news (EU-TOTAL, EU-COVERED, and EU-MISSING) and their editorial customization across different media-channels all over the globe.

It is apparent that the external image of the EU is not uniform and significantly different than its home representation. In some of the regions (e.g., CENTRAL-ASIA, NORTH-AFRICA) we find a positive image of the EU whereas in regions like NORTH-AMERICA or SOUTH-AFRICA we encounter a negative view of the EU. On the other hand, a more positive representation of EU-TOTAL news over EU-COVERED ones reveals interesting patterns: (i). EU-media originated news are strongly criticised in the other regional outlets, (ii). regional media cover several news about the EU which are missing in their home-based outlets and further, such news are described in an overly positive manner. Unfortunately, audiences only consume the EU-TOTAL feed of news media and don’t necessarily know about or have access to what is covered in the EU. This might become a source of misperception about the EU among regional audiences. Starting from 2015, UNHCR [8] also observed a heavy trend of migration towards the EU, especially from MIDDLE-EAST, CENTRAL-ASIA, WEST-ASIA, NORTH-AFRICA, and WEST-AFRICA. Biased news consumption might play an important role in such developments and eventually also result in the creation of threats and risks caused by misconceptions of opportunities and requirement. *In this paper, we carried out a comparative analysis of the news representation. However, further study of the perception of the regional audiences is needed to better understand possible causal relations [59]. We will study this aspect in future research. Nevertheless, this paper already highlights the need for the EU to develop dissemination strategies tailored to each region that reinforce the EU’s internal messages and provide continuous feedback on the respective reception and reflection in media. Such strategies could counteract editorial policies which might create a misrepresented image of EU countries if left unchecked.*

## 6 CONCLUDING DISCUSSION

In this paper, we performed a detailed analysis of the **representation** of EU-related news around the world. We conducted a detailed study to check the view of the EU in the media of other geographical regions and their differences with the internal reporting inside the EU. We were able to observe how the EU is covered differently in the news of other areas and found that the sentiment polarity for EU-related topics differs from that of EU-media. Furthermore, only a small subset of the EU-related news is directly adopted. We further extended this analysis over different topics and sub-regions of ASIA, AFRICA, and AMERICA. The results of this study are an important first step towards a better understanding of the **perception** of the EU among audiences across the globe. It will be interesting and beneficial to check the influence on the targeted audiences, which will be part of our next steps. An open interview with the audiences might be helpful in exploiting the correlation between **representation-perception**.

Still, the present study exhibits some limitations and opens interesting lines for future research. We identify some of them here:

- 1) We consider the news posted in both international and local news outlets of different regions. Non-English content is translated into English for the purpose of analysis. However, we did not check the differences in the news coverage between the local and international (usually in English) outlets of a specific region. International outlets are mostly used to reach global audiences (or expats), not the local population. Local audiences usually consume news posted in their native languages. Hence, it is essential to analyze such representation gaps, if any.
- 2) We consider eleven broad categories of topics in this paper. Some of the sub-topics such as ‘Human Trafficking’ or ‘Immigration’ may be of great concern. A detailed and thorough understanding of the representation of fine-grained sub-topics is required to analyze the influence of the media in their perception and in the impact on specific events.
- 3) We explore the differences in representation in the news across different regions. However, it will also be interesting to check the differences in the perceptions generated due to such representations. In this paper, we assume the role of mass media to be critical in forming an image or a perception of a region. However, we fully acknowledge that many other aspects also play a crucial role in personal decisions and certainly in policy making. Side by side, citizens may form their perception/view by collecting information from several different sources. We will analyze this in detail in future research.

Digital availability of news from around the world allows the study of perception at a large and global scale. We have seen how narratives (through selection of events and sentiment polarity) of EU-related issues may change when reported from outside its borders. These narratives and the way the EU is portrayed can reasonably be expected to have an impact on the perceptions and beliefs of the audiences (but they are by no means the only influencing factors). The

rapid detection of topics with diverging perceptions as well as of developments leading to such differences will allow to identify misunderstandings and potential threats caused by misconceptions in an efficient manner. The insights gained through such analysis empower decision-makers and policy-makers to better understand the global context and to act accordingly in a timely manner.

**Acknowledgement:** Funding for this project was in part provided by the European Union's Horizon 2020 research and innovation programme under grant agreement No 832921 (MIRROR).

## REFERENCES

- [1] K. E. Boulding, "National images and international systems," *The Journal of Conflict Resolution*, vol. 3, no. 2, pp. 120–131, 1959. [Online]. Available: <http://www.jstor.org/stable/173107>
- [2] C. M. Koch, I. Moise, D. Helbing, and K. Donnay, "Public debate in the media matters: evidence from the european refugee crisis," *EPJ Data Science*, vol. 9, no. 1, p. 12, 2020.
- [3] M. E. McCombs and D. L. Shaw, "The agenda-setting function of mass media," *The Public Opinion Quarterly*, vol. 36, no. 2, 1972. [Online]. Available: <http://www.jstor.org/stable/2747787>
- [4] J. Bahamonde, J. Bollen, E. Elejalde, L. Ferres, and B. Poblete, "Power structure in chilean news media," *PLOS ONE*, vol. 13, no. 6, pp. 1–18, 06 2018. [Online]. Available: <https://doi.org/10.1371/journal.pone.0197150>
- [5] N. Chaban, M. Knodt, Š. Liekis, and I. NG, "Narrators' perspectives: communicating the EU in ukraine, israel and palestine in times of conflict," *European Security*, pp. 1–19, Aug. 2019. [Online]. Available: <https://doi.org/10.1080/09662839.2019.1648256>
- [6] S. Kelly, "New zealand elite perceptions on the eu: A longitudinal analysis," *Baltic Journal of European Studies*, vol. 3, no. 3, pp. 153–174, 2013. [Online]. Available: <https://content.sciendo.com/view/journals/bjes/3/3/article-p153.xml>
- [7] M. Georgiou and R. Zaborowski, "Media coverage of the "refugee crisis": A cross-european perspective." *Council of Europe report (DG1(2017)03)*. [Online]. Available: <http://eprints.lse.ac.uk/id/eprint/84672>
- [8] UN Refugee Agency, "UNHCR - the un refugee agency," 2019. [Online]. Available: <https://www.unhcr.org/>
- [9] N. Chaban and M. Holland, "Seeing the eu from outside its borders: Changing images of europe," *Baltic Journal of European Studies*, vol. 3, no. 3, pp. 3 – 12, 2013. [Online]. Available: <https://content.sciendo.com/view/journals/bjes/3/3/article-p3.xml>
- [10] Europe Direct, "EUROSTAT - your key to european statistics," 2019. [Online]. Available: <https://ec.europa.eu/eurostat/data/database>
- [11] G. J. Borjas, "Economic theory and international migration," *The International Migration Review*, vol. 23, no. 3, pp. 457–485, 1989. [Online]. Available: <http://www.jstor.org/stable/2546424>
- [12] S. Castles, H. De Haas, and M. J. Miller, *The age of migration: International population movements in the modern world*. Macmillan International Higher Education, 2013.
- [13] M. Machill, M. Beiler, and C. Fischer, "Europe-topics in europe's media: The debate about the european public sphere: A meta-analysis of media content analyses," *European Journal of Communication*, vol. 21, no. 1, p. 57–88, 2006. [Online]. Available: <https://doi.org/10.1177/0267323106060989>
- [14] C. Timmerman, K. Hemmerichs, and H. M.-L. D. Clerck, "The relevance of a "culture of migration" in understanding migration aspirations in contemporary turkey," *Turkish Studies*, vol. 15, no. 3, pp. 496–518, 2014.
- [15] N. Chaban and M. Holland, *Communicating Europe in times of crisis: External perceptions of the European Union*. Springer, 2014.
- [16] L. Adam, "The significance of eu topics in national media. has there been an europeanization of reporting in the national media," in *Bruges Political Research Papers No. 27*. College of Europe, 2012.
- [17] A. Frech, "'mass migration', crime and 'decent people': The portrayal of polish migrants in british newspapers," 2008. [Online]. Available: <https://reutersinstitute.politics.ox.ac.uk/our-research/mass-migration-crime-and-decent-people-portrayal-polish-migrants-british-newspapers>
- [18] T. Groeling, "Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news," *Annual Review of Political Science*, vol. 16, no. 1, pp. 129–151, 2013.
- [19] T. Groseclose and J. Milyo, "A measure of media bias," *The Quarterly Journal of Economics*, vol. 120, no. 4, pp. 1191–1237, 2005. [Online]. Available: <http://www.jstor.org/stable/25098770>
- [20] M. McCluskey and Y. M. Kim, "Moderatism or polarization? representation of advocacy groups' ideology in newspapers," *Journalism & Mass Communication Quarterly*, vol. 89, no. 4, pp. 565–584, 2012. [Online]. Available: <https://doi.org/10.1177/1077699012455385>
- [21] F. Ribeiro, L. Henrique, F. Benevenuto, A. Chakraborty, J. Kulshrestha, M. Babaei, and K. Gumadi, "Media bias monitor: Quantifying biases of social media news outlets at large-scale," 2018.
- [22] C. Budak, S. Goel, and J. M. Rao, "Fair and balanced? quantifying media bias through crowdsourced content analysis," 2016.
- [23] A. Dallmann, F. Lemmerich, D. Zoller, and A. Hotho, "Media bias in german online newspapers," in *HT*, 2015.
- [24] E. Elejalde, L. Ferres, and E. Herder, "On the nature of real and perceived bias in the mainstream media," *PLOS ONE*, vol. 13, no. 3, pp. 1–28, 03 2018. [Online]. Available: <https://doi.org/10.1371/journal.pone.0193765>
- [25] F. Hamborg, K. Donnay, and B. Gipp, "Automated identification of media bias in news articles: an interdisciplinary literature review," *International Journal on Digital Libraries*, Nov 2018.
- [26] D. Bourgeois, J. Rappaz, and K. Aberer, "Selection bias in news coverage: Learning it, fighting it," in *Companion Proceedings of the The Web Conference 2018*, ser. WWW '18, 2018, pp. 535–543.
- [27] J. Gruenewald, J. Pizarro, and S. M. Chermak, "Race, gender, and the newsworthiness of homicide incidents," *Journal of criminal justice*, vol. 37, no. 3, pp. 262–272, 2009.
- [28] D. Saez-Trumper, C. Castillo, and M. Lalmas, "Social media news communities: gatekeeping, coverage, and statement bias," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 1679–1684.
- [29] T. Groseclose and J. Milyo, "A measure of media bias," *The Quarterly Journal of Economics*, vol. 120, no. 4, pp. 1191–1237, 2005.
- [30] M. Sanderson, "Duplicate detection in the reuters collection," " *Technical Report (TR-1997-5) of the Department of Computing Science at the University of Glasgow G12 8QQ, UK*", 1997.
- [31] Z. Papacharissi and M. de Fatima Oliveira, "News frames terrorism: A comparative analysis of frames employed in terrorism coverage in us and uk newspapers," *The international journal of press/politics*, vol. 13, no. 1, pp. 52–74, 2008.
- [32] P. K. Bhowmick, A. Basu, and P. Mitra, "Reader perspective emotion analysis in text through ensemble based multi-label classification framework." *Computer and Information Science*, vol. 2, no. 4, pp. 64–74, 2009.
- [33] S. Pinto, P. Balenzuela, and C. Dorso, "Setting the agenda: Different strategies of a mass media in a model of cultural dissemination," *Physica A: Statistical Mechanics and its Applications*, vol. 458, 12 2015.
- [34] J.-M. Eberl, "Lying press: Three levels of perceived media bias and their relationship with political preferences," 2019.
- [35] V. Niculae, C. Suen, J. Zhang, C. Danescu-Niculescu-Mizil, and J. Leskovec, "Quotus: The structure of political media coverage as revealed by quoting patterns," in *Proceedings of the 24th International Conference on World Wide Web*, ser. WWW '15, 2015, p. 798–808.
- [36] C. Tan, H. Peng, and N. A. Smith, "'you are no jack kennedy": On media selection of highlights from presidential debates," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18, 2018, p. 945–954.
- [37] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter," in *Proceedings of the 20th International Conference on World Wide Web*, 2011, pp. 695–704.
- [38] R. Abebe, J. Kleinberg, D. Parkes, and C. E. Tsourakakis, "Opinion dynamics with varying susceptibility to persuasion," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '18, 2018, p. 1089–1098.
- [39] R. Rotabi, C. Danescu-Niculescu-Mizil, and J. Kleinberg, "Competition and selection among conventions," in *Proceedings of the 26th International Conference on World Wide Web*, ser. WWW '17, 2017, p. 1361–1370.

- [40] G. K. Zipf, "Some determinants of the circulation of information," *The American Journal of Psychology*, vol. 59, no. 3, pp. 401–421, 1946. [Online]. Available: <http://www.jstor.org/stable/1417611>
- [41] —, "The  $P_1P_2/D$  hypothesis: on the intercity movement of persons," *American Sociological Review*, vol. 11, no. 6, pp. 677–686, 1946. [Online]. Available: <http://www.jstor.org/stable/2087063>
- [42] Elejalde, Erick, Ferres, Leo, and Schifanella, Rossano, "Understanding news outlets' audience-targeting patterns," *EPJ Data Sci.*, vol. 8, no. 1, p. 16, 2019. [Online]. Available: <https://doi.org/10.1140/epjds/s13688-019-0194-8>
- [43] S. M. Mohammad, M. Salameh, and S. Kiritchenko, "How translation alters sentiment," *Journal of Artificial Intelligence Research*, vol. 55, pp. 95–130, 2016.
- [44] I. Flaounas, M. Turchi, O. Ali, N. Fyson, T. De Bie, N. Mosdell, J. Lewis, and N. Cristianini, "The structure of the eu mediasphere," *PLoS one*, vol. 5, no. 12, pp. 1–6, 2010.
- [45] T.-A. Hoang, K. D. Vo, and W. Nejdl, "W2e: A worldwide-event benchmark dataset for topic detection and tracking," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, ser. CIKM '18, 2018, p. 1847–1850.
- [46] P. J. Shoemaker, M. Eichholz, E. Kim, and B. Wrigley, "Individual and routine forces in gatekeeping," *Journalism & mass communication quarterly*, vol. 78, no. 2, pp. 233–246, 2001.
- [47] S. N. Soroka, "The gatekeeping function: Distributions of information in media and the real world," *The Journal of Politics*, vol. 74, no. 2, pp. 514–528, 2012.
- [48] L. Hänsel, F. C. Feyerabend, W. Glos, K. Dümmel, J. D. Rey, M. Lange, Z. Ilievski, S. Aleksoski, D. Pasoski, H. M. Sieg, and N. Berkmann-Dierkes, "The influence of external actors in the western balkans," 2018. [Online]. Available: [https://www.kas.de/c/document\\_library/get\\_file?uuid=194afc48-b3be-e3bc-d1da-02771a223f73&groupId=252038](https://www.kas.de/c/document_library/get_file?uuid=194afc48-b3be-e3bc-d1da-02771a223f73&groupId=252038)
- [49] S. E. Garding, "Bosnia and herzegovina: Background and u.s. policy," 2018. [Online]. Available: <https://fas.org/sgp/crs/row/R45691.pdf>
- [50] V. L. Morelli, "Kosovo: Background and u.s. relations," 2018. [Online]. Available: <https://fas.org/sgp/crs/row/R44979.pdf>
- [51] M. A. Bujang and N. Baharum, "Sample size guideline for correlation analysis," *World*, vol. 3, no. 1, 2016.
- [52] C. Wang, M. Zhang, L. Ru, and S. Ma, "Automatic online news topic ranking using media focus and user attention based on aging theory," in *Proceedings of the 17th ACM conference on Information and knowledge management*, 2008, pp. 1033–1042.
- [53] Welch's unequal variances t-test, "Welch's t-test," 2020. [Online]. Available: [https://en.wikipedia.org/wiki/Welch's\\_t-test](https://en.wikipedia.org/wiki/Welch's_t-test)
- [54] W. P. Eveland Jr, K. Marton, and M. Seo, "Moving beyond "just the facts" the influence of online news on the content and structure of public affairs knowledge," *Communication Research*, vol. 31, no. 1, pp. 82–108, 2004.
- [55] K. Kobayashi, "The influence of topic knowledge, external strategy use, and college experience on students' comprehension of controversial texts," *Learning and Individual Differences*, vol. 19, no. 1, pp. 130–134, 2009.
- [56] E. Elejalde, L. Ferres, E. Herder, and J. Bollen, "Quantifying the ecological diversity and health of online news," *Journal of computational science*, vol. 27, pp. 218–226, 2018.
- [57] P. Müller and A. Schulz, "Alternative media for a populist audience? exploring political and media use predictors of exposure to Breitbart, Sputnik, and Co." *Information, Communication & Society*, pp. 1–17, 2019.
- [58] S. Sanovich, "Computational propaganda in Russia: the origins of digital misinformation," 2017.
- [59] J. P. Chang, J. Cheng, and C. Danescu-Niculescu-Mizil, "Don't let me be misunderstood: Comparing intentions and perceptions in online discussions," in *Proceedings of the 29th International Conference on World Wide Web*, ser. WWW '20, 2020.



**Koustav Rudra** is presently postdoctoral researcher at L3S Research Center, Leibniz University Hannover. He completed his Ph.D. from the Department of CSE, Indian Institute of Technology, Kharagpur, India. He received his M.Tech. from the same department, and a B.E. in Computer Science from BESU, Shibpur (which is presently known as IEST Shibpur). His research interest lies in information retrieval, social computing, AI for social good.



traditional and social media in the context disaster-communication.

**Gerhard Backfried** is one of the founders and currently holds the position of Chief Scientific Officer at HENSOLDT Analytics. He received a master's degree in computer science (M.Sc.) from the Technical University of Vienna with speciality in AI and Linguistics and a Ph.D. degree in Computer Science from the University of Vienna. Prior to joining HENSOLDT, Gerhard worked in the fields of expert systems for the finance sector and personal dictation systems. More recently he has been focusing on the combination of



retrieval, data engineering and data analysis in a social context.

**Miroslav Shaltev** is currently part of the Research Data Infrastructure group of the Leibniz Centre for Tropical Marine Research, Bremen, Germany. He received his Diploma in Physics in 2009 and Dr. rer. nat. degree in 2013 from Leibniz University, Hannover, Germany. During his doctoral studies at Albert Einstein Institute in Hannover he did data analysis of gravitational wave data searching for continuous wave signals. During his time at L3S Research Center he was interested in various aspects of information



is involved in the European project MIRROR, where she and her team are working on media-analysis methods for better understanding perceptions and misperceptions in the context of migration.

**Dr. Claudia Niederée** works as a research group leader Research group leader at the L3S Research Center in Hannover. She holds a master in Computer Science and was awarded a doctoral degree in Computer Science by the Technical University Hamburg-Harburg (Germany). Her main research interests are diversity in knowledge, digital forgetting, social media analysis, entity-centric technology and AI-based systems. She has published more than 100 scientific articles and papers. Currently she



**Erick Elejalde** obtained his Ph.D. in Computer Science from the University of Concepción, Concepción, Chile. He also holds an M.Sc. in Computer Science from the same university. He works now as a postdoctoral researcher at the L3S Research Center, Leibniz University Hannover, Germany. His research interests include computational social science, online media-behavior modeling, and social networks.