Climate change and public opinion: demonstrations contributed to spark awareness among Twitter users.

Michele De Marinis

Abstract

This paper aims to assess the impact that movements protesting about lack of political action against climate change have had on the public debate in recent years, specifically on Twitter. In order to do so, this data analysis takes into consideration tweets (i.e. comments made on Twitter) from 2013 to 2021 and looks at the volume of tweets with particular keywords on a specific time interval; furthermore it searches for a correlation between the amount of tweets referring to climate protesters and tweets talking about climate issues, to see if these movements have contributed to increase public awareness on such topics. Data show that the volume of climate related tweets increased significantly after the beginning of strikes and demonstrations in 2018; moreover, the analysis reveals a correlation between the amount of tweets referring to climate change and those referring to climate protests.

Introduction

Climate change - and in particular global warming - is probably the biggest challenge mankind has to face in this present time and in the next decades. Many people reckon that governments, institutions and companies have failed so far to acknowledge the problem and to implement the necessary measures to solve it. For this reason, several protests and demonstrations have taken place in recent years, especially after the Swedish schoolkid Greta Thunberg started her *Skolstrejk för klimatet* (Swedish for *skipping school for climate*), in Stockholm at the end of august in 2018, and many followed her example. In the latter months of 2018 and for the whole length of the following year, the *Fridays for Future* protests (named after the weekday in which the strikes take place) have started spreading around the globe pushing milions of people to take to the streets [1].

The goal of this study is to assess whether these protests have produced the hoped-for outcome, which is both sparking interest and awareness on such issues and lobbying people in charge to take action. In order to do so, this analysis uses Twitter as a sample of the general public, basing the validity of this assumption on the fact that Twitter is one of the largest social network in the world with hundreds of millions of active users. Yet, it is important to remember that comments made on social media cannot reproduce public opinion entirely and therefore it is key to look at the results in view of this consideration.

Methods

To scrape tweets from Twitter I used an open-source Python tool called *snscrape* [1], which collects all tweets that satisfy a particular query. Both data and code written for this analysis are shared on a public repository [2].

The first data set includes all tweets written in English from 2013-01-01 to 2021-12-31, containing the keyword 'climate change'. Moreover, only tweets with a minimum of 100 retweets have been scraped, for two reasons: 1) to be sure tweets had reached some engagement (and therefore were more representative of the

[1] a chronological list of school strikes that have taken place so far can be found here: `https://en.wikipedia.org/wiki/List_of_school_climate_strikes`

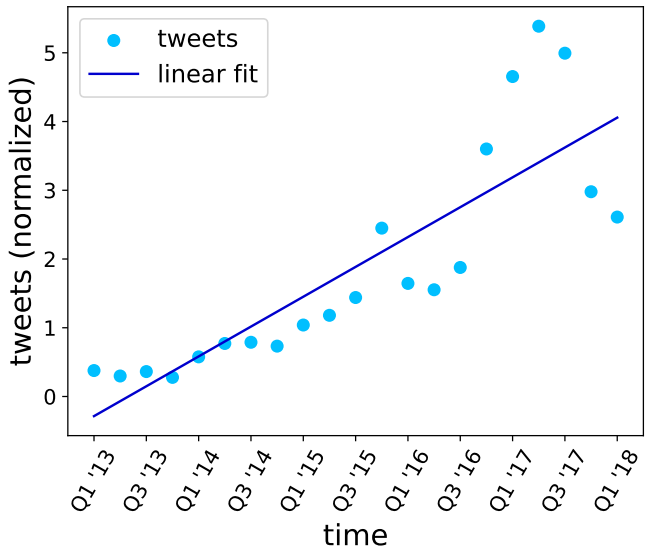

Figure 1: Number of tweets referring to climate change in each quarter. Data have been normalized considering the number of Twitter active users, so numbers on y-axis represent number of tweets divided by hundreds of milions of users and therefore should be read as $\frac{tweets}{users} \times 10^{-6}$. The slope of the trend is $a = 0.21710$ ($\chi^2 = 1.55068$ with $P = 1.00000$, $KS = 0.19048$ with $P = 0.85308$).

general sentiment) and 2) to make the computational problem executable in a reasonably short amount of time. This first data set includes 49048 tweets. That being done, data have been divided into quarters in order to count how many tweets were published in each quarter. This number was later normalized by the number of active Twitter users in that period [3]. Then the analysis took into account data before the third quarter of 2018 (namely when Greta Thunberg started striking) in order to look for a linear trend. This trend is shown in 1 and can be considered as a null model that describes how the volume of tweets referring to climate change evolves throughout the years.

In order to assess if there is a correlation between the number of tweets talking about climate issues and tweets talking about environmental demonstrators this analysis uses four different data sets, each one made of tweets from 2018-09-01 to 2019-12-31 (that is when protests spiked):

1. tweets talking about climate change without mentioning Greta Thunberg nor the Fridays for Future movement;
2. tweets talking about global warming with the same conditions as in 1.;
3. tweets talking about Greta Thunberg without mentioning climate change nor global warming;
4. tweets talking about Fridays for Future with the same conditions as in 3.

Exact queries are specified in [2]. Queries have been set this way in order to consider completely separate data sets, without any element in common; and also because my interest is to see if talking *only* about the demonstrations correlates with discussions focusing *exclusively* on climate issues. Additionally, this way, there is a higher chance of avoiding tweets that report news or statements which are not representative of the public sentiment. Then, tweets in each data set were grouped by a two-week-long period and, just as done before, elements in each sub-period were counted. A two-week-long period was chosen because it is a reasonable length of time of a public discussion held on social media, for instance: on the occasion of a global strike, presumably, people will start talking about it some days before it and will keep commenting it for the following week.

Results

Volume of tweets written after Greta Thunberg started striking do not follow the expected trend based on data from 2013 to 2018, as it is shown in 2, in particular more tweets with climate change related content have been written after this event.

The correlation analysis shows that there is a moderate-strong correlation between the keyword 'climate change' and both the keywords 'Greta Thunberg' and 'Fridays for Future'. On the other hand, there seems to be no correlation between the latter keywords and 'global warming'. Results are shown in 3.

Discussion

Data show that there is definitely a correlation between the spreading of protests and demonstrations and the volume of tweets referring to

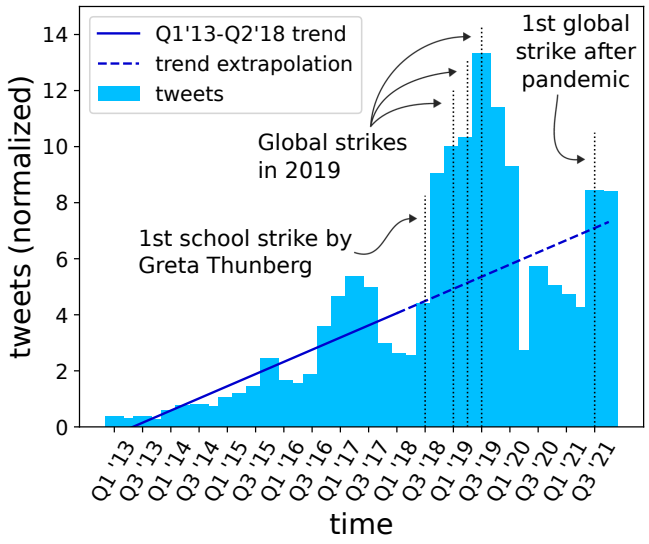

Figure 2: **Data after Q3 '18 do not fit into the extrapolated trend line**, rejecting the null model, as the Kolmogorov-Smirnov test returned $P = 0.0350$ which is below the acceptance limit of 0.05. Moreover, volume of tweets talking about climate change seems to be higher when major demonstration events take place.

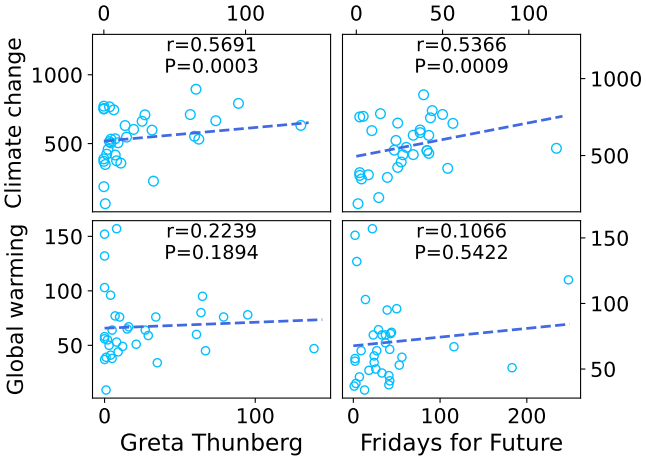

Figure 3: **Tweets talking about climate change correlate with tweets talking about Greta Thunberg and Fridays for Future demonstrators**, the correlation does not hold for tweets talking about global warming. (P is the $p-value$ for the null hypothesis that no correlation exists and r is the *Pearson correlation coefficient*).

climate change. This was shown both analysing the volume of tweets before and after the start of the protests and carrying out a correlation test between specific terms. Of course, this is not enough to establish a causation model, meaning that it is not possible to state that *because of* the protests people have started talking more often about climate on Twitter. Nevertheless, by looking at 2, we can see that the number of tweets spiked when major demonstrations took place, and even after the pandemic - during which, presumably, people were concerned about other issues - volume of tweets increased again near a global strike.

Also, it is hard to tell why the keyword 'global warming' does not correlate with the keywords 'Greta Thunberg' and 'Fridays for Future' whilst the keyword 'climate change' does. A further analysis on how these two terms are perceived and used might be helpful to answer this question.

Lastly, it is important to notice that this data analysis is focused only on the *volume* of tweets and therefore it does not answer the question on *how* protesters contributed to spark interest among the general public. That is to say, it is not possible, from this analysis, to assess people's sentiment about climate change and global warming after the protests and in what terms they talked about it on Twitter.

References

[1] `https://github.com/JustAnotherArchivist/snscrape`, accessed 10 april 2023.

[2] `https://github.com/slimmythepig/dc1`.

[3] Data about active Twitter users were found here (2013-2018): `https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/`, and here (2019-2021): `https://www.statista.com/statistics/303681/twitter-users-worldwide/`.