

# Study of the fidelity of media, as representations of the society state

Lucas Burget

Sinan Gökçe

Juliane Dervaux

lucas.burget@epfl.ch sinan.gokce@epfl.ch juliane.dervaux@epfl.ch

## Abstract

In actual societies, media have a huge role. This paper aims to use the GDELT V2 data set to retrieve characteristics of media articles (tone, source origin) and of their related events (localization, Goldstein scale). These data are combined with objective indicators such as happiness score and democratic index, and a statistical analysis is undertaken to assess whether media are realistic representations of the society state.

## 1 Introduction

Our daily life as twenty-first century earth residents is constantly punctuated by the overabundance of news and we like to make opinions on all kinds of hot topics. The sources of these news thus have a dramatic impact on our conception of the world, and it is fundamental for us to be aware of their bias, and be cautious when assimilating all the information they spread.

In this project, it has been decided to focus on several characteristics of media articles in different countries and their correlation with the level of happiness of their inhabitants. It's also aimed to find a link between the political regime and the media characteristics concerning specific events, such as protests. These analysis will, at the end of the report, will be profitable to have an answer to the following question: Do media represent the current state of the society regarding the well-being and some important political factors namely, corruption, confidence in the government and democracy rank?

## 2 Method

### 2.1 Dataset

#### 2.1.1 GDELT

The main data set used in this project in order to study the emotions conveyed by media is the GDELT V2 (Global Data on Events, Location and Tone).

This massive open database contains over a quarter-billion event records, since the end of 2014, obtained from broadcasts, prints and online news sources from all around the World. Powerful translation tools are used to process in almost real-time articles from more than 100 languages. Each record is associated to precise geographic references, both concerning the action location and its actors, thanks to the pre-processing using full text geo-coding tool based on Leetaru (2012). The global tone of the article in which the event was mentioned is assessed using an algorithm from Shook et al. (2012). This algorithm computes a tonal score by subtracting the percentage of words in the text found in a negative lexicon to the percentage of words in a positive lexicon.

In order to study the vision of one country's media on another, both locations are needed. Thankfully, one field of the mention table is the URL of the source, and a table linking each URL to its source country is provided on the GDELT blog (2018). On the other hand, the field *Actiongeo\_countrycode* in the event table has been used to assess the event location. For further explanation of the different field of the table, please refer to the GDELT Event Codebook V2.0 (2015).

There are a few major statistics to keep in mind when dealing with this data set. There are more than 1.2 billion pair mentions-event but the data

set is US-centered: 45% of the mentions and 27% of the events concern them. For the other countries, for both mentions and event number, 50% of them appear between  $10^5$  and  $10^6$  times. Finally, the distribution of the different event types is not even: around 40% of them are about *Consultation* and *Making public statement*, and they are rather positive.

### 2.1.2 World Happiness Report 2018

The second data set used in this project, is the World Happiness Report (2018). The following national variables were selected for this study : happiness score (answer to a subjective question on personal well-being and life quality), confidence in the national government and perception of corruption. These three measures are from the Gallup World Poll (2017).

### 2.1.3 Democracy Index

The Democracy Index (2017) provides a metrics for the democracy worldwide for 165 independent states and two territories. This data set has been used over the metric contained in the World Happiness Report (2018) because it allows grouping countries into four different categories namely full democracy, flawed democracy, hybrid regime and authoritarian.

## 2.2 Statistical approach

In order to obtain relevant statistical variables, a bootstrapping method was implemented. This method consists in sampling the subset data with replacement several times, in order to get a distribution of the feature of interest, and to be able to locate its true value at the population-level in a confidence interval.

All correlation values cited below result from a bootstrapping step, by taking the median value and illustrating the 95%-confidence interval of each distribution.

## 3 Results and Discussion

### 3.1 Tone of the world's media

Have you ever heard acquaintances saying 'I don't read the news anymore, they are way too depressing!' ? Indeed, the world's media are negative, no matter the topic, as we can see in Figure 1. One might wonder if this result is biased toward the US, due to its over-representation in the data. But the result is similar with or without the US media sources.

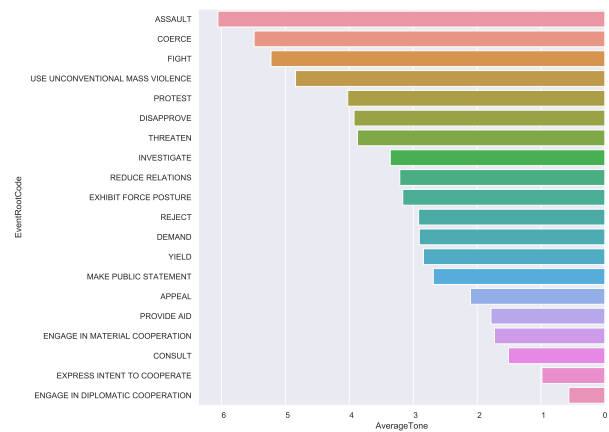


Figure 1: Bar plot representing the average tone of all the categories of events

### 3.2 Well-Being

To investigate deeper, it seemed useful to examine whether people leaving in a country where media sources are negative on average, are indeed less happy. All GDELT event mentions were sorted on the basis of the location of their sources and the average tone per country was computed, and the resulting correlation is -0.20, as illustrated in the Figure 2. Thus, very surprisingly, the happier the inhabitants of a country, the more negative its media sources and the popular belief about the anxiety-provoking effect of media might be wrong. The media, in order to sell, will rather exaggerate and appeal to reader's emotions, but they are able to put the information into perspective.

Another studied aspect was the link between the overall media tone concerning all events happening in a specific country and its happiness score. GDELT mentions were sorted on the basis of the event location and the average tone was computed. The resulting correlation is of 0.31 (Figure 2). The interpretation of this result is that the overall tonal image emerging from media representation concerning a country is positively correlated with the well-being of its inhabitants.

One could also think that the stability of a country impacts the well-being experienced internally. The Goldstein Scale is an indicator of the impact of an event on the stability of a country, in a range [-10;10] (see GDELT Event Codebook V2.0 (2015)). It associates a Goldstein Scale value to each CAMEO event type, which allows the computation of the average Goldstein Scale value over

all events that happened in a specific country during the last years. As an example, the Palestinian territory was found to have the lowest average Goldstein Scale value (-1.57) and on the other hand the highest average value (2.48) belongs to the Tuvalu island country.

The correlation between average Goldstein Scale and happiness score is 0.31, showing a positive trend between the stability of a country as it is represented in the media, and the well-being.

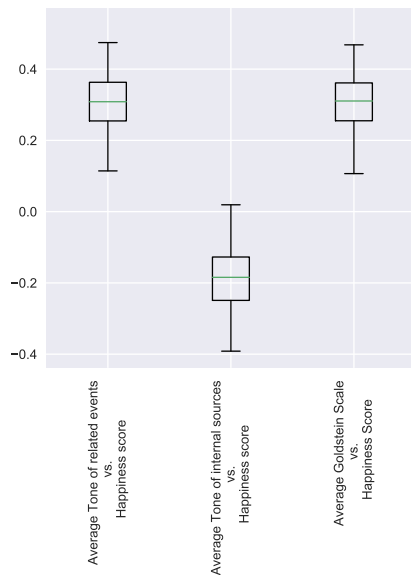


Figure 2: Bar plots representing the bootstrapping resulting 95%-confidence intervals of the correlation values between : a. Average tone over all event mentions and Happiness score; b. Average tone of all internal sources and Happiness score; c. Average Goldstein Scale over all events and Happiness score.

### 3.3 Political regime

Another important component of a society's state, is its political regime. Therefore, the relationship between some political indicators and the media content was questioned. First, using the Democracy Index (2017), the correlation between the democratic quality of a country and the average tone of its media sources as been computed. The resulting correlation was -0.27. Using the classification of the countries in four different democratic categories, this trend was confirmed. However, this average may not be fully relevant as it comprises mentions of events of various CAMEO types, thus one might miss more specific information related to the event type. Therefore,

event mentions related to five different EventRootCodes of interest (Engage in diplomatic cooperation, Engage in material cooperation, Protest, Exhibit force posture and Fight) were selected and their average tone for each democratic category was computed. Figure 3 illustrates similar results: media sources located in authoritarian regimes are talking about these five event topics in a more positive manner. The interpretation of this finding is not straightforward but one could conclude that in democratic countries, media are more free to express their disagreement and to criticize.

Then, one can wonder whether the image a country spreads about itself is more positive than the image expressed by external media sources. Thus, the average tone in internal and external media was computed for each country. Then, the interest was focused on the difference between these two metrics, as illustrated by Figure 4, and its link with the democratic quality of the country (Figure 5). Interestingly, a median correlation of -0.52 was found. This result means that the more a country is authoritarian, the more its media sources expose a positively-biased representation of it.

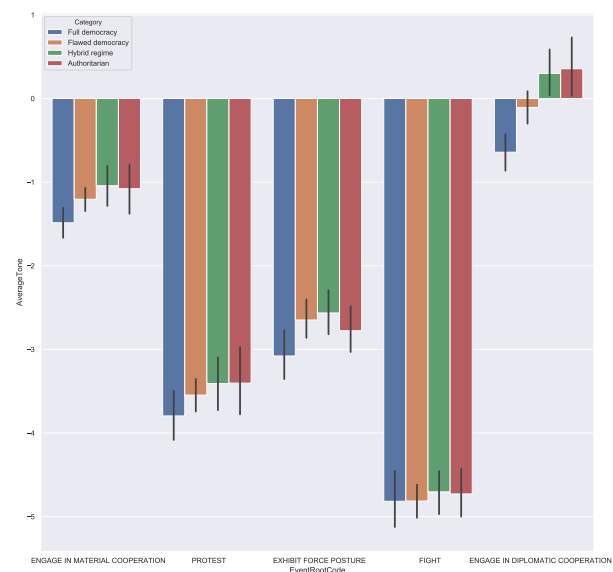


Figure 3: Bar plot representing the average tone of each democratic category, concerning five relevant event types

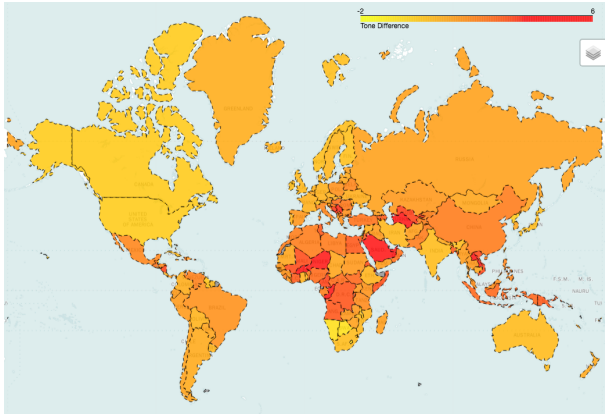


Figure 4: World map illustrating the difference in average tone between Internal and External media sources.

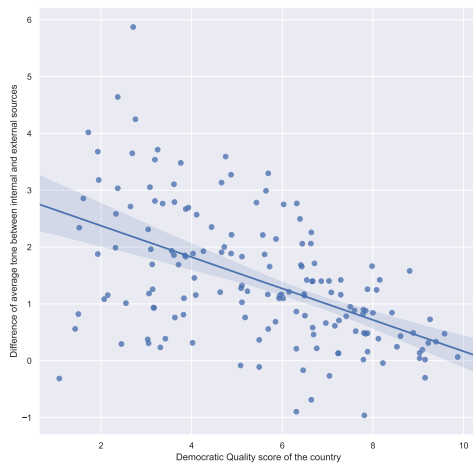


Figure 5: Scatter plot representing the relationship between the internal vs. external tone difference and the democratic quality.

### 3.4 Protests

In this section, the attention has been turned specifically to protests for two reasons. Firstly, they are reflecting the citizen's judgment against to regime in the country and are strongly related to the happiness and politic factors by their nature. Secondly, protests are the events which have considerable impact on the stability of countries.

The analysis starts with explaining the correlation between the number of protest mentions and the actual number of protest events, occurring in 233 countries.

There is a perfect correlation of 0.98 between protest events and their reflection in the media. Thus a further analysis can be done in order to investigate if the media represent the current states of countries in term of protests.

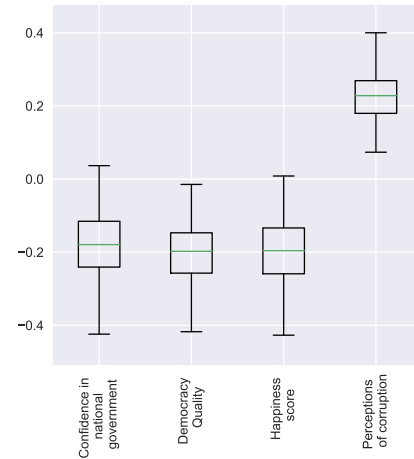


Figure 6: Box plots for correlations between normalized protest mentions and different indicators (95% confidence interval)

The following analysis considers the number of protest mentions normalized by the amount of all mentions related to a country, to get a notion of importance of protest mentions. In the interest of observing the relation between the previously mentioned indicators and protest news, the normalized number of protest news and the indicator scores have been gathered for each country. This fusion caused to reduce the number of countries to 135, since some information misses for some countries. The results are summarized in Figure 6.

The results from Figure 6 justify the general intuition stating that the importance of protest news in a more corrupt country should be higher and lower in a happier country.

## 4 Conclusion

After analyzing the different links between vision of the world by the media and some objective indicators, one can draw some important conclusions. First of all, there is overall a negative media "offset", even if it doesn't seem to have an anxiogenic impact. However, one can still retrieve several society characteristics in the flow of news. Indeed, the global image of a country over all media news represents accurately its internal state, in term of well being and protest events. Finally there is a bias with the democratic quality: media of more authoritarian countries see themselves in a more positive way than others do.

## References

- Kalev H. Leetaru. 2012. *Fulltext Geocoding Versus Spatial Metadata for Large Text Archives: Towards a Geographically Enriched Wikipedia*. D-Lib Magazine, volume 18, number 9/10.
- E. Shook, K. Leetaru, G. Cao, A. Padmanabhan, and S. Wang. 2012. *Happy or not: Generating topic-based emotional heatmaps for Culturomics using CyberGIS*. 8th International Conference on EScience (IEEE), pages 1-6.
- Mapping The Media: A Geographic Lookup Of GDELT's Sources  
<https://blog.gdeltproject.org/mapping-the-media-a-geographic-lookup-of-gdelts-sources/>
- GDELT Event Codebook V2.0  
<http://data.gdeltproject.org/documentation/GDELT-Event.Codebook-V2.0.pdf>
- John F. Helliwell, Richard Layard and Jeffrey D. Sachs. [https://s3.amazonaws.com/happiness-report/2018/WHR\\_web.pdf](https://s3.amazonaws.com/happiness-report/2018/WHR_web.pdf)
- Gallup World Poll. <https://bit.ly/2UOVg1R>
- The Economist Intelligence Unit  
<https://bit.ly/2Q2kGGf>