Evolution of Media Coverage on Climate Change and Environmental Awareness: an Analysis of Tweets from UK and US Newspapers

Gianpaolo Zammarchi, Maurizio Romano, and Claudio Conversano

Abstract Climate change represents one of the biggest challenges of our time. Newspapers might play an important role in raising awareness on this problem and its consequences. We collected all tweets posted by six UK and US newspapers in the last decade to assess whether 1) the space given to this topic has grown, 2) any breakpoint can be identified in the time series of tweets on climate change, and 3) any main topic can be identified in these tweets. Overall, the number of tweets posted on climate change increased for all newspapers during the last decade. Although a sharp decrease in 2020 was observed due to the pandemic, for most newspapers climate change coverage started to rise again in 2021. While different breakpoints were observed, for most newspapers 2019 was identified as a key year, which is plausible based on the coverage received by activities organized by the Fridays for Future movement. Finally, using different topic modeling approaches, we observed that, while unsupervised models partly capture relevant topics for climate change, such as the ones related to politics, consequences for health or pollution, semi-supervised models might be of help to reach higher informativeness of words assigned to the topics.

Keywords: climate change, Twitter, environment, time series, topic modeling

Gianpaolo Zammarchi (✉)
University of Cagliari, Viale Sant'Ignazio 17, 09123, Cagliari, Italy,
e-mail: gp.zammarchi@unica.it

Maurizio Romano
University of Cagliari, Viale Sant'Ignazio 17, 09123, Cagliari, Italy,
e-mail: romano.maurizio@unica.it

Claudio Conversano
University of Cagliari, Viale Sant'Ignazio 17, 09123, Cagliari, Italy, e-mail: conversa@unica.it

P. Brito et al. (eds.), *Classification and Data Science in the Digital Age*,
Studies in Classification, Data Analysis, and Knowledge Organization,
https://doi.org/10.1007/978-3-031-09034-9_43

1 Introduction

Climate change is one of the biggest challenges for our society. Its consequences which include, among others, glaciers melting, warming oceans, rising sea levels, and shifting weather or rainfall patterns, are already impacting our health and imposing costs on society. Without drastic action aimed at reducing or preventing human-induced emissions of greenhouse gasses, these consequences are expected to intensify in the next years. Despite its global and severe impacts, individuals may perceive climate change as an abstract problem [1]. It is also a well-known fact that the level of information plays a crucial role in the awareness about a topic (e.g. healthy food [2] and smoking [3]) . Media represent a crucial source of information and can exert substantial effects on public opinion, thus helping to raise the awareness on climate change. For instance, media can explain climate change consequences as well as portraying actions that governments, communities and single individuals can take. For this reason, it is important to distinguish themes that might have gained popularity from those that may have seen a decrease of interest. Nowadays, social media have become a reliable and popular source of information for people from all around the world. Twitter is one of the most popular microblogging services and is used by many traditional newspapers on a daily basis. While we can hypothesize that in the last few years the media coverage on climate change might have risen, due for instance to international climate strike movements, the recent emergence of the coronavirus disease 2019 (COVID-19) pandemic might have led to a decrease of attention on other relevant topics.

Aims of this work were to: (1) assess trends in media coverage on climate change using tweets posted by main international newspapers based in United Kingdom (UK) and United States (US), and (2) identify the main topics discussed in these tweets using topic modeling.

2 Dataset and Methods

We downloaded all tweets posted from 2012 January 1st to 2021 December 31st from the official Twitter account of six widely known newspapers based in UK (The Guardian, The Independent and The Mirror) or US (The New York Times, The Washington Post and The Wall Street Journal) leading to a collection of 3,275,499 tweets. Next, we determined which tweets were related to climate change and environmental awareness based on the presence of at least one of the following keywords: “climate change”, “sustainability”, “earth day”, “plastic free”, “global warming”, “pollution”, “environmentally friendly” or “renewable energy”. We plotted the number of tweets on climate change posted by each newspaper during each year using R v. 4.1.2 [4].

We analyzed the association between the number of tweets on climate change and the whole number of tweets posted by each newspaper using Spearman’s correlation analysis. For each year and for each newspaper, we computed and plotted the differences in the number of posted tweets compared to the previous year, for either (a)

tweets related to climate change and (b) all tweets. Finally, we used the changepoint R package [5] to conduct an analysis aimed at identifying structural breaks, i.e. unexpected changes in a time series. In many applications, it is reasonable to believe that there might be *m* breakpoints (especially if some exogenous event occurs) in which a shift in mean value is observed. The changepoint package estimates the breakpoints using several penalty criteria such as the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). We estimated the breakpoints using the Binary Segmentation (BinSeg) method [6] implemented in the package.

Lastly, we used tweets posted by The Guardian to perform topic modeling, a method for classification of text into topics. Preprocessing (including lemmatization, removal of stopwords and creation of the document term matrix) was conducted with tm [7] and quanteda [8] in R. We used two different approaches: 1) Latent Dirichlet Allocation (LDA) implemented in the textmineR R package [9]; and 2) Correlation Explanation (CorEx), an approach alternative to LDA that allows both unsupervised as well as semi-supervised topic modeling [10].

3 Results

3.1 Analysis of Tweet Trends and Breakpoints

Among 3,275,499 collected tweets, we identified 11,155 tweets related to climate change and environmental awareness. Figure 1A shows the number of tweets on climate change posted by each of the analyzed newspapers from 2012 to 2021, while Figure 1B the total number of tweets posted by each newspaper.

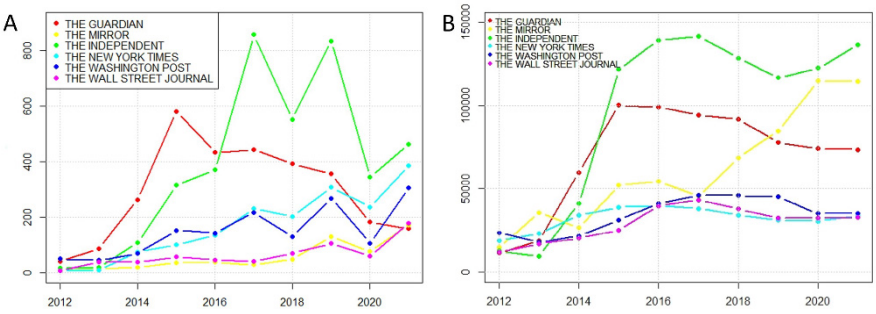

Fig. 1 Number of tweets on climate change (A) or total number of tweets (B) posted by the six newspapers from 2012 to 2021.

For the majority of newspapers, the number of tweets on climate change increased from 2014 to 2019, saw a sharp decrease in 2020, in correspondence of the emergence of the COVID-19 pandemic, and a subsequent rise in 2021. On the other hand, the

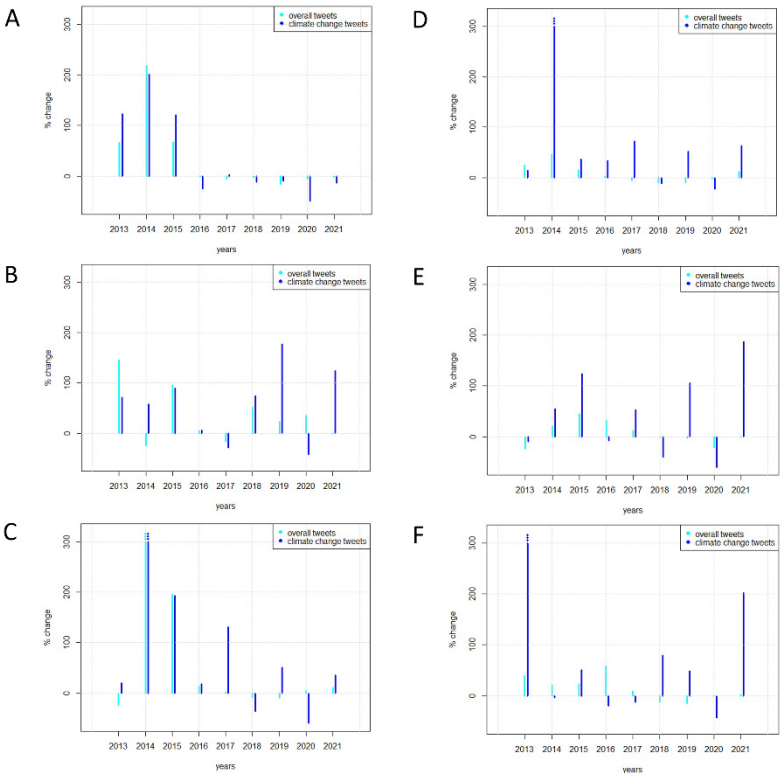

Fig. 2 Year-over-year percentage changes of overall tweets and tweets on climate change. A: The Guardian, B: The Mirror, C: The Independent, D: The New York Times, E: The Washington Post, F, The Wall Street Journal.

number of tweets on climate change posted by The Guardian showed a peak during 2015 and a subsequent decrease. However, it must be noted that The Guardian is also the newspaper that showed a more pronounced decrease in the overall number of tweets.

The number of tweets on climate change was significantly positively correlated with the overall number of tweets posted from 2012 to 2021 for four newspapers (The Guardian, Spearman's rho = 0.95, $p < 0.001$; The Mirror, Spearman's rho = 0.95, $p < 0.001$; The Independent, Spearman's rho = 0.76, $p = 0.016$; The Washington Post, Spearman's rho = 0.70, $p = 0.031$) but not for The New York Times (Spearman's rho = 0.18, $p = 0.63$) or The Wall Street Journal (Spearman's rho = 0.49, $p = 0.15$). Year-over-year percentage changes among either tweets related to climate change or all posted tweets can be observed in Figure 2.

Looking at Figure 2, we can observe a great variability in the posted number of tweets during the years, both for the total number of tweets and for the number of tweets on climate change. While the analysis aimed at identifying structural changes

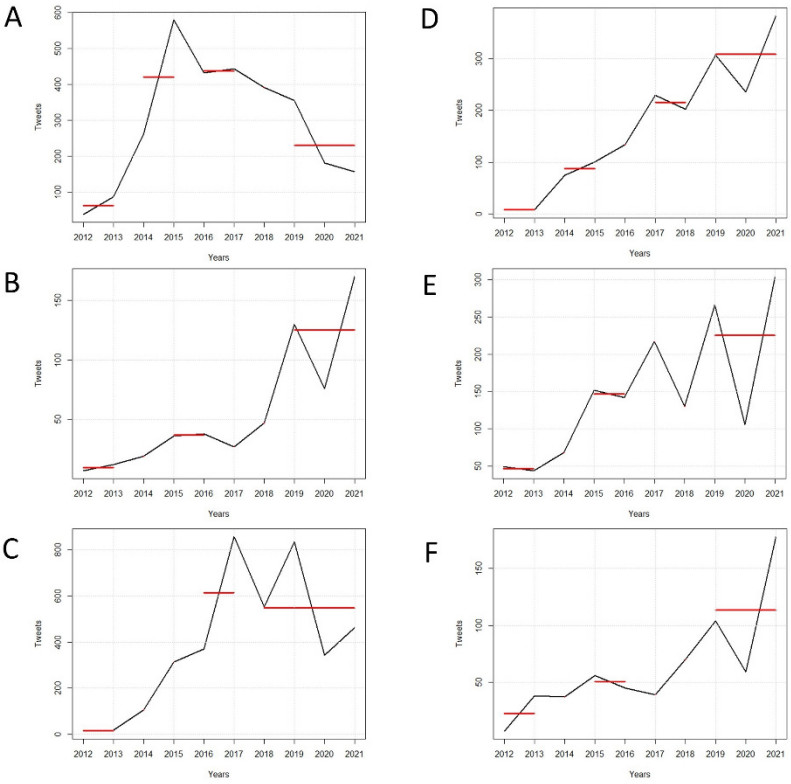

Fig. 3 Structural changes in the time series of tweets related to climate change. A: The Guardian, B: The Mirror, C: The Independent, D: The New York Times, E: The Washington Post, F, The Wall Street Journal. The red line represents the years between two breakpoints.

in the time series comprising tweets on climate change identified three or four breakpoints for all newspapers, wide variability was observed regarding the specific year in which these structural changes were identified (Figure 3). Despite the great variability, Figure 3 shows that even if a common breakpoint cannot be identified, 2019 was a key year for five out of six newspapers (except for The Independent).

3.2 Topic Modeling

Finally, we exploited the topic modeling approach to identify and analyze the main topics discussed by newspapers in their tweets. Due to space limitations, we focus only on The Guardian since this newspaper showed a trend in contrast with the others. Data comes from 2,916 tweets posted by The Guardian analyzed using LDA and CorEx. For LDA, a range of 5-20 unsupervised topics was tested, with the most

interpretable results obtained with 10 topics (Table 1). The topic coherence ranged from 0.01 to 0.34 (mean: 0.13). For each topic, bi-gram topic labels were assigned with the labeling algorithm implemented in textmineR. We can observe that topics are related to politics or leaders (Topics 3, 7 and 10), environmental scientists or climate journalists (Topics 1 and 5), energy sources (Topics 4 and 8) and effects of climate change (Topics 2, 6 and 9). The intertopic distance map obtained with LDAvis is shown in Figure 4. The area of each circle is proportional to the relative prevalence of that topic in the corpus, while inter-topic distances are computed based on Jensen-Shannon divergence.

Table 1 Top terms for the ten topics identified with LDA.

dana_nuccitelli	air_pollution	barack_obama	renewable_energy	john_abraham
dana	pollution	fight	energy	john
dana_ nuccitelli	air	obama	renewable	trump
nuccitelli	air_pollution	trump	renewable_energy	australia
live	study	plan	uk	tackle
trump	finds	battle	sustainability	abraham
air_pollution	donald_trump	fossil_fuel	extreme_weather	pope_francis
pollution	trump	report	world	pollution
air	schoolstrike	fossil	paris	study
air_pollution	school	ipcc	leaders	tackling
uk	great	warns	talks	pope
tackle	donald	stop	deal	scientists

Fig. 4 Intertopic distance map.

Finally, we conducted a semi-supervised topic modeling analysis based on anchored words using CorEx. When anchoring a word to a topic, CorEx maximizes the mutual information between that word and the topic, thus guiding the topic model towards specific subsets of words. A model with 5 topics and three anchored words for each topic (Table 2) showed a total correlation (i.e. the measure maximized by CorEx when constructing the topic model) of 4.36. This value was higher compared to the one observed with an unsupervised CorEx analysis with the same number of topics (total correlation = 0.97, topics not shown due to space limits). Topics related to politics (Topic 3) and science (Topic 5) were found to be the most informative in our dataset based on the total correlation metric.

Table 2 Topics with anchored words and examples of tweets.

Topic	Topic words	Examples of tweets per topic
1	**school, strike, march,** schoolstrik, climatestrikeuk, ukschoolstrik, schoolstrikeclim, climatemarch, arabia, saudi	EPA wipes its climate change site day before march on Washington
2	**ocean, ice, environment,** john, dana, nuccitelli, air, abraham, sea, reed	Chasing Ice filmmakers plumb the 'bottomless' depths of climate change - new clip from @GuardianEco
3	**trump, obama, lead,** donald, barack, ivanka, brighton, repli, administr, pick	Trump administration pollution rule strikes final blow against environment
4	**plastic, fuel, oil,** fossil, compani, pictur, wast, big, bay, photo	Engaging with oil companies on climate change is futile
5	**studi, scientist, research,** find, link, say, show, death, prematur, speci	Microplastic pollution revealed 'absolutely everywhere' by new research

The anchored words are reported in bold.

4 Discussion

The present study aims to evaluate how some of the most relevant British and American newspapers have given space to the topic of climate change on their Twitter page in the last decade. Apart from The Guardian, which shows a decreasing trend in the number of tweets related to climate change, all the other newspapers showed an overall growing trend, except during 2020. During this year, the number of tweets related to climate change declined for all six newspapers. This was most probably due to the COVID-19 outbreak that was massively covered by all media. By analyzing the breakpoints in Figure 3, it is possible to observe that 2019 was a relevant year for climate change. This is plausible considering that, starting from the end of 2018, the strikes launched by the Fridays for Future movement to raise awareness on the issue of climate change, gained high media coverage.

Our topic modeling analysis showed that the main topics defined using unsupervised models such as LDA are mostly related to politics, environmental scientists, energy sources and effects of climate change. While unsupervised models capture relevant topics, using CorEx we found a semi-supervised model to be able to reach a higher total correlation, which is a measure of informativeness of the topics, compared to an unsupervised model with the same number of topics.

As future developments, we plan to extend our analyses to newspapers from other countries. We believe our work to be useful to gain more knowledge and awareness about the climate change topic and on how much space relevant newspapers have given to this issue on social media. Increasing the knowledge about the nature of the topics covered by newspapers will lay the basis for future studies aimed at evaluating public awareness on this highly relevant challenge.

References

1. Van Lange, P. A. M., Huckelba, A. L.: Psychological distance: How to make climate change less abstract and closer to the self. Curr. Opin. Psychol. **42**, 49–53 (2021)
2. Wakefield, M., Flay B., Nichter M., Giovino G.: Role of the media in influencing trajectories of youth smoking. Addiction, **98**, 79-103 (2003)
3. Dumanovsky, T., Huang, C. Y., Bassett, M. T., Silver, L. D.: Consumer awareness of fast-food calorie information in New York City after implementation of a menu labeling regulation. American Journal of Public Health, **12**, 2520-2525 (2010)
4. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2020). Available via `http://www.R-project.org`
5. Killick, R., Eckley, I. A.: changepoint: An R Package for changepoint analysis. J. Stat. Softw. **58**, 1–19 (2014)
6. Scott, A.J., Knott, M.: A cluster analysis method for grouping means in the analysis of variance, Biometrics, **30**, 507-512 (1974)
7. Feinerer, I., Hornik, K., Meyer, D.: Text mining infrastructure in R. J. Stat. Softw. **25**, 1–54 (2008)
8. Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A.: quanteda: An R package for the quantitative analysis of textual data. J. Open Source Softw. **3**, 774 (2018)
9. Jones, T., Doane, W.: Package 'textmineR'. Functions for Text Mining and Topic Modeling (2021). Retreived from `https://cran.r-project.org/web/packages/textmineR/textmineR.pdf`
10. Gallagher, R., Reing, K., Kale, D., Ver Steeg, G.: Anchored correlation explanation: Topic modeling with minimal domain knowledge. Trans. Assoc. Comput. **5**, 529-542 (2017)