Climate change and populists in geolocated Twitter

Francisco Porras-Bernardez, Georg Gartner

Research Group Cartography, TU Wien, Vienna, Austria.
francisco.porras.bernardez@tuwien.ac.at, georg.gartner@tuwien.ac.at

Abstract. Surveys have been one of the traditional tools to collect public opinions. However, social media are an important alternative to surveys, being a source of information easily available, in high volume and at low cost. There is plenty of literature dealing with the study of different social, political or environmental topics through social media such as Twitter. Climate change is one of these topics and has major relevance in our current society. In addition, politics is a common element of analysis in the platform. Nevertheless, there is not enough insight about the overall quantitative relevance of climate change compared with other topics such as politicians. Moreover, some of the literature focus specifically on geolocated tweets, which are a small fraction of the total posts generated. This work in progress deals with the identification and semantic analysis of geolocated posts in social media. We analyse and compare the presence of climate change with populist politicians in the platform. These political figures often have a controversial stance on climate change while enacting policies affecting millions of citizens. We aim to study how suitable is the platform for spatiotemporal analysis of public opinion on climate change, and how relevant is the topic on it compared to the presence of some populists. We also aim to provide guidance for further research based on geolocated tweets by estimating how much geolocated data is produced by which countries. More than 170 M geolocated tweets were extracted and analysed. Those tweets containing terms related to climate change in the official languages of the 14 most popular countries in the dataset, as well as the names of several politicians were filtered. Then, an analysis was performed to characterise the spatial and temporal global distribution of these posts during most of the past decade. This was compared with the dates of major events related with climate change and politics. Additionally, sentiment analysis was used to characterise the polarity of the posts. This paper presents an estimation of the relative presence of climate change in Twitter based on probably one of the largest geolocated tweets datasets existing.

This work will also offer a semantic analysis of the posts including a graph of the main terms used by country, as well as the polarity of the sentiments associated with climate change. This study has the potential to benefit policy makers, non-governmental organizations, activists, journalists and social media researchers worldwide.

Keywords. Twitter, Climate change, Populism, NLP, Sentiment Analysis

1. Introduction

Twitter is one of the most popular SM platforms offering geolocated content generated by its users. This and other SM have been used for the public discussion of topics of diverse relevance. This medium is a platform used by the general public, activists, governments, NGOs or public figures such as celebrities or politicians. Twitter has been widely used in research for the analysis of different phenomena including social, political or environmental issues (Doğu, 2019; Milleville et al., 2019; Yaqub et al., 2017). Climate change is an extremely relevant topic nowadays that is in the intersection of society, politics and environment.

Climate change presence in Twitter has been explored in previous studies attending to different topics such as its impacts, its causes, or even the acceptance of the phenomenon itself (Chen et al., 2019; Pathak et al., 2017). Additionally, many works have concentrated on analysing the sentiments associated to the tweets mentioning the term (Cody et al., 2015; Veltri & Atanasova, 2017). Moreover, other studies have focused on geolocated tweets to analyse regional differences or language and culture-based disparities (Bennett et al., 2021; Singleton et al., 2018). Nevertheless, most works have been limited to a few number of regions and languages and have built their conclusions on top of relatively limited datasets, spatially and temporally.

The role of politics in the platform has been analysed in terms of international events, political actions or direct communication related with climate change (Ebrey et al., 2020; Sanford et al., 2021). However, there has not been a specific focus on populist politicians and how prominent is their presence in the platform, compared directly to climate change. This is relevant due to their often-controversial stance towards the topic.

To the best of our knowledge, a comparative analysis of the overall presence of both topics in Twitter based on a large global geolocated dataset has not been carried out until now. There are differences in the rate of tweets geolocated among countries and the proportion of tweets geolocated is estimated in a 1% (Tasse et al., 2017). However, this kind of analysis could shed some light on the relative relevance of climate change in the overall public interest within Twitter. A differentiated analysis by countries and languages based on

geolocated tweets could contribute to build a better picture across regions and cultures.

For this study, we have processed more than 170 million geolocated tweets generated worldwide (2011-2020). The evolution of tweets produced in the countries with more users is being analysed and compared with the number of tweets dealing with climate change or four relevant populist politicians. We aim to estimate the overall presence of climate change in the platform compared to these political figures. This analysis has a wide coverage temporally, spatially and linguistically by taking into account terms in 14 languages. We are performing a basic analysis of the sentiments and the main terms of interest on these tweets in order to depict the topics commonly related to climate change and these politicians in this medium.

This study has the potential to benefit policy makers, non-governmental organizations, activists or journalists by providing an insight into the public debate about climate change in Social Media. Additionally, academics could benefit by an enlarged knowledge about the quantitative impact of some populist politicians along different countries within a specific platform. Moreover, by focusing on and quantifying geolocated tweets we provide additional hints for further research aiming to target specific regions.

2.   Method

2.1.   Data preparation

Given the Twitter API limitations, we decided to use an existing large collection of tweets built by The Internet Archive. The Twitter Stream Grab[1], contains more than 10 billion tweets mined from the Twitter Stream API. This API delivers a 1% random sample of their public daily global tweets[2]. It is available as thousands of TAR and JSON files that contain the tweets accounting for approximately 4TB of data. We developed multiple Python scripts to mine all the tweets available in the collection at that moment. Then we extracted and further processed only those posts geolocated. For one part of the geolocated tweets only a bounding box was available, thus a centroid was computed in such cases to obtain a point location.

2.2.   Tweet filtering by terms

For filtering tweets related to "climate change" or "global warming" in our dataset, the terms were used translated to 14 languages. These include the top 13 languages in number of tweets in our dataset. In a second analysis

[1] https://archive.org/details/twitterstream
[2] https://developer.twitter.com/

phase, four world leaders considered populists by multiple sources[3] were selected. These politicians were presidents of their countries during the development of this work: Donald Trump, Vladimir Putin, Jair Bolsonaro and Recep Tayyip Erdoğan. These four figures were chosen because of their controversial public profiles, and the potential for global and regional analysis in the platform.

2.3. Sentiment Analysis

In this phase, we perform sentiment analysis on the filtered tweets in order to estimate the polarity of the posts generated by the users, analyse its temporal evolution and explore possible relations with real world events. The sentiment analysis determines sentiment orientation and classifies the tweets into classes of polarity: positive, neutral or negative. For the analysis, we use the VADER model (Hutto & Gilbert, 2014).

2.4. Topics extraction and visualization

We will follow a simple approach using NLP in order to analyse and visualize the most frequent words in the tweets. We can consider these words topics of interest (TOIs). For our objectives, we use a script in R based on the packages tm and igraph. The script cleans and counts the frequency of appearance of the words. First, the text is cleaned. Then, a corpus is built with all the cleaned words and a term-document matrix is generated considering each tweet a document. Those words with a frequency of appearance above a threshold are selected. Finally, graphs are created with nodes and edges representing the terms and their relations. Graphic elements are proportional to the total frequency of appearance and co-appearance.

3. Preliminary results

We present here a sample of some current results. *Figure 1* shows the amount of tweets in our dataset by country of posting for the top countries. In *table 1* is shown the amount of tweets generated in the top 24 countries and which share it represents from the total dataset.

---

[3] https://ips-dc.org/the-rising-tide-of-the-populist-right/
https://www.diepresse.com/5695513/trump-bolsonaro-erdogan-die-stunde-der-populisten
https://www.theguardian.com/commentisfree/2020/apr/26/trump-to-erdogan-men-who-behave-badly-make-worst-leaders-pandemic-covid-19
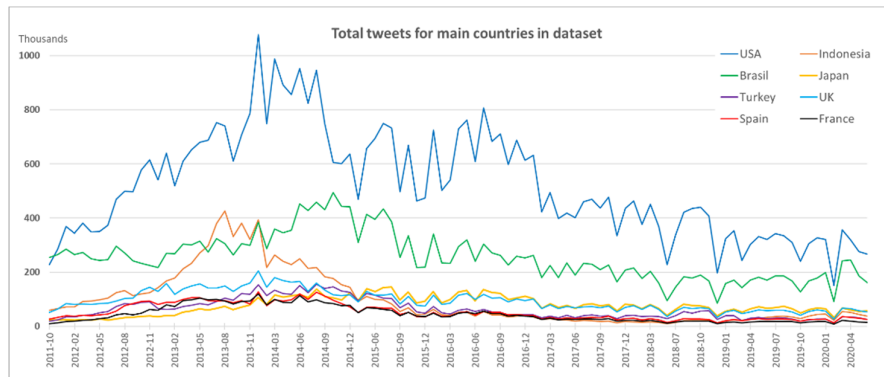
Figure 1. Evolution of tweets from top 8 countries in the dataset.

	Country	Tweets	% of total		Country	Tweets	% of total
1	USA	51.819.811	30,34	13	Russia	2.545.795	1,49
2	Brazil	25.592.957	14,98	14	Canada	2.172.296	1,27
3	Indonesia	9.666.053	5,66	15	Italy	1.720.884	1,01
4	UK	9.444.855	5,53	16	Saudi Arabia	1.668.118	0,98
5	Japan	7.622.546	4,46	17	Thailand	1.554.671	0,91
6	Turkey	6.275.875	3,67	18	Chile	1.494.398	0,87
7	Argentina	6.244.591	3,66	19	Colombia	1.471.555	0,86
8	Spain	5.204.510	3,05	20	India	1.456.484	0,85
9	France	4.415.451	2,58	21	The Netherlands	1.242.471	0,73
10	Malaysia	4.276.622	2,50	22	South Africa	1.113.650	0,65
11	Mexico	3.791.270	2,22	23	Portugal	1.070.508	0,63
12	Philippines	3.702.589	2,17	24	Germany	865.517	0,51
					TOTAL dataset	**170.813.099**	**100**

Table 1. Number of tweets in the dataset per country of generation.

Figure 2 compares the evolution of all tweets produced worldwide (blue line) in our dataset with the proportion of those mentioning *Trump* (red line).
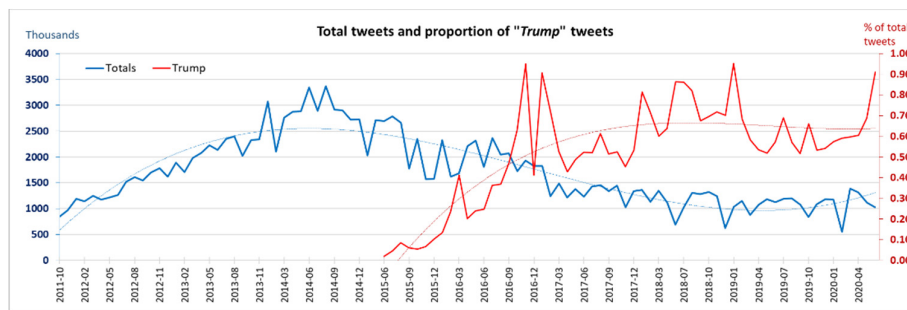

Figure 2. Share of global tweets mentioning Trump vs. total.

Figure 3 compares the evolution of geolocated tweets produced worldwide with the proportion of those mentioning *climate change.*
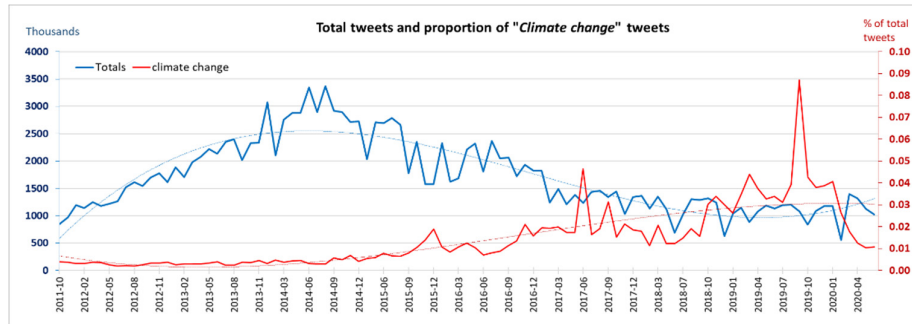

Figure 3. Share of global tweets mentioning Climate Change vs. total.

In *figure 4* we can see the evolution of the polarity of the *Trump* tweets in English after performing sentiment analysis. There is an overall positive polarity until May 2016 when he became his party's nominee. After this event, most of the months present a slightly negative mean polarity.
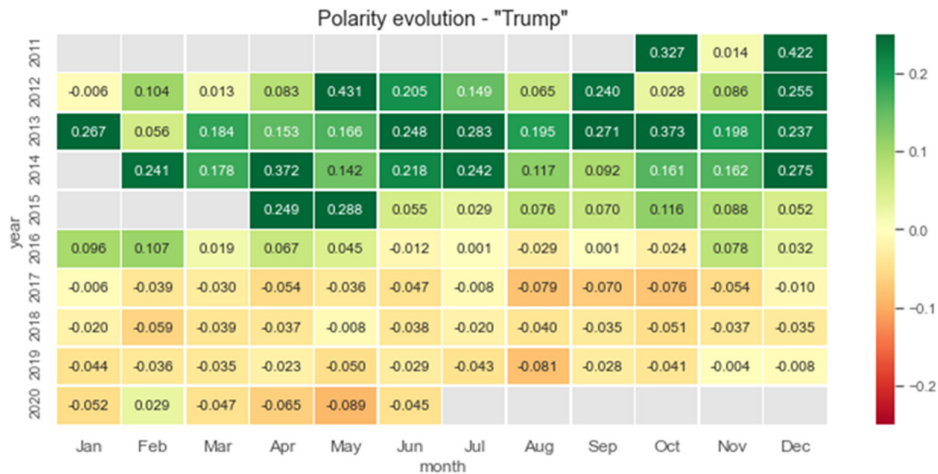

Figure 4. Polarity of tweets mentioning "trump".

Figure 5 compares the share of tweets for *Trump* and *climate change* in several countries. Distance between *climate change* and *Trump's* curves is lower in the European countries considered, compared to the USA and its neighbour Canada.
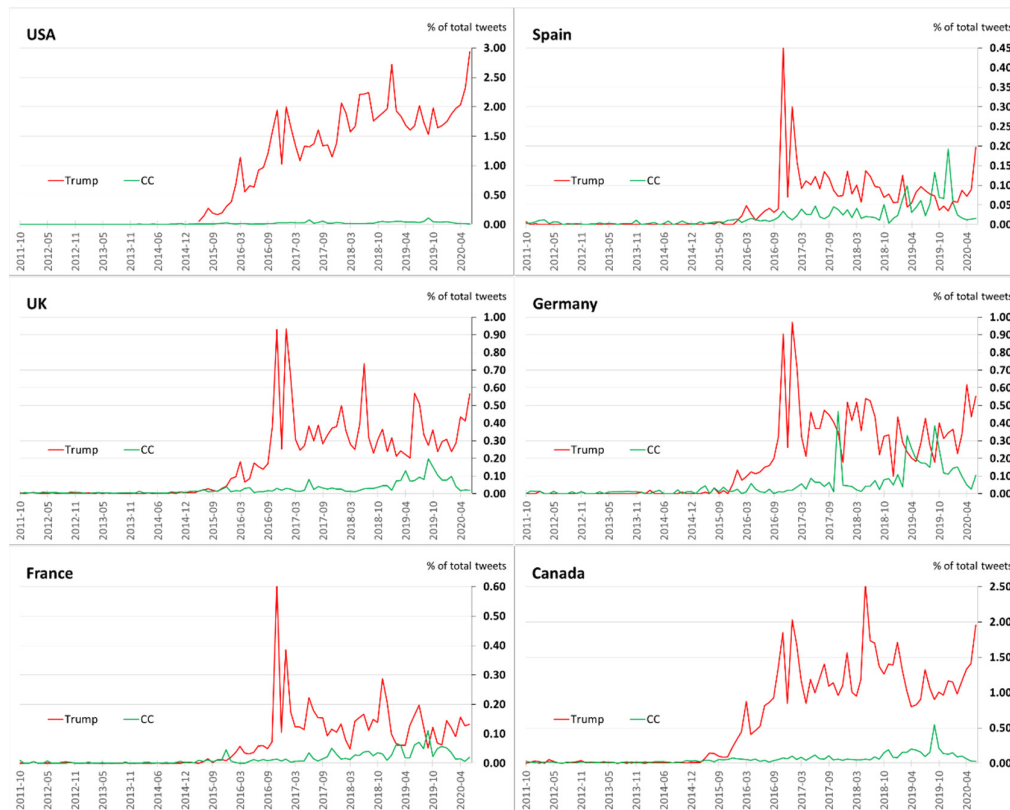

Figure 5. Share of tweets in different countries mentioning Trump vs. climate change.

Figure 6 shows in logarithmic scale the evolution of each group of tweets. Trump's global presence is outstanding. *Erdogan* and *Putin* show a similar impact during time, whereas *Bolsonaro* evolves from a modest global presence to overpass both. At a global level, all of the politicians have had a much bigger direct presence in Twitter than climate change.
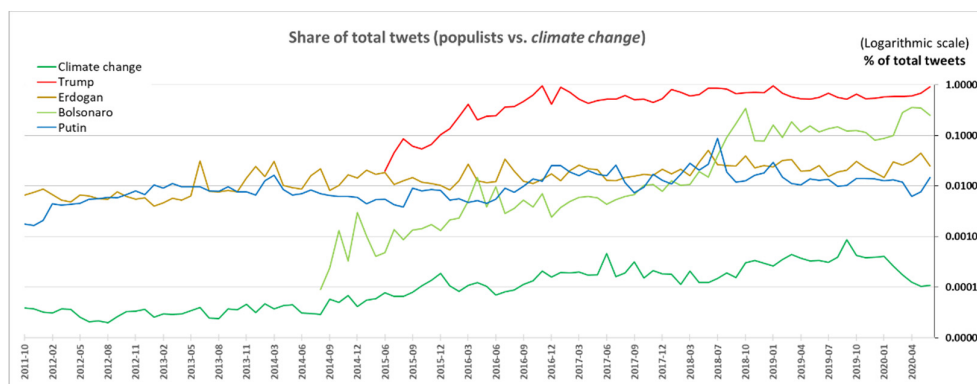
Figure 6. Global tweets dealing with climate change and selected politicians

4. Conclusion

The selected politicians are far more present in the platform than *climate change. Climate change* is comparatively more mentioned in the European countries considered, even overpassing *Trump* in short periods. We are mining a big and unique geolocated dataset that allows for a large data analysis spatially and temporally. Once completed, this work might allow us to provide a better insight into the geolocated conversation in the platform.

References

Bennett, J., Rachunok, B., Flage, R., & Nateghi, R. (2021). Mapping climate discourse to climate opinion: An approach for augmenting surveys with social media to enhance understandings of climate opinion in the United States. *PLOS ONE*, *16*(1), e0245319. https://doi.org/10/gmpv2q

Chen, X., Zou, L., & Zhao, B. (2019). Detecting Climate Change Deniers on Twitter Using a Deep Neural Network. *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, 204–210. https://doi.org/10/gmpv68

Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll. *PloS One*, *10*(8), e0136092. https://doi.org/10/gj2w3k

Doğu, B. (2019). Environment as Politics: Framing the Cerattepe Protest in Twitter. *Environmental Communication*, *13*(5), 617–632. https://doi.org/10/gmn6rg

Ebrey, R., Hall, S., & Willis, R. (2020). Is Twitter Indicating a Change in MP's Views on Climate Change? *Sustainability*, *12*(24), 10334. https://doi.org/10/gmpv69

Hutto, C. J., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. 10.

Milleville, K., Ali, D., Porras-Bernardez, F., Verstockt, S., Van de Weghe, N., & Gartner, G. (2019). *WordCrowd: A location-based application to explore the city based on geo-social media and semantics*. 231–236.

Pathak, N., Henry, M. J., & Volkova, S. (2017). *Understanding Social Media's Take on Climate Change through Large-Scale Analysis of Targeted Opinions and Emotions*. 2017 AAAI Spring Symposium Series.

Sanford, M., Painter, J., Yasseri, T., & Lorimer, J. (2021). Controversy around climate change reports: A case study of Twitter responses to the 2019 IPCC report on land. *Climatic Change*, *167*(3), 1–25.

Singleton, S., Kumar, S. A. P., & Li, Z. (2018). Twitter Analytics-Based Assessment: Are the United States Coastal Regions Prepared for Climate Change*f*. *2018 IEEE International Symposium on Technology and Society (ISTAS)*, 150–155. https://doi.org/10/gmpv2k

Tasse, D., Liu, Z., Sciuto, A., & Hong, J. (2017). State of the Geotags: Motivations and Recent Changes. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 250–259.

Veltri, G. A., & Atanasova, D. (2017). Climate change on Twitter: Content, media ecology and information sharing behaviour. *Public Understanding of Science*, *26*(6), 721–737. https://doi.org/10/gf6dqd

Yaqub, U., Chun, S. A., Atluri, V., & Vaidya, J. (2017). Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Government Information Quarterly*, *34*(4), 613–626. https://doi.org/10/gcqjp8