CHARLES UNIVERSITY

Faculty of Social Sciences

Institute of Sociological Studies

Department of Sociology

Master's Thesis

2020

Bc. et Bc. Jana Bicanová

CHARLES UNIVERSITY

Faculty of Social Sciences

Institute of Sociological Studies

Department of Sociology

The Effect of Scientific Argumentation on Climate Activism on Twitter

Master's Thesis

Author: Bc. et Bc. Jana Bicanová

Study programme: Sociology

Supervisor: Mgr. Jan Urban, Ph.D.

Year of the defence: 2020

Bibliographic Record

Bicanová, J. *The Effect of Scientific Argumentation on Climate Activism on Twitter.* Master's Thesis. Prague: Charles University.

Length of the thesis: 91,034 characters including spaces

Keywords

Climate activism, social movements, global warming, climate change, Twitter

Klíčová slova

Klimatický aktivismus, sociální hnutí, globální oteplování, klimatická změna, Twitter

Title

The Effect of Scientific Argumentation on Climate Activism on Twitter

Název práce

Vliv užití vědecké argumentace na klimatický aktivismu na Twitteru

Abstract

The aim of this thesis is to test the main principle of the Gateway Belief Model (GBM) on Twitter data, as suggested and experimentally validated by other authors. van der Linden et al. (2015 and 2019). The GBM predicts that the perception of scientific consensus on anthropogenic climate change increases the probability of support for public action against or in favor of the mitigation of the climate change. In this work I analyse a random sample of 115,940,434 tweets gathered over the course of the first six months of 2020. The big data is pre-processed using unsupervised (Latent Dirichlet Allocation) and supervised (Naïve Bayes Classifier) machine learning algorithms in order to generate keywords for filtering environmentally themed tweets and to classify either absence or presence of the climate activism. Within the dataset, 5,857 environmentally themed tweets were detected, finding that only 94 out of them were explicitly linked to the message of scientific consensus about anthropogenic climate change. The harvested dataset proved to be unsuitable for testing the GBM, not only because of the small number of tweets which contain the message about 97 % of climatologists reaching the consensus, but also because the majority of these tweets deny the consensus and therefore, do not represent a perception of it. Considering these unanticipated circumstances, the research aim was modified and the principle of the GBM was tested on a more general level. Instead of the scientific consensus, any science-related content of environmentally themed tweets was studied. This allows me to suggest that during the studied period, the twitterverse does not behave according to the GBM. On the contrary, my results reveal that chance the environmentally themed tweets using the scientific argument are six times higher not to be an expression of the climate activism.

Abstrakt

Cílem této práce je na datech z Twitteru otestovat hlavní princip Gateway Belief Model (GBM) tak, jak navrhli a experimentálně ověřili již jiní autoři. van der Linden a kol. (2015 a 2019). GBM předpokládá, že vnímání vědeckého konsensu o antropogenní změně klimatu zvyšuje pravděpodobnost veřejného zájmu a akce proti nebo ve prospěch zmírněného tématu změny klimatu. V této práci analyzuji náhodný vzorek 115 940 434 tweetů stažených v průběhu prvních šesti měsíců roku 2020. Tato data jsou předběžně zpracována typem strojového učení bez učitele (Latent Dirichlet Allocation) a následně strojovým učením s učitelem (Naïve Bayes Classifier) tak, aby byla vygenerována klíčová slova pro získání pouze klimaticky tematizovaných tweetů ze všech vytěžených dat. A dále aby bylo možné klasifikovat buď přítomnost, nebo nepřítomnost klimatického aktivismu. V rámci datového souboru bylo zjištěno 5 857 ekologicky tematických tweetů, z nichž pouze 94 bylo explicitně spojeno s myšlenkou vědeckého konsensu o antropogenní změně klimatu. Získaný datový soubor se tedy ukázal jako nevhodný pro testování GBM, a to nejen kvůli malému počtu tweetů obsahujících zprávu o tom, že 97 % klimatologů dosáhlo konsensu, ale také proto, že většina tweetů obsahujících tuto zprávu konsenzus zpochybňuje. A proto nepředstavují nezávisle proměnnou z GBM. Ve světle těchto neočekávaných okolností byl výzkumný cíl pozměněn a princip GBM byl testován na obecnější úrovni. Místo vědeckého konsensu byl studován veškerý obsah vědeckých tematických tweetů. To mi umožňuje navrhnout, aby se během studovaného období choval twitterverse podle GBM. Naopak, mé výsledky ukazují, že klimatické tweety, které používají vědecký argument, mají šestkrát vyšší šanci, že nebudou výrazem klimatického aktivismu.

Declaration

1. I hereby declare that I have compiled this thesis using the listed literature and resources only.

2. I hereby declare that my thesis has not been used to gain any other academic title.

3. I fully agree to my work being used for study and scientific purposes.

In Prague on 31. 07. 2020 Bc. et Bc. Jana Bicanová

Contents

Introduction

The high amount of information available on Twitter means additional data source for the social scientific discipline. Despite the different character of the Twitter data compared to the conventional research approaches, previous research (Hodges, Stocking 2016) proves that the Twitter data can be utilized for the testing and development of existing social scientific theories. This innovative approach is challenging because of the fact that conventional social scientific theories use operational definitions, measures, and analytical procedures that are not easily applicable to big data from social networks given sheer volume, highly structured content, and incompleteness of such big data (Salganik 2018, p. 24).

In this work, I will explore how Twitter data can be used to gain insight into factors of climate activism. The main reasons for this topic are, firstly, to theoretically grounded and to test a particular social theory on big data. Secondly, concretely Twitter has been chosen as it is one of the mostly used platforms among the social networks (Sharma et al. 2019) and the social media plays a vital role in influencing peoples´ daily decisions regarding a great variety of topics (Sharma et al. 2019). Finally, Twitter data is a promising source for social scientific research hides behind its nonreactive nature (Salganik 2018, p. 23).

The main contribution of this research lies in its aim to be theoretically grounded and to test a particular social theory on big data. For my work, I have chosen the theoretical framework of Gateway Belief Model (GBM, van der Linden et al., 2015; van der Linden et al., 2019), that is an important theory about antecedents of climate activism. So far, the validity of the GBM has been tested in randomized experiments carried out on individuals; this theory has never been tested on Twitter data. On top of that, the authors of the theoretical model themselves, while suggesting the direction for future research after a large-scale replication, recommend validating the GBM using real-world data (van der Linden et al., 2019). Twitter data can potentially provide such empirical data on real-life interactions.

The Twitter platform deserves the attention of social scientific researchers because social media plays a vital role in influencing peoples´ daily decisions in political, social and economic domains and Twitter is one of the mostly used platforms among the social networks (Sharma et al. 2019). The population of Twitter users sharing tweets in English will be studied because influencers using Twitter choose this language to target as many people (i.e. Twitter users) as possible. The English language is accessible to most of the people, thus discussions in English are the most influential ones (Kirilenko, Stepchenkova 2014).

Another reason why Twitter data is a promising source for social scientific research hides behind its nonreactive nature (Salganik 2018, 23). Researchers using conventional social scientific approaches usually must deal with the fact that people can change their behavior, when they are being observed, interviewed, etc. by researchers (Webb et al. 1966). Nonreactivity is advantageous for research although it does not ensure that all the Twitter users are always free of social desirability bias, the tendency of people to present themselves in the best possible way. However, Twitter data is not better than traditional data sources of social scientific research in every single aspect.

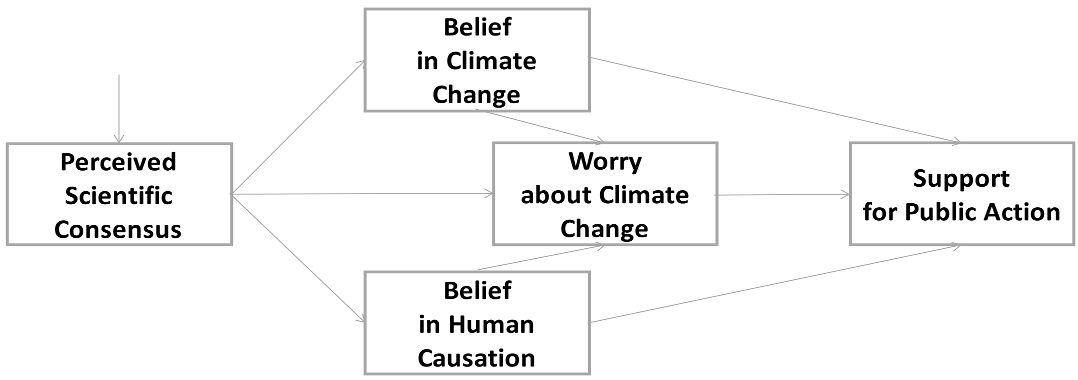
The thesis will first present theoretical background based on the Gateway Belief Model and the conceptualization of the model´s dependent variable - support for public action. The climate activism will be elaborated with regards to its possible forms in the Twitter environment. Some implications of the specifics of the Twitter platform for performance of internet political activism will be discussed as well. The second section will describe the methods applied on the harvested dataset. The chosen methods were derived from the specifics of the big data, or more precisely the Twitter data. So, the particularities of the studied data, its advantages and disadvantages it poses for the presented research will be discussed. The succeeding part will present the key findings including the difficulties the research had to face in the light of global events of the first six months when the coronavirus pandemic paralyzed most of the world. The last section will discuss the findings and suggest explanations for the difficulties the research faced as well as possible directions of the future research.

Theoretical Background

The GBM suggests that perception of a domain-specific un/certainty is an important heuristic that informs one´s personal views. According to GBM, the perception of scientific agreement is a key factor of social activism related to climate change. More specifically, the GBM (see *Figure 1: Gateway Belief Model (GBM)*) predicts that scientific consensus affects policy support related to climate change mitigation through several mediators e.g., beliefs about the existence of global climate change and worry about global climate change (van der Linden et al., 2015).

The model describes a process of judgment and attitude change. The first stage

Figure 1: Gateway Belief Model (GBM)



involves a change in perceived scientific consensus which predicts cascading changes in other key beliefs about the issue. A change in perceived scientific consensus acts as a “gateway” in the sense that it predicts changes in personal beliefs and attitudes about climate change. Change in these central beliefs predicts support for public action. The influence of perceived scientific consensus on support for public action emerges indirectly in the GBM (van der Linden et al., 2015) but considering the complexity of Twitter data and limitations of this research, only the main simplified idea of the GBM is tested. Thus, the key interest of the presented work lays in the correlation between perceived scientific consensus and support for public action, while the model mediators (i.e. beliefs about existence of global climate change and worry about global climate change) are omitted. In order to appropriately detect all the model variables and the causality between them, controlled intervention in the form of consensus messaging would be required. If this work proved the main principle of the GBM to be valid within the twitterverse[1], it will justify further research effort in this direction.

Perceived Scientific Consensus

Informing about scientific consensus is a non-persuasive communication, which means it only conveys the consensus that most climate scientists have concluded that humans are causing global warming but does not directly advocate solutions or policy-support.

Support for Public Action

For purposes of this work, the act of support for public action is conceptualized as attempts to educate the public through information provision, promote fundraising, encourage direct political action, promote (individualized) expression of climate action, and attract followers of the same ideology. Requesting an action or asking for a donation can be described as a traditional strategy, which has not been fundamentally altered by the emergence of modern technology and social media. Although there have been hopeful voices labelling some events that brought social changes (e.g. Arab Spring uprisings) as Twitter Revolutions, others argue that the ability to mobilise the public by requesting action or donations has not increased thanks to social media. The doubts about social media being a game-changer that provides a magical key how to get the indifferent public sincerely involved in social movements are expressed well by Rosen (2011): "*Revolutions happen when they happen. Whatever means are lying around will get used.*" And indeed, social media can be an important arena for activism because it can help with logistical support or information distribution (Rosen in Murphy, 2018).

Important limitation of the internet activism in democratic states is its low-stakes involvement and minimal demands on one´s activity. Looking specifically on Twitter the vast majority of users might only retweet an outrageous content and can be seen as a global individual or an empathetic cosmopolitan who is immersed in public good afterwards, but it needs to be emphasised that it takes just a second to (re)tweet something. Rather than deep engagement, these masses of users are only superficially engaged. As long as the user stays passive in the comfort of his or her chair, it is nothing more than the "slacktivism"; other labels

[1] Twitterverse refers to all Twitter users. The term captures all members of the online social media network.

used for such low-stakes activism include “latte activists,” or “armchair warriors” (Murphy, 2018, pp. 90-91).

Online Activism Based on Informational Tweets and Citizen Journalism

In contrast to action or donation requests, the remaining dimensions of climate activism are directly linked to an innovative character of social networking and other modern communication technologies. Early theorists proposed that information propagation would be a primary function of online activism (Denning 2001). Social media in general facilitates a new possibility how of make one’s voice heard without traditional news; Twitter, in particular, is known for being the place, where news-sharing is happening.

Empirical expertise by Mark Boukes, for example, shows that in contrast to the frequent use of Facebook, more frequent usage of Twitter positively affects the acquisition of current affairs knowledge (Boukes, 2019, pp. 12-13). The specifics of Twitter come from its dynamics based on hashtag categories that aggregate any tweet with the same hashtag into a communal meta-thread (Murphy, 2018, p. 67). Some of these communal meta-threads become *trending hashtags* and represent what Twitter users consider worthy of attention. Qualitative research by Boczkowski, Mitchelstein and Matassi (2018) reveals the common practice of Twitter users who contribute to the self-fulfilling prophecy of trending topics, when checking what is trending because “*everybody is tweeting about it*” (Boczkowski, Mitchelstein and Matassi, 2018, p. 3531).

The hashtag-based communal news space facilitates a phenomenon that can be described as “citizen journalism”. There are many examples (e.g.:, The Miracle on the Hudson; see Murphy, 2018, p. 63-65), when ordinary citizens happened to participate at or witness an important event and by tweeting about it in real time, their tweets went viral. The role of traditional news has been modified because its former position of gatekeepers, who decide what story is worth of covering, has been violated by citizen journalists (Allcott and Gentzkow, 2017, p. 214). The ability of citizen journalists experiencing an issue to be right in the heart of the situation, had to be incorporated in the new model of journalism. Sometimes the press may not be able to mobilize its resources quickly enough and without the citizen journalists the story might be missed. It has become an ordinary journalistic procedure of traditional media itself to recognize influential Twitter news content (Murphy, 2018, p. 51-69). The Twitter platform and its users create signals about agendas, frames, opinion, and behavior that are interpreted by political elites, news business, and publics themselves (Bimber and Gil de Zúñiga, 2020, p. 708) and thus became an important component of the public sphere.

The "citizen journalism" has potentially advantageous democratizing capacity and can give voice to the unheard and a pure grassroot agenda setting can be facilitated. The t*rending topics* tend to be a diverse mix of the profound and the banal, which proves its organic origins. And Dhiraj Murphy in his book *Twitter: social communication in the twitter age* claims that Twitter has been able to "*retain more grassroots/bottom-up feel as "native content" than other platforms (most notably, Facebook)*" (Murphy, 2018, p. 160).

Online Activism Based on Promotional or Ideology Reinforcing Tweets

Twitter offers a tool that helps its users to participate in influential debates. The platform was at its beginnings unique social networking environment, because it allows users to interact with others with whom they are not formally connected, not only thanks to hashtags as discussed above but also through replies and mentions. And thus, possibly reach far greater audience while trying to promote environmental activities, worldviews, ideologies, etc. Tweets tweeted for the sake of climate action promotion and ideology/identity reinforcement are included in the concept of online engagement in the line with research by Hodges and Stocking (2016).

Number of diverse researches studied unconventional forms of political activities (for example Harris 2008; O'Loughlin and Gillespie 2012) and proves that it is needed to broaden what is understood by the term of political engagement. Communities who face diminished prospects for effective participation in formal political processes tend to seek alternative means and new political environments, where the participatory practice is established on their own terms. The newly emerging action repertoires are characterized by its extra-parliamentary realm, non-hierarchical and informal networks, and in a variety of sporadic campaigns that are not institutionalized (Stolle and Hooghe, 2011, p. 120).

Considering specifically the environmental issue, the youth is to be recognized as a marginalized group seeking for alternative political arenas. Prospects of the deepening climate emergency potentially cause a sense of despair and feelings of helplessness because of the complex character of required solutions or mitigations. Since the policies and decisions made today will influence outcomes over the remainder of this century and beyond, the younger generations are highly involved (Patridge, 2008). Although the global-scale protests of Fridays For Future happening during 2019 lead by the youth and inspired by Greta Thumberg have gained attention and secured the voice of the youth in the public environmental discussion, the agency of the youth keeps facing doubts and tries to delegitimize their readiness by arguments like: '*They are only children who should be focused on their education, what do they know about life?*' (Petrov 2019) and most importantly they are not acknowledged as an autonomous political actor with a meaningful participation in the current decision-making process (Hart 2008, Checkoway 2011, Taft and Gordon 2013).

The political activism reacting to the climate emergency has been manifested not only by organizing awareness-raising events, sustainability campaigns or volunteering in organizations such as 350.org, Global Power Shift, Friends of the Earth, Gen Zero, Climate Youth (O'Brien, Selboe and Hayward, 2018, p. 42) but also by an individualized mode of political action. Promoting recycling, consuming more environmentally friendly products, following vegetarian or vegan diets are all individual activities, but they still deserve to be acknowledged as a full-bodied part of environmental social movement. The individualized mode of climate activism is another valid strategy how the youth politically expresses themselves by politicizing morality and everyday life (Manning, 2012).

Environmental activism can be also based on strategies challenging business-as-usual economic and its emphasis on economic growth to shift political and economic power away from the fossil fuel industries and carbon polluters. Expressions of these strategies (particularly popular with young people who commit to climate change) include for example decision not to get driver´s licenses (Sivak and Schoettle 2016) or to boycott products harmful to the environment.

Some of the above-mentioned expressions of dissent may not be primarily perceived as a threat to the status quo but they can have an impact eventually. The popularity of sufficiency in everyday life makes the youth pioneers of a consistently ecological and sustainable lifestyle (Aljets and Ebinger, 2016, p. 6). Practices such as energy saving, sharing, gifting, and rejecting packaging are examples of how people can manifest their political positions without directly participating in politics. These activities and their public promotion potentially alter the framing of what is perceived as 'cool' towards the direction of climate activists´ political aims and potentially lead to a grassroots support of environmentalism. The dimensions of internet activism, conceptualized as the promotion of environmental activities and ideology/identity reinforcement, reveals the individualized mode of political action. In this sense, this research recognizes the limited resources and opportunities of (particularly young) climate activists who can only raise their voices using strategies established on their own terms such as divestment campaigns, boycotts or politicizing morality and everyday life .

Methods

Hypotheses

The main idea of the GBM applied on Twitter environment has led to the hypothesis:

H: Tweets interacting with the message of scientific consensus about anthropogenic global warming are more likely to be a performance of environmental activism.

Data

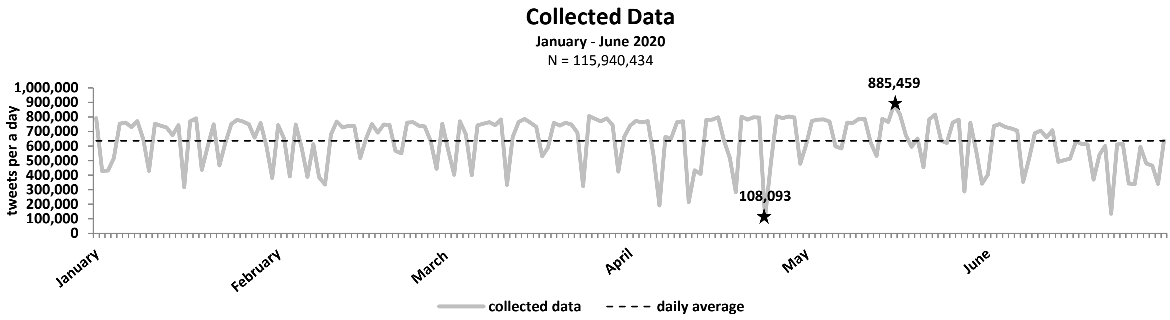
Data Harvesting

The dataset consists of 5,857 environmentally themed tweets from the beginning of January 2020 until the end June 2020, collected by using the Twitter Stream Application Programming Interface (API). The Twitter Stream API gives an access to a random sample representing about 1 % of all tweets posted in the time period, which means the streaming is happening in real time. (Dahal et al. 2019; 24) The Twitter stream has been accessed through a Python package *Tweepy.* Such Twitter harvesting was unstable due to rate limits set by Twitter and other unexpected discontinuations of the data stream.

While listening to Twitter stream, 115,940,434 Tweets in English language was collected. The average and maximum number of tweets streamed per day was 637,035 and 885,459, respectively, with the maximum reached on May 16th as shown in Figure 2: Number of tweets collected per day during the studied period). Although there are some days with extremely low numbers of tweets, in general, the daily stream of tweets was relatively steady. Anomalies such as the minimum number of collected data on April 24th (108,093 tweets), were caused by the Twitter streaming limitations or by technical problems with the internet connection or the electricity.

These deviances must take into consideration at the time of making associations between significant events and the tendencies observed in the dataset which is vital in order to avoid systematic errors and understand what shapes the dynamics of the studied sample. But apart from that, these deviances do not pose any further difficulties for the research. Since the technical issues associated with the tweets streaming occurred on a random basis without any systematic effects, the collected dataset is still a random sample.

Figure 2: Number of tweets collected per day during the studied period



The collected dataset covers approximately one percent of all the tweets shared on Twitter during the first six months of 2020.

Generating Keywords

An unsupervised machine learning topic modelling was applied to generate more nuanced set of keywords than the primal keyword search constituting of expressions "climat", "environ" and "global_warming" and thus filter irrelevant tweets that have been detected by the primal keyword search but have nothing to do with climate change. Extraction of the environmentally themed sample within the collected dataset was based on a list of keywords detected and a list of influential environmental hashtags. Although the list of the most popular hashtags used in tweets containing expressions "climat", "environ" or "global warming"produced helpful insight on how to recognize an environmentally-themed tweet, not all tweets include hashtags. In order to gain dataset without preferring only hashtagged content, Latent Dirichlet allocation (LDA) was applied on tweets containing expressions "climat", "environ" or "global warming" for each month separately. The produced topics were a probabilistic mixture of words that represented word co-occurrence trends in the dataset (Steinskog et al. 2017).

As Steinskog and colleagues (2017) admit: "*statistical methods cannot model a human's perception of the coherence in a topic model perfectly*", in line with common practice human judgement was used to evaluate the generated topics (Steinskog et al. 2017, 79). The process of the topic assessment was supported by the cluster dendograms demonstrating if/how closely the topics are related. The set of topics with the best measure of coherence score was examined in detail. Each of topic constituted of a list up to 20 words and their relevance was assessed based on criterion of exclusively environmental determinant while combined with either an "environ" or a "climat" expression. The unclear cases were decided based on random examples of tweets where the expression occurred. When some of these tweets were discovered to be non-environmentally themed debates, the word was omitted. The outcome of topic modelling revealed the most prevalent set phrases occurring in climatic discourse throughout the course of the first six month of 2020 and time specifics. For example, an expression of Earth Day featured in an April topic modelling output, when the annual event is celebrated.

Thanks to collecting the whole random sample instead of setting the filtering criterion at the very beginning, key words could be selected in the way to make sure they reflect what happened to be the environmental agenda during the studied period. The “always-on” nature of Twitter data enables the research to be flexible and study/work with unanticipated events. Once something relevant happens, it is possible to *travel back in time*, directly observe changes before, during and after that particular event and quantify its effect (Budak, Watts 2015). Thus words representing a significant environmental topic and words often used in environmental debates together with expressions “climat” or “environ” were selected as additional filtering criterion.

The final keyword search criterion defining a tweet to be environmentally themed when at least one out of three criterions was fulfilled. The first part of the filter selected tweets containing a combination of either an expression “climat” or “environ” and one of the listed expressions: “action”, “activis”, “arctic”, “australia”, “bicycle”, “bill nye”, “biodiversity”, “burning”, “bushfires”, “carbon”, “cars”, “catastrophe”, “change”, “clean”, “coal”, “crisis”, “deal”, “denial”, “denier”, “disaster”, “earth”, “emergency”, “emissions”, “environmental day”, “espanto”, “expert”, “fossil fuel”, “friendly”, ”fuel”, “green”, “great”, “hoax”, “hoshi”, “hyundaixbts”, “ice”, “industry”, “inequality”, “justice”, “left”, “licypriya”, “morrison”, “movement”, “nakate”, “narrative”, “nature”, “planes”, “planet”, “plant”, “pollution”, “protect”, “research”, “risk”, “scien”, “solar”, “strike”, “sustainable”, “threat”, “thunberg”, ”trash”, “vanessa nakate”, “vegan”. The second reason why a tweet was detected as a part of environmental debate was when it contained the set phrase “global warming”. The third evidence of the environment topic was when the tweet was tagged with at least one of these hashtags: “#biodiversity”, “#bushfires”, “#climateaction”, “#climatechange”, #climatecriminals”, “#climatecrisis”, “#climatehoax”, “#climatejustice”, “#climateprotest”, “#climatestrike”, “#climatestrikeonline”, “#cop26”, “#digitalstrike”, “#earthday”,

"#extinctionrebellion", "#flattenthecurve", "#fornature", "#fridaysforfuture", "#globalwarming", "#greennewdeal", "#hyundaixbts", "#jakarta", "#kangaroos", "#nature", "#pollution", "#pollution", "#run4climate", "#savecongorainforest", "#science", "#siberia", "#solar", "#sustainability", "#timefornature", "#worldenvironmentday" or "#worldoceansday".

Character of Harvested Data

Data collected from Twitter falls into a category of *observational data* because it was gained by observing a social system without intervening in any way. The creation and original collection of the analysed tweets were initially for purposes other than research. From the perspective of this research the collected random sample of tweets is a *found* dataset that is going to be repurposed (Salganik 2018, 13-14).

The modification of the original purpose requires utilize the advantageous aspects of the data while taking into the consideration features of the data that cannot be influenced during its creation. Carrying out a research on *found* data brings challenges in the form of fact that, the available dataset significantly differs from ideal dataset which would be targeted, while designing a survey. The authors of the investigated tweets cannot be asked any additional questions which means the research needs to settle for a limited amount of information 1) tweet ID; 2) text of the tweet; 3) time when the tweet was shared; 4) author of the tweet´s user name; 5) number of followers the user has; 6) description of the user written by the user and made available on his or her profile, which is meant to introduce the user; 7) user location in the case that the user allowed this piece of information to be shared ; 8) time when the profile of the user was created and number of likes the tweet had by the time of collection which is always zero because the streaming is at the very moment of the sharing act.

Apart from the incompleteness of the data (Salganik 2018, 24), it needs to be approached with regards to who designed Twitter and what are Twitter´s main goals. Its target in the first place is to provide a service to its users and to make a profit (Salganik 2018, 15). As well as other big data systems, Twitter is constantly changing its features how users can use it in order to keep up with its competitors and improve its service. On the top of that, features set by Twitter differs from one country to another (Poblete et al. 2011). It makes it hard to use the data to reliably measure long-term trends, because in order to do that, the measurement system itself must be stable. Once Twitter, for example, introduces a change of tweet length, any longitudinal study of tweet updates will be vulnerable to artefacts caused by this change (Salganik 2018, 33-34).

Nonreactive character of research data collected on Twitter (as discussed above) is not to be confused with data capturing natural interactions between users that has been only amplified and made public. The analysed Twitter users did not simply move their ordinary conversations from an offline environment to the cyber space. The way how Twitter is designed also significantly influences the way how Twitter users themselves behave and what kind of communicative strategies they choose. The ways in which users communicate via social media in general are qualitatively different from traditional face-to-face communication. Looking specifically on Twitter, before tweeting any content the users are likely to think about what hashtags to include with regards to trending topics, or who should be @-mentioned (Murphy, 2018, p. 8).

Users´ intentional adaptation is partly caused by Twitter designers who aim at inducing specific types of behavior such as clicking on ads, retweeting or posting content. The goals of Twitter designers, materialized in algorithms deciding which user sees what content, can also introduce patterns into data and its implications on the data can be described as being *algorithmic confounding* (Chaney et al. 2018). Dealing with algorithmic confounding is particularly difficult because the algorithms are largely invisible and kept secret as Twitter´s know-how. Many features of Twitter are proprietary and poorly documented.

Another challenging feature of Twitter data lies in its often censor-less debates and weakly regulated dynamics. Although this particular feature might be given the credits for the thriving phenomena of citizen journalism (as discussed above) and ability of giving a voice to marginalized groups, on the other hand the liberation of news agenda gatekeeping gives questionable content the opportunity to spread (Vosoughi, Roy and Aral, 2018).The victory of the speed of information dissemination over fact-checking and verification (Murphy, 2018 pp. 70-73) creates perfect circumstances for armies of bots to manipulate the social signals about what people are thinking and talking about, how they are talking about public affairs, what they want, and what they are doing (Bimber and Gil de Zúñiga, 2020, p. 708).

Research has managed to provide a lot of evidence about a massive presence of social bots operating, not only on Twitter but on social media in general, with intentions to undemocratically influence public discussion by creating the wrong impression about social reality (Bradshaw and Howard, 2019, p. 11). The so-called army of bots do not get access to the social media platforms from front-end websites like real-life users usually do, because the activity of bots is driven by a code-to-do connection. The computational tool of botnets has not been designed exclusively for political purposes and has many other possible uses. The word botnet is derived from "robot" and "network" which covers its main idea, since it can be defined as: "*collection of connected computers with programs that communicate across multiple devices to perform some task*" (Howard and Kollanyi, 2016, p. 1). The particular task of the computer-generated programs in the case of bots on Twitter platform would be posting, tweeting, or messaging of their own accord. The design of the coordinated activity aims at creating false signals about public discourse (Bimber and Gil de Zúñiga, 2020, p. 708). For the sake of completeness, similar effects (the creation of impression of organic support for a political candidate, ideology or issue; the appearance of opposition; or for distraction or agenda change) can be achieved by an organized team of human participants as well (Allcott and Gentzkow, 2017, p. 217).

Apart from the external malicious attempt to run misinformation campaigns, the mechanisms applied on Twitter by its developers proved to be ripe conditions for these types of attacks (Howard and Kollanyi, 2016, p. 1). An empirical evidence about what type of content is likely to become the trending topic on Twitter proves that rumours, unverified or untruthful information has more potential for spreading because it attracts attention and produces emotional responses, which are desired traits by Twitter algorithms. Thus, false information is spread better and faster by Twitter algorithms than verified truthful information (Vosoughi, Roy and Aral, 2018).

Twitter data can be loaded with questionable content such as rumours which might lead to emotions like fear and anxiety. Once authoritative source confirmed a statement as false or fabricated, then it is labelled as a *rumour* (Wu et al. 2015). False or inaccurate information, communicated intentionally or unintentionally, with an attempt to present it as being true is *misinformation*. It spreads without an intent to deceive the reader and can be caused by a lack of understanding, attention or misrepresentation of an original piece or true information (Habib et al. 2019). On the contrary *fake news* is written and published with the intent to mislead its readers in order to gain the reader´s attention in bad faith, e.g. with political intentions. *Fake news* is characteristic for its sensationalist, exaggerated, or patently false headlines (Allcott and Gentzkow 2017). The last distinguished possibility of questionable content is called a hoax, which is a false story to deceive the truth, meaning “to cheat” usually spread through the news outlet to gain political and financial benefits. These electronic messages with evil intention to misguide recipients are consist of audio, text and multimedia content. Hoaxes are being spread with political or some other special interests, and due to the lack of media literacy, people are unable to distinguish between real news and hoaxes spread on social media (Hunt 2016).

Pre-processing

Responding to above described dirtiness of Twitter data, bots detection methods was convenient to be applied because otherwise the analysis would treat automated content that makes no sense as something valid and worthy of analysis. With regards to the limits of this research, only basic strategies how to detect automatically produced content on Twitter were applied and thus signs of anomalies were explored. Anomaly means to behave differently from normal behaviour, defined as statistically predictable value. The pattern of normality is in this method base on the content and number of tweets sent by the user and the comments made on the tweet (Liu et al. 2012). Another accessible approach was to focus on aspects related to the moment of account creation – the time of creation and username. The detection scheme to filter potentially automated account groups is based on the assumption that the differences between account names were created algorithmically (Lee, Kim 2014).

Original format of harvested data needed to be cleaned in order to retrieve only the part of variable (labelled as *raw_text*) that is the actual content of the tweet and at the same time not to lose the additional pieces of information contained in the variable. The variable *raw_tweet* can take different standardised formats depending on whether the twitter action is an original tweet, a re-shared tweet (i.e. retweet) or a comment (see Table 1 Standardized format of a retweet and Table 2. Standardized format of a tweet containing mentions and a link).

Table 1 Standardized format of a retweet

id_str	raw_text	created	user_name	fav_count
1224801284819050000	**RT @ConservationOrg:** Scientific communities overwhelmingly agree that the climate is in crisis and we've got 10 years to drastically cut carbon emissions	04/02/2020 21:05:56	NthCoastMedia	0

Retweets are characterized by beginnings of "*RT @* (....):", for example:

Table 2. Standardized format of a tweet containing mentions and a link

id_str	raw_text	created	user_name	fav_count
1231804753287360000	**@ClimateAction15 @zalisteggall** Thanks for that , what affect do you or nasa believe the sun has on our temperature ? Or the underwater volcano's on ocean temp ?? **https://t.co/qaEiRdSsGc**	24/02/2020 04:55:13	LeeroySmith68	0

The underlined part of the retweet by user called NthCoastMedia is to be eliminated before analysing the content of the tweet and at the same time it is a valuable piece of information to be stored as a new variable. Comments reacting to a tweet and its previous comments begin with a set of @-signs accompanied by usernames who previously participated in the interaction:

Picture 1 Illustration how the context of collected comment (framed part) looks like

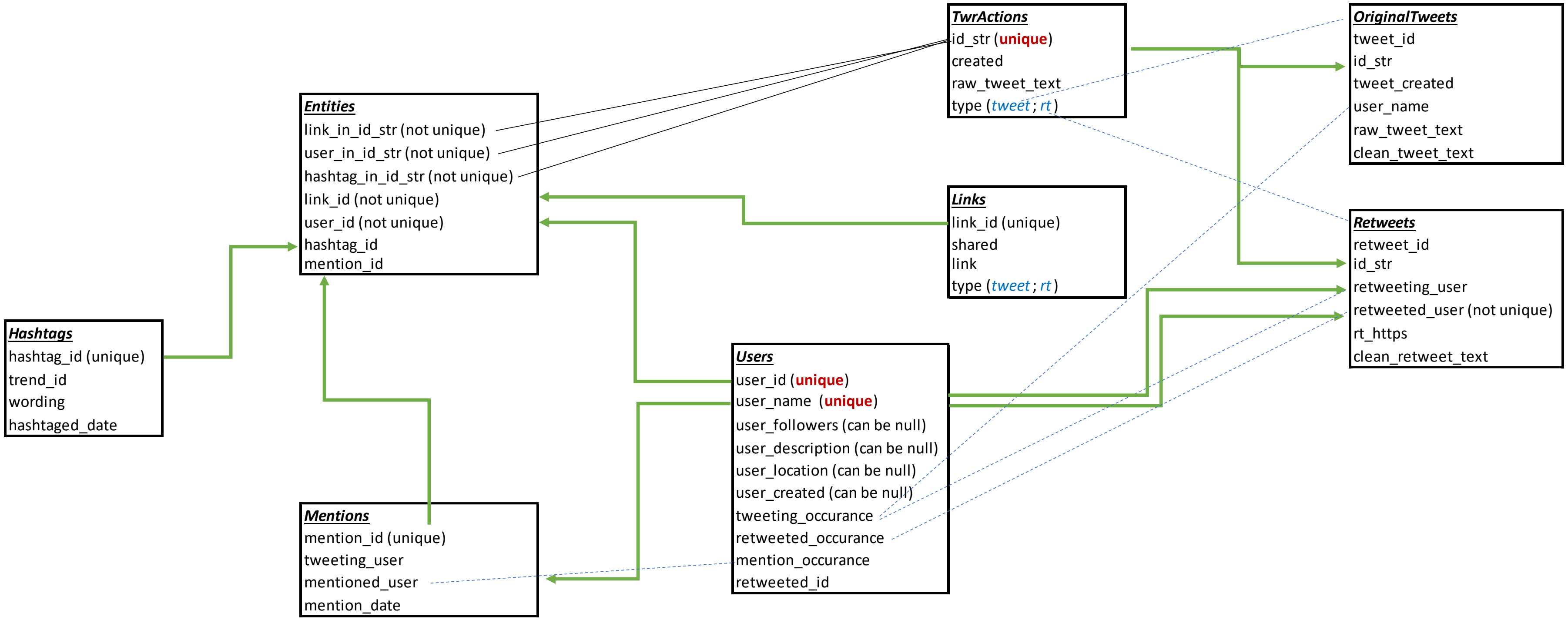
The collected action identified by number 1231804753287360000 represents a comment by LeeroySmith68 replying to a comment by a user called ClimateACtion15 (see picture 1, the collected action is framed). The link at the end of the variable *raw_text* is url to ClimateACtion15´s comment.

Original tweet can also contain @-signs accompanied by usernames (and the tweet does not have to begin with it), which means the author of the tweet mentioned another user. The endings of twitter actions may be incorporated by links in the format characterized by its beginning of "https://t.co/" and accompanied by a particular combination of characters.

Parts of the *text_raw* variable representing retweeted users, commented users, mentioned users and links were detected by string functions using SQLite and retrieved as new variables. The *text_raw* variable was cleaned from all these entities and stored as *clean_tweet_text*. Wide format of the dataset was afterwards transformed into a relational database (see **Chyba! Nenalezen zdroj odkazů.**) because it enables exploration of data from various perspectives and is an optimal strategy how to manipulate with larger datasets.

.

Figure 3. Relational database of collected tweets: table scheme



Coding

Considering the size of the studied dataset, supervised machine learning algorithm was utilized as a strategy how to overcome the time-consuming process of manual coding. Supervised learning method is based on prediction that is made on the basis of labelled examples. Once the algorithm is given rules, it is capable of automatically learning without being explicitly programmed. (Habib et al. 2019) With regards to the research aims, the desired information about environmentally themed tweets was their character – whether they represent an act of climate activism or not.

In order to obtain this information for all 5,857 environmental tweets, 20 % of them were manually coded. The Coding rules were based on the four types of Twitter activism as previously used by Hodges and Stocking (2016). Informational twitter activism was coded as present when a tweet was shared for the sake of educating the public about environmental issue and an illustrative example is: "*RT @botanyone: A new study in Philosophical Transactions of the Royal Society B argues that plant dispersal will lag rates of climate change - leading to a loss in biodiversity."*

An action request was recognized during the manual coding when a tweet was shared with manifested intention to mobilize the public about environmental issue. Attempt to get other users to do something concrete (such as to sign a petition) as well as more abstract requests simply expressing the need of action were coded as this type of activism. This category covered tweets such as: *"Treat every crisis like a crisis. we need urgent action and we are demanding #climatejustice"* or "*RT @ExtinctionR: An estimated 80% of the entire forest area – and 30 villages – may be lost. We need #climatejustice now."* or "*I just voted for #ClimateAction! Join me and play #Mission1Point5 to make your voice heard on #ClimateChange*"

The third category of promotional/supportive activism represented tweets shared with intention to promote activities of the user himself or herself as well as the activity of other climate activists. An example can be: *"RT @KimKardashian: What's your EcoResolution? This year I am going on a journey of climate action with @MyEcoResolution - a platform that enables people to step up rather than shut down in the face of our climate and ecological crisis. Join us!*"

The last recognized type of Twitter activism was category labelled as ideologically/identity reinforcing activism because it incorporates an effort to promote user´s personal view on environmental issue, disseminate his or her interpretations and attract followers of the same ideology. An example of this type is: "*RT @kenklippenstein: Big thanks to the richest man on earth for spreading awareness of the need for climate action instead of just funding."* The types are not exclusive and one tweet could be coded as more types at the same time.

The manually coded environmental tweets were used as a training dataset for training the Naïve Bayes classifier. the Naïve Bayes algorithm uses the presence or absence of words to estimate the probability that a given tweet classifies as climate activism. This method was suitable for applying on the studied tweets thanks to the fact that it does well with noisy and missing data. It is the standard for text classifications which requires relatively few examples for training, but also works well with very large numbers of examples. (Xu et al. 2017)

On the other hand, its weak points lie in the assumptions it makes. The part of its name containing ´Naïve´ has been derived from some of the "naive" assumptions about the data which are made by the algorithm. The Naïve Bayes algorithm relies on an often-faulty assumption of equally important and independent features (i.e. it assumes that all the words in the tweets are equally important an independent on each other). In general, even when these assumptions are violated, the algorithm performs quite well. Although it is not ideal for dataset containing many numeric features, but it was not the case when it comes to the Twitter data. (Nigam et al. 2000)

Putting the Naïve algorithm into practice was possible by using an R package called *e1071*. The package generates a classifier that can be used to make predictions. Before using its predictions as likely values of climate activism in further analysis, evaluation of the model performance was done. The performance of the Naïve algorithm was tested on the test subset constituted of 25 % of manually coded tweets which were not seen by the algorithm. (Lantz 2019, 122) After using the classifier on the test subset to generate predictions, comparison between the predicted values and the manually coded values was made. An R package called *gmodels* provides a function *CrossTable* which facilitated computation of misidentified percentage of the dataset.

Analysis

The independent variable entering the logistic regression was intended to be perception of scientific consensus on anthropogenic cause of climate change detected by keywords search but as a response to the lack of this particular content in the harvested dataset, the research aim has been modified. The main idea of the GBM is tested on a more general level and thus any environmental tweet containing expressions related to science entered the logistic regression as an independent variable representing the argument of scientific perspective on climate change. This strategy enables to find out whether any more sophisticated research approach (containing for example a controlled intervention in order to make sure the scientific consensus message is present in the data) is worthy of effort in the future or the whole dynamics of the environmental debate using the argument of science in the twitterverse incline to non-activist content and is used by climate change deniers). The dependent variable is information if the environmentally themed tweet was coded/classified as an act of climate activism or not.

Results

Anomaly within the random sample of tweets was recognized when a user was extremely active, so the content by such a user was harvested (i.e. randomly chosen) more than 700 times over the course of six months. Based on exploration of the whole harvested dataset, the amount of 700 tweets proved to be not so widespread number of harvested tweets per user. Detailed examination of attributes of the users with 700 and more harvested tweets revealed a frequent occurrence of bot features. Detection of attempts to make a particular hashtag to be a trending hashtag was enabled by this strategy. It revealed an activity pattern of bots who were tweeting content with an identical hashtag accompanied by randomly chosen and constantly changing words that made no sense together.

When speaking about the users from whom at least one environmentally themed tweet was harvested, 16 users produced more than 700 tweets which were randomly chosen to be streamed within the one percent sample. These producers of an environmental content with suspiciously too frequent activity did not manifest any evident signs suggesting that they were bots and the produced content was anyhow unsuitable for entering the analysis. The highest average number of harvested tweets per day was 12.67 from a user who was tweeting often enough to be randomly chosen 2,268 times over the course of the studied period. None of the extremely frequent producers of a Twitter content was tweeting predominantly about the environmental issue. The number of environmental tweets harvested by the extremely frequent content producers ranged between one and two tweets per user (see detailed information about producers of an environmental content with suspiciously too frequent activity in Table 3).

Table 3. Producers of an Environmental Content with Suspiciously too Frequent Activity

Producers of an Environmental Content with Suspiciously too Frequent Activity								
User name	User created	User location	User description	User followers	Number of env. tweets harvested	Content of environmental tweets produced by the user	Total number of tweets harvested over the course of half a year	Average number of harvested tweets per day
clairebotai	23/04/2016 17:43			4,355	2	RT @AyanDasGATech: Interesting application of #MachineLearning for studying trends of #biodiversity in #ancient history... #DataScience RT @ShiCooks: #Bangladesh is a world leader in off-grid home #solarpanels #ClimateCrisis #ClimateEmergency #ClimateAction #Renewables	2,268	12.67
femtech_	25/06/2019 14:36	Berlin, Deutschland	I'm a friendly bot retweeting female developers, engineers and scientists. #girlswhocode #womenintech #womenwhocode #womeninstem #momswhocode Made by @frankanka	11,501	1	RT @Ralph_Jacobson: Earth Day 2020, Taking Action on Climate Change #womenowned #womenintech https://t.co/yuASWmc8Su	1,754	10.63
Tomthunkit	28/04/2011 01:56	Atlanta, GA	Tomthunkit's Mind Diner is open 24/7. All-you-can-eat smorgasbord of hand-picked deliciously sweet tweets for your political palate. Donations Welcome.	65,519	1	2019 was the year young people around the world stood up and demanded real climate action. These were the key moments https://t.co/Z2hJjtiJYH	1,095	6.19
newsfilteric	30/05/2019 10:49	New York, NY	Run by investors, for investors. Real-time market news: - Earnings - M&As - FDA approvals - Insider trading - Analyst upgrades - SEC filings (8K, 10Q/K, ...)	940	1	Coronavirus crisis shows world could take climate-change action, says UK's Prince Charles #coronavirus #COVID2019 https://t.co/cgzvWei2f6	1,087	6.39
autismbot2	09/07/2019 17:28	Mars	This is a BOT! The purpose is to find #autism papers on twitter and retweet them. No Replies! #autismus #autismo #asperger #autistic #autisme	864	1	RT @Bugsey111: A plus for Global Warming! COVID-19 Drops Air Pollution & Autism Awareness Month https://t.co/z0zzgefyte via @YouTube	1,017	6.97
CoronaUpda	13/03/2020 14:11		I retweet important updates about #coronavirus Follow Us for all the related updates worldwide	660	1	RT @business: The coronavirus pandemic may have cleared skies and halted cities, but it isn´t slowing global warming https://t.co/cQMxfGx1H5	916	9.64
DIYMikes	19/11/2015 15:00	Salt Lake City, UT	Artist. Mad Scientist. Designer. Upcycler. DIYMike	11,618	1	RT @Ceilidhann: "For all the talk about climate change and a need for action, an awards show itself must have a massive carbon footprint."	843	4.93
dev_discour	05/12/2017 08:31	National Capital Region	Devdiscourse: World's leading Website for International Development News, Opinions, Interviews and Breaking News.	74,649	2	Japan, U.S. lead survey's corporate climate change action 'A List' https://t.co/iGSFepaOig All nations must come together to bring down global warming; question of planet's survival: Merkel https://t.co/jRhPVw5nHi	842	4.98
All435Reps	03/05/2019 22:54		Collecting and retweeting tweets by all members of the US House of Representatives. Archived at: http://bit.ly/all435reps	484	1	RT @RepBarragan: With the WH working against #ClimateChange solutions, cities and states must lead the way for #ClimateActionNow Calif.	816	4.80
mayur_shin	07/10/2013 08:11	Pune, India	Staff Software Engineer	512	1	RT @elonmusk: @teslaownersSV @iam_preethi I def believe in the ethical treatment of animals & taking action of climate, but these are mostl	792	7.07
EdinburghW	20/06/2016 22:03	Edinburgh, Scotland	I randomly RT content from around Edinburgh to present a slice of Lothian life. Sister account to @Glasgow_Watch. Created by actual human @ardavey.	9,426	2	RT @CAPplanners: CAP Global partner @BruceStiftel sharing the relationship between climate action planning & good City planning @WUF_10 RT @Reniour: Mass extinction 450 million years ago 'triggered by volcanic eruptions and global warming' https://t.co/edV89yOtLQ	788	4.58
BillEsteem	15/11/2016 23:17	United States of America	USA, POTUS, POP, World, Tech, Follow, fb https://t.co/eX7jVa0YBS https://t.co/nCOv1YnPJ0 https://t.co/emAuivpSlv https://t.co/VpbHgBdVpZ https://t.co/VGCVK4bxkb -- Jumanji Windows App https://t.co/jxErMjJD7w --	5,020	1	RT @nytimes: Antarctica set a record high temperature on Thursday, underscoring the global warming trend, researchers said	782	4.47
TayyabaWa	03/10/2018 11:48	In the time machine		12,965	1	RT @emel0371: Global warming also threatens this wonderful formation. The world is under a lot of threats. We can still laugh at something	715	5.30
dismisstrum	25/03/2017 23:43			1,445	1	RT @natsecaction: Trump on coronavirus sounds a lot like Trump on climate change. His war on science and facts puts us all at risk.	712	4.29
yojudenz	13/05/2013 04:10	New Zealand	Kiwi Deplorable who loves & respects America. #MAGA #IStandWithPresTrump #BlueLivesMatter #VETERANS #WeStandWithFlynn #Trump2020Landslide #NRA #MILITARY #MAWA	26,967	1	RT @PoliticalIntent: @alexsalvinews @OANN @ChanelRion @whca Global warming is a huge issue yet CNN still has Acassta releasing more	711	4.34
jurylady5	01/12/2011 05:19	Southern hemisphere	Interested in world affairs, communication, education and peace.	16,256	1	RT @profdmcinnes: Free copy of Sustainable Planet #science #environment #conservation #naturalcapital #biodiversity #nature #strategy	706	4.10

Exploration of the most active producers of environmental content did reveal some bots. Within ten the most active producers of an environmental content, four users (ecology_tweets, OurFutureBot, DumbFActs5 and ThatReallyNice1) described themselves in the user description section on the Twitter profile as automatically ran entity. The content produced by these users did not include tweets composed with malicious intentions to manipulate the social signals. The description of users called ecology_tweets and OurFutureBot suggests that both of them perform climate activism because they state to retweet content hashtagged with #EcologyTwitter (in the case of ecology_tweets user) and hashtagged with either #FridaysForFuture or #PlasticFree (in the case of OurFutureBot). The number of user followers indicates that all the detected automated users share a content that some Twitter users found useful to receive. The number of follower (ranging from 2 to 79) was most likely not artificially exaggerated (for example by other bots) in order to create an impression of influential/popular Twitter user with thousands of followers (see detailed information about the top ten most active producers of an environmental content in Table 4).

Figure 4. *TOP 20 Users Most Represented in the Environmental Debate*shows the 20 users with the highest representation in the environmental debate. The user representation can be defined as the production of original tweets (tweeting), being retweeted by other users or being mentioned in other users' tweets (for clarification see the relational database of collected tweets in **Chyba! Nenalezen zdroj odkazů.**). In general, the most common way of representation among the top 20 users was being retweeted (76% over the number of tweets in the environmental debate of the top 20 users). The most represented user in the environmental debate was a user called INDIWASHERE which is a profile of a persona without a clear association to a publicly recognized person or institution outside of the twitterverse. The fact that the environmental content shared by the user INDIWASHERE gained attention of the Twitter users and attracted so many of them to the point they retweeted it more than environmentally themed tweets by the U.S. Senator Bernie Sanders proved that Twitter platform can serve the climate activists as a powerful tool to influence other Twitter users.

Table 4. TOP 10 Most Active Producers of an Environmental Content

The Most Active Producers of an Environmental Content					
User name	**User location**	**User description**	**User followers**	**Number of env. tweets harvested**	**Total number of tweets harvested over the course of half a year**
Eco1stArt	USA	Featuring the most stunning eco friendly items on the planet: fine art, furniture, jewelry & apparel.Connecting ecological & innovative artists worldwide.	1,303	71	277
EUClimateAction		The Directorate-General for #ClimateAction (DG CLIMA) is responsible for the @EU_Commission's international & domestic activities fighting #climatechange	61,856	24	30
imagine_garden	NJ	"My grandmothers' ingenuity overcame gender discrimination and economic exploitation in a hostile society to prove their genius." #earthsoldierbook E.S.C.S	2,724	18	246
ecology_tweets	everywhere and everywhen	Retweeting all things #EcologyTwitter. Fully automated. Maintained by @MikeMahoney218	64	18	43
OurFutureBot	Italy	I'm a retweetter-bot of #FridaysForFuture and #PlasticFree My creator @GiuseppeHuman	79	17	243
RichardMunang	Nairobi, Kenya	@UNEP Innovation Award Winner; Author of the book Making #Africa Work through the Power of #innovativevolunteerism; OPINIONS are MINE and NOT my organization's	21,130	14	97
DumbFacts5		bot project created by Tarik	2	9	110
ThatReallyNice1	Sadly, anywhere	Bot attempting to be as opinionated as that one guy you run into a bit too often and who makes you clench your fist like Arthur's	3	9	97
Roger25538443		Heat in atmosphere is caused by the movements of all its gas molecules. Methane gas is 1.7in 1,000,000 moving molecules producing heat. perezcasadiego@gmail.com	6	9	46
yuvi_nation	Munsyari, India	#climatechange #zeroplastic plastic collector boy	4,249	7	148

Only three users (EUClimateAction, Eco1stArt and QTAnon1) contributed during the studied period with a significant number of original tweets so that their tweets were harvested more than 20 times. This form of representation constitutes only 9.5% of the tweets in the top 20. Some users were highly represented in the environmental debate in the form of mentions in other user´s tweets, this is the case of *GretaThunberg*, *IPBES*, *UN* and *Espanto2001.* All these four profiles have a clear association to a publicly recognized person or institution outside of the twitterverse. The representation by mentions constituted 14.3% of the tweets in the top 20.The list of the twenty most often mentioned users in the contexts of the environmental debate reveals a mix of influential people, celebrities, the news and recognized institutions such as GretaThumber, Espanto2001, ScottMorrisonMP, UN, realDonalTrump (refer to FiguresFigure *4*

Figure 5,

Figure 6, andFigure *7* for more detailed information about users represented in the environmental debate).

Figure 4. TOP 20 Users Most Represented in the Environmental Debate

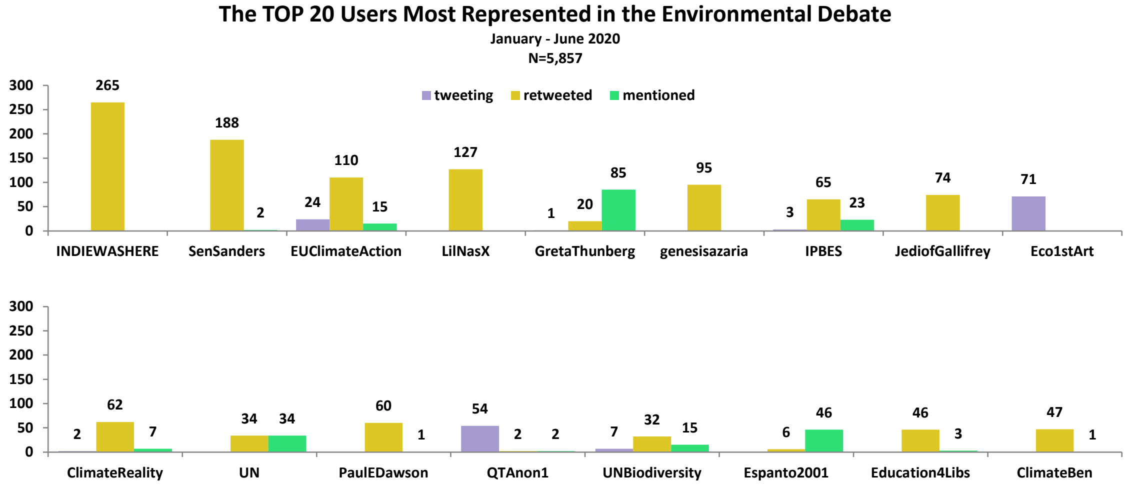


Figure 5. TOP 20 Most Often Tweeting Users within the environmental Debate

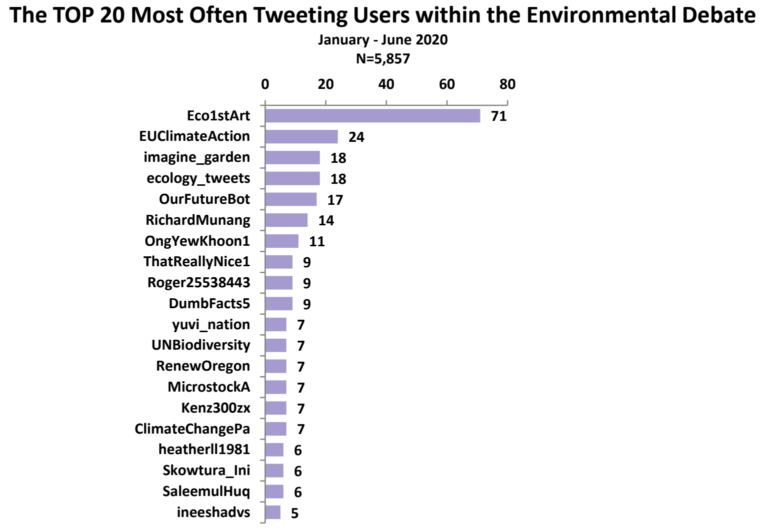


Figure 6. TOP 20 Most Often Retweeted Users within the environmental Debate

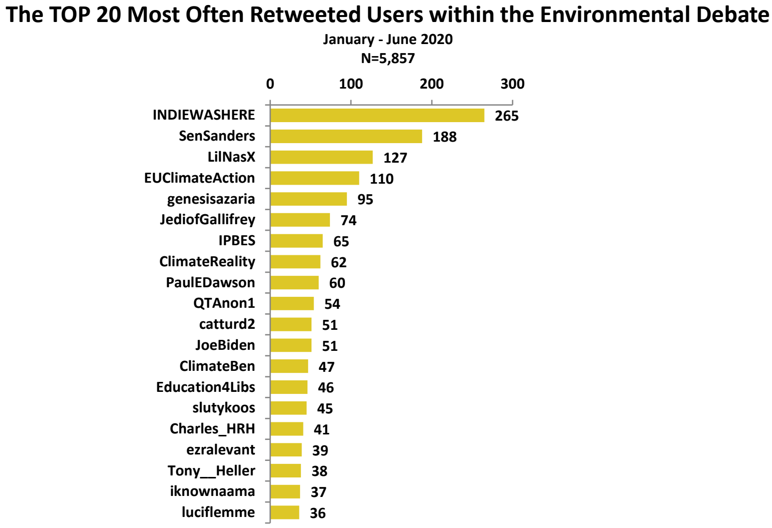
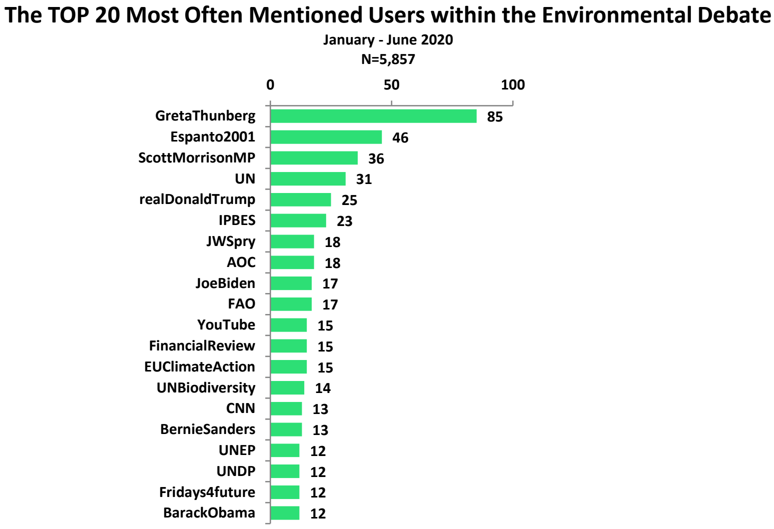


Figure 7. TOP 20 Most Often Mentioned Users within the environmental Debate



Envi Debate

- = definování obsahu, který potřebuji k testování teorie → absence zprávy o konsensu
- → klesající dynamika (→ další sign. události) → aktivismus

When the primal keyword search criterion consisting of expressions “climat”, “environ” and “global warming” was applied, the most popular hashtags provided a helpful insight into the topics that were shaping the environmental discussion. As

Table 5 shows, most popular hashtags in January reflected that the bushfires in Australia were discussed with regards to the climate change (the top 20 hashtags included "#AusPol", "bushfires", "#Australia", #AustraliaFires, #kangaroos, #AustralianFires). Hashtags related to the coronavirus outbreak got into the top twenty in March (#COVID19, #coronavirus, #COVID-19, #COVID2019, #Covid_19, #DigitalStrike) and since that month on, the coronavirus related hashtags stays a permanent part of popular

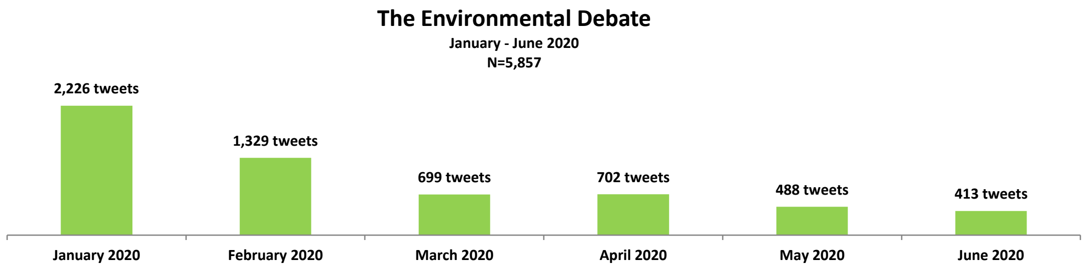
Table 5. Most popular hashtags used in tweets containing expressions “climat”, “environ” or “global warming”

The Top 20 Most Popular Hashtags within Tweets containing either an expression "*climat*", "*environ*" or "*global warming*" over the course of the first six months of 2020												
ranking	**January**	**frequency**	**February**	**frequency**	**March**	**frequency**	**April**	**frequency**	**May**	**frequency**	**June**	**frequency**
1	#ClimateChange	1945	#ClimateChange	1126	#ClimateChange	746	#darkselfie	717	#ClimateChange	463	#WorldEnvironmentDay	1236
2	#ClimateEmergency	678	#ClimateEmergency	457	#ClimateCrisis	284	#ClimateChange	578	#covid19	252	#ClimateChange	293
3	#ClimateCrisis	552	#climate	378	#COVID19	251	#EarthDay	434	#climate	193	#WorldEnvironmentDay2020	287
4	#climate	399	#ClimateCrisis	370	#climate	233	#COVID19	373	#environment	176	#environment	150
5	#ClimateAction	383	#environment	263	#coronavirus	229	#climate	181	#ClimateStrikeOnline	154	#ForNature	139
6	#Environment	275	#ClimateStrike	261	#ClimateEmergency	213	#environment	177	#ClimateAction	146	#climate	138
7	#AusPol	269	#ClimateAction	247	#environment	193	#ClimateCrisis	176	#ClimateCrisis	144	#ClimateCrisis	117
8	#ClimateStrike	233	#FridaysForFuture	126	#ClimateStrikeOnline	170	#coronavirus	173	#FridaysForFuture	120	#COVID19	105
9	#FridaysForFuture	192	#AusPol	111	#ClimateAction	152	#EarthDay2020	169	#coronavirus	95	#ClimateAction	103
10	#bushfires	159	#SchoolStrike4Climate	88	#FridaysForFuture	136	#ClimateAction	161	#ClimateEmergency	87	#NurtureTheNature	77
11	#Jakarta	135	#SOTU	76	#ClimateStrike	109	#ClimateEmergency	123	#TrumpIsNotADoctor	71	#biodiversity	63
12	#Australia	134	#ClimateJustice	75	#PFW	85	#ClimateStrikeOnline	107	#StayAtHome	68	#HyundaixBTS	55
13	#SchoolStrike4Climate	132	#DemDebate	66	#COVID-19	57	#TrumpIsNotADoctor	56	#SchoolStrike4Climate	63	#ClimateStrikeOnline	53
14	#AustraliaFires	93	#ClimateActionNow	60	#COVID2019	55	#FridaysForFuture	34	#FlattenTheCurve	56	#ClimateEmergency	53
15	#WEF20	92	#ActOnClimate	57	#auspol	53	#DigitalStrike	34	#EnvironmentConservation	45	#FridaysForFuture	51
16	#ClimateChangeIsReal	87	#ClimateHoax	49	#DemDebate	50	#environmental	32	#DigitalStrike	41	#EnvironmentDay	47
17	#ClimateCriminals	83	#GlobalWarming	47	#ClimateLaw	48	#StayAtHome	31	#ActOnClimate	40	#ClimateFriendlyFood	43
18	#kangaroos	80	#cdnpoli	45	#Covid_19	45	#nature	31	#NieR	34	#SchoolStrike4Climate	40
19	#AustralianFires	75	#environmental	41	#environmental	43	#ClimateStrike	31	#SaveCongoRainforest	27	#Nature	32
20	#DemDebate	68	#TrumpBudget	40	#DigitalStrike	43	#ActOnClimate	30	#SaturdayMorning	27	#StayAtHome	29

The events during the first two months of 2020 temporarily shifted the public discussion further from environmental issues but it does not mean this research loses its relevance completely. Considering the circumstances, it is especially valuable that the data collection targeted random sample of the whole twitterverse and not only tweets closely related to the research topic. The outbreak of coronavirus global pandemic effected almost every single aspect of social reality, not excepting the environmental discussion/activism and thus the collected dataset is loaded with indicators of how the pandemic penetrated the whole twitterverse in general and the environmental Twitter debate in particular.

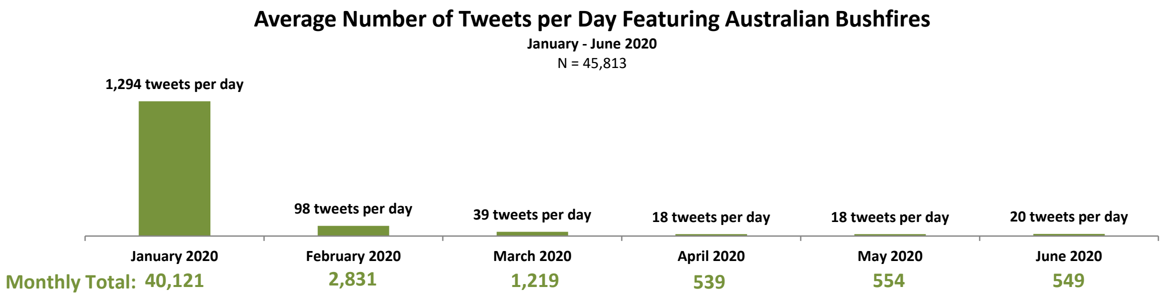
The environmental debate detected in the dataset presents an evident decreasing trend. January was the most active month with 2,226 environmentally themed tweets (see monthly frequency of tweets classified as a part of the environmental debate on Twitter in Figure 8*. Monthly frequency of tweets classified as a part of the environmental debate on Twitter*).

Figure 8. Monthly frequency of tweets classified as a part of the environmental debate on Twitter



The dynamics of the very first month can be attributed to the bushfires in Australia. Although the fires predominantly featured in the environmental debate, the event went beyond environmentally themed tweets. In January, the whole sample contained on average 1,294 tweets per day related to this event (see the evolution of the average number of tweets featuring the bushfires in Australia over the course of the studied period in Figure 9).

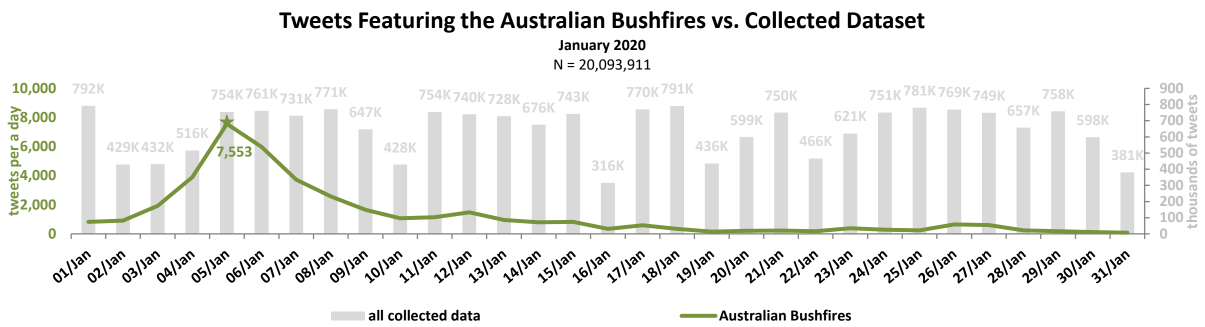
Figure 9. Evolution of average number of tweets featuring the bushfires in Australia over the course of the first six months of 2020 present not only in environmentally themed tweets



(Tweets featuring the event were filtered from the whole harvested dataset by using criterion of containing either "Australia" + "fire" or "Australia" + burn or "#AustraliaBurn" or "#AustraliaFire" or "#AustraliaWildFire" or "#AustraliaOnFire".)

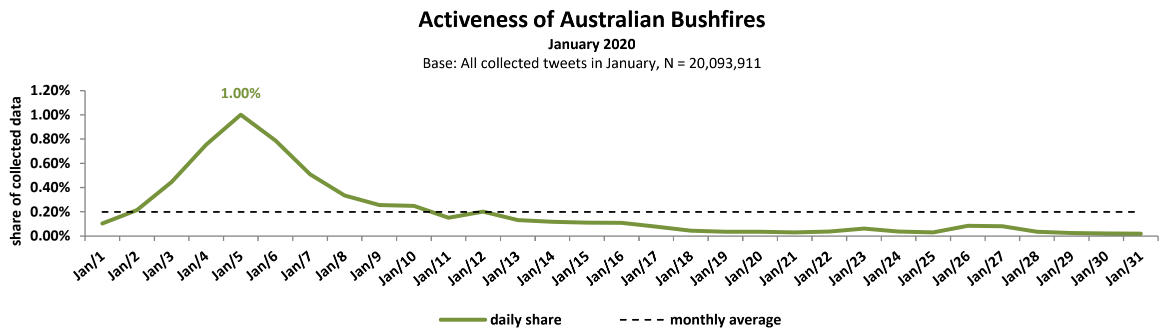
The discussion of the emergency caused by the bushfires in Australia peaked in the twitterverse on January 5th (see frequency of tweets featuring the bushfires in Australia collected in January 2020 in Figure 10). On this day, 7,663 tweets featuring this topic were collected which constitutes 1% of the whole sample and 5 times (402%) more activity compared to the average share of January on the same topic (0.1997%) (See the activeness of Australian bushfires in January 2020 in Figure 10 and Table A3a).

Figure 10. Frequency of tweets featuring the bushfires in Australia collected in January 2020 compared to all collected tweets



(Tweets featuring the event were filtered from the whole harvested dataset by using criterion of containing either "Australia" + "fire" or "Australia" + burn or "#AustraliaBurn" or "#AustraliaFire" or "#AustraliaWildFire" or "#AustraliaOnFire").

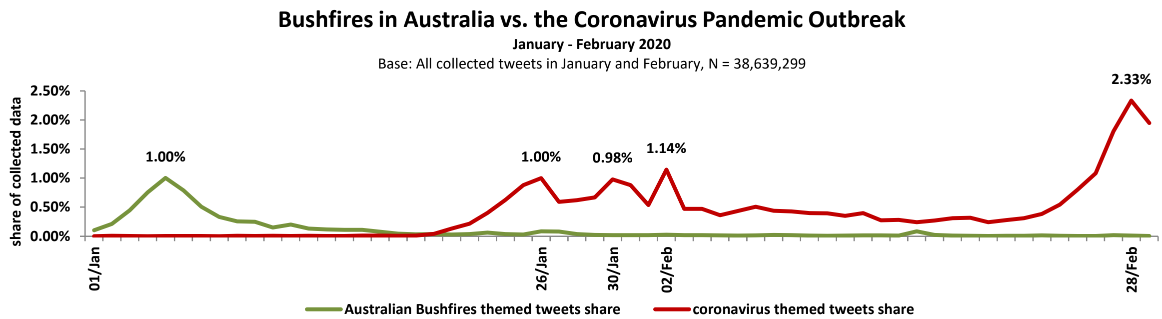
Figure 11. Activeness of Australian bushfires in January 2020



(Tweets featuring the event were filtered from the whole harvested dataset by using criterion of containing either "Australia" + "fire" or "Australia" + burn or "#AustraliaBurn" or "#AustraliaFire" or "#AustraliaWildFire" or "#AustraliaOnFire").

The attention of twitterverse started to get lost by the end of January, and by February it had almost gone most likely as a result of the coronavirus pandemic outbreak. On January 26th (1%), January 30th (0.98 %) and February 2nd (1.14 %) the topic of coronavirus recorded comparable attention to that the bushfires in Australia received during its climax on January 5th (see activeness of Australian bushfires in January 2020 compared to the Coronavirus pandemic outbreak in Figure 12). However, the maximum share corresponding to the coronavirus outbreak over the course of the half year of study is much higher than the share of the tweets featuring the bushfires.

Figure 12. Activeness of Australian bushfires in January 2020 compared to the Coronavirus pandemic outbreak

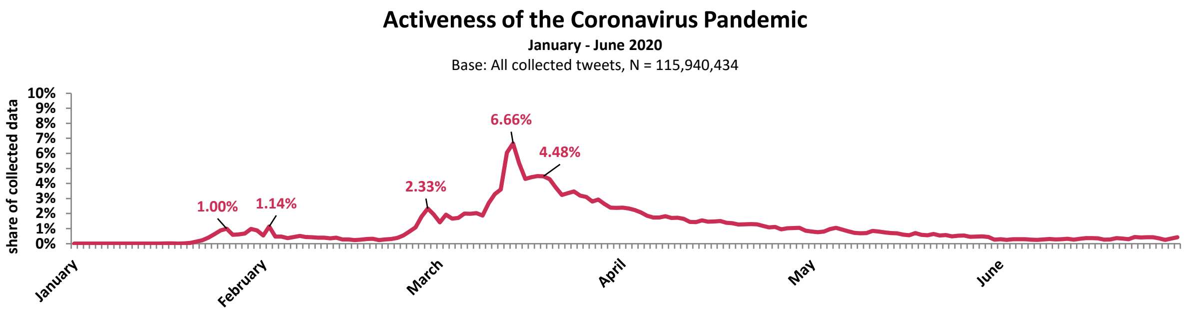


(Tweets featuring the bushfires in Australia were filtered from the whole harvested dataset by using criterion of containing either "Australia" + "fire" or "Australia" + burn or "#AustraliaBurn" or "#AustraliaFire" or "#AustraliaWildFire" or "#AustraliaOnFire". Tweets featuring the coronavirus pandemic outbreak were filtered from the whole harvested dataset by using criterion of containing either "corona" or "covid" or "pandemic" or "stay" + "home" or "stay" + "safe" or "#FlattenTheCurve" or "Wuhan" or "2019-ncov" or "SARS-cov-2").

The transition between January and February registered the first burst of tweets about the virus following the events in China. The first alarming news coming from China indicating the seriousness of the situation appeared on January 23rd (cancelation of large-scale Lunar New Year celebrations and partial lockdown of transport in and out of Wuhan, the epicentre of the spread). Twitter activity related to the global pandemic intensified at the end of February when the situation in Italy was worsening. On February 28th it reached 2.33 % share of the whole sample which corresponds to 286% more activity compared to the monthly average share (0.60 %). See *Table* A1b for detailed data about the activeness of the coronavirus pandemic during February, and Figure 13 for the activeness of the coronavirus pandemic over the course of the first six months of 2020.

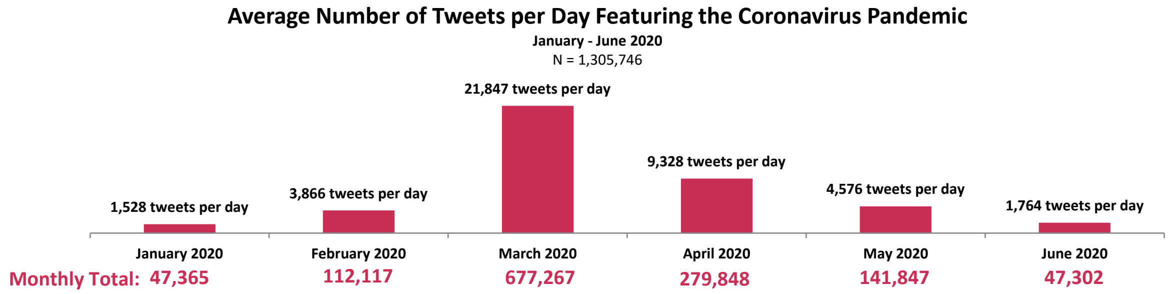
Twitterverse joined the worldwide consternation caused by the global pandemic of coronavirus especially during March 2020. On March 13th its share of the whole sample reached the highest figure of 6.66 % which represents 489% more activity compared to the half year average share (1.13%) See *Table* A1c for detailed data about the activeness of the coronavirus pandemic during March. The average number of tweets featuring the coronavirus pandemic during March was 21,847 per day. Refer to Figure 14 to the see evolution of monthly average number of tweets featuring the coronavirus pandemic over the course of the first six months of 2020.

Figure 13. Activeness of the coronavirus pandemic over the course of the first six months of 2020



(Tweets featuring the coronavirus pandemic outbreak were filtered from the whole harvested dataset by using criterion of containing either "corona" or "covid" or "pandemic" or "stay" + "home" or "stay" + "safe" or "#FlattenTheCurve" or "Wuhan" or "2019-ncov" or "SARS-cov-2").

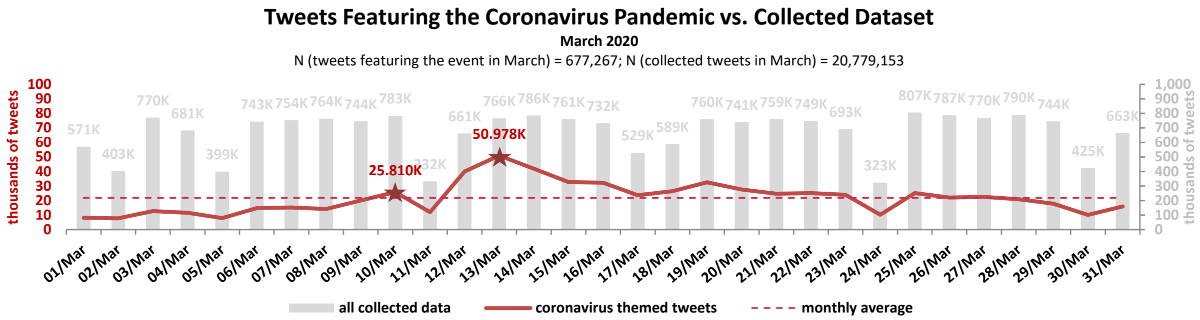
Figure 14. Evolution of monthly average number of tweets featuring the coronavirus pandemic over the course of the first six months of 2020



(Tweets featuring the coronavirus pandemic outbreak were filtered from the whole harvested dataset by using criterion of containing either "corona" or "covid" or "pandemic" or "stay" + "home" or "stay" + "safe" or "#FlattenTheCurve" or "Wuhan" or "2019-ncov" or "SARS-cov-2").

The final escalation of the number of tweets related to the pandemic took place in the middle of March. This rapid growth coincides with an ongoing critical situation in Italy. (Travel restrictions on the entire Lombardy region was placed on March 8th.) The whole country went into lockdown on March 9th and the day after, 25,810 tweets per day were recorded in the collected dataset. In the following days, the Twitter activity intensified reaching its peak on March 13th which was the most active day of the whole dataset. On this day, the highest number of tweets featuring the coronavirus pandemic reached 50,978 which coincides with the declaration of national emergency in the USA on this day. After reaching the peak on March 13th, the number of tweets related with the pandemic started to decrease gradually throughout the rest of the sampling period. Refer to Figure 15 to see the frequency of tweets featuring the coronavirus pandemic collected in March 2020.

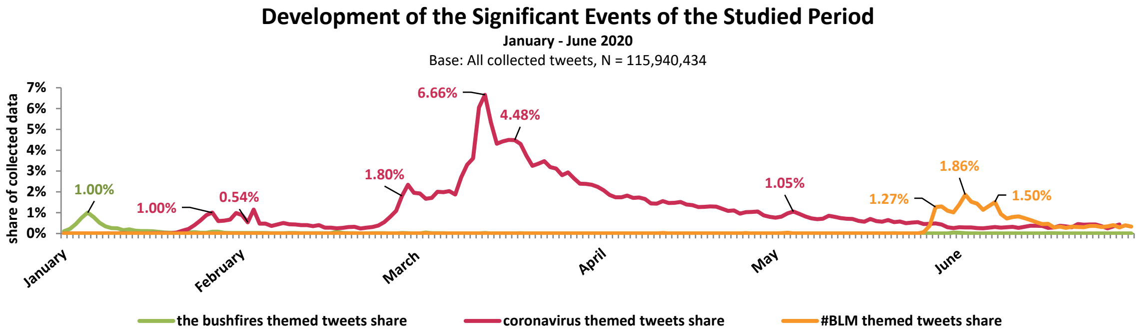
Figure 15. Frequency of tweets featuring the coronavirus pandemic collected in March 2020 compared to all the collected tweets



(Tweets featuring the coronavirus pandemic outbreak were filtered from the whole harvested dataset by using criterion of containing either "corona" or "covid" or "pandemic" or "stay" + "home" or "stay" + "safe" or "#FlattenTheCurve" or "Wuhan" or "2019-ncov" or "SARS-cov-2").

April and May presented a steady decline of the pandemic-related tweets; however, the attention received throughout April still exceeded significantly the amount of attention the climate debate received in the middle of the environmental disaster caused by bushfires in Australia in January. However, at the end of May, the topic of coronavirus global pandemic in the twitterverse was surpassed by the #BLM movement. Refer to Figure 16 to see the development of the significant events of the studied period.

Figure 16. Development of the significant events of the studied period



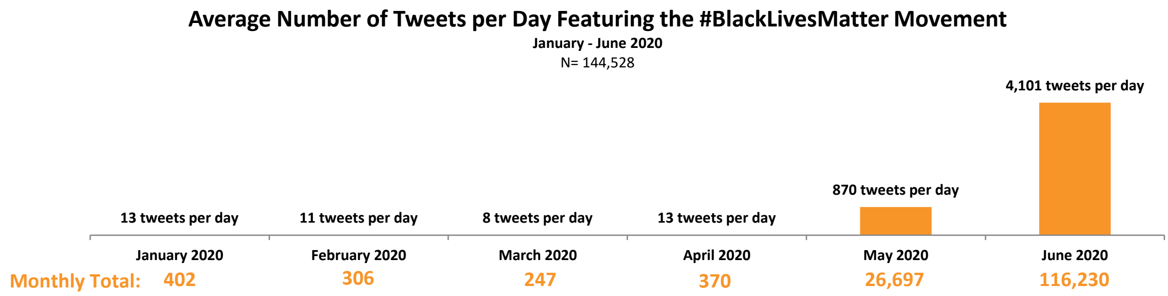
(Tweets featuring the event were filtered from the whole harvested dataset by using criterion of containing either "Australia" + "fire" or "Australia" + burn or "#AustraliaBurn" or "#AustraliaFire" or "#AustraliaWildFire" or "#AustraliaOnFire". Tweets featuring the coronavirus pandemic outbreak were filtered from the whole harvested dataset by using criterion of containing either "corona" or "covid" or "pandemic" or "stay" + "home" or "stay" + "safe" or "#FlattenTheCurve" or "Wuhan" or "2019-ncov" or "SARS-cov-2". #BLM themed tweets contain "blm" or "BlackLivesMatter" or "black" + "lives" + "matter" or "George" + "Floyd" or "#ICantBreath" or "#SayTheirNames" or "#DefundThePolice".)

Tweets containing the set phrase „black lives matter", its acronym („blm") or its hashtag were present throughout the course of the studied six months. The #BlackLivesMatter hashtag has been used for years and its origin dates back to February 2012 when an unarmed black 17-year-old American Trayvon Martin was shot. Its general purpose is to highlight how black lives have been marginalized institutionally in the USA. Besides the youth movements committed to climate change discussed above, the social movement using the hashtag #BLM is another example of marginalised groups who seek an alternative political platform on social media.

By creating a focused outrage on Twitter (and other social media platforms), the #BLM movement tries to fight against and create awareness about the fact that African Americans have historically had little voice in mainstream media. Since 2012 the hashtag has been redeployed during subsequent shootings such as in Ferguson in August 2014 when 18-year-old Mike Brown was shot by Darren Wilson or other acts of injustice towards people with black skin (Murphy, p. …).

The hashtag #BLM resurged dramatically during the studied period. While in the first four months of 2020 the monthly average of tweets per day featuring #BLM had only one or two-digit figures, May and June 2020 experienced a rapid growth of #BLM-related tweets. In May 2020, the average number of tweets per day was 870 which corresponds to a total of 26,697 #BLM-related tweets in that month (see the evolution of the monthly average number of #BLM themed tweets over the course of the first six months of 2020 in Figure 17).

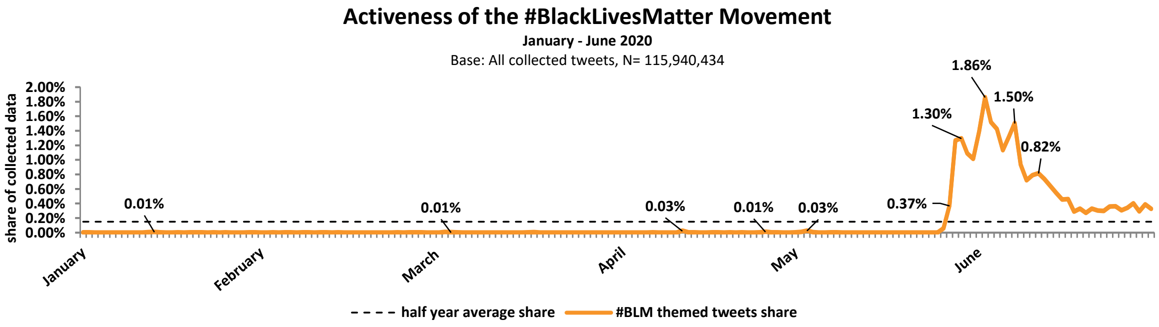
Figure 17. Evolution of the monthly average number of #BLM themed tweets over the course of the first six months of 2020



(#BLM themed tweets contain "blm" or "BlackLivesMatter" or "black" + "lives" + "matter" or "George" + "Floyd" or "#ICantBreath" or "#SayTheirNames" or "#DefundThePolice".)

Proportionately, in May 2020 the anti-racism and anti-police brutality content received 0.13 % share of the overall Twitter attention in the studied period. This high share was particularly influenced by the activity of the very last days of the month. The burst of tweets began on May 27th when only on that day #BLM themed tweets represented 0.37 % of all the collected sample equivalent to 190% more activity compared to the monthly average. Two days later, on May 29th, the monthly maximum value was recorded reaching 1.30% share of the whole dataset which represents 916% more activity compared to the monthly average. This tendency continued and the stream of tweets intensified in June. On June 2nd, the share of #BLM themed tweets climbed to 1.86 % share of all collected tweets representing 1,326% more activity compared to its half year average (0.13%). Refer to Figure 18 to see the activeness of the #BlackLivesMatter movement over the course of the first six months of 2020, and to Table A1f for detailed information of the activeness in June 2020.

Figure 18. Activeness of the #BlackLivesMatter movement over the course of the first six months of 2020



(#BLM themed tweets contain "blm" or "BlackLivesMatter" or "black" + "lives" + "matter" or "George" + "Floyd" or "#ICantBreath" or "#SayTheirNames" or "#DefundThePolice".)

The dynamics of twitterverse in May and June 2020 reflects the development of real-life events, triggered by the death of 45-year-old African-American George Floyd on May 25th. George Floyd died after being handcuffed and pinned to the ground by a police officer. The incident took place in Minneapolis while bystanders captured video of the police officer using his knee to pin George Floyd by neck. The video proves that George Floyd repeatedly said „I can´t breathe“ and was widely shared on social media afterwards.

Twitterverse got flooded with content related to #BLM movement in the following days and so did streets around the world. Protests in support of the #BLM movement, against racism and police brutality were held mainly in the USA. Considering the Twitter data feature caused by algorithmic confounding, it is worth mentioning that the hashtag #BlackLivesMatter was constantly trending during June. On top of that, the Twitter itself directly stood up in support for the BLM movement. In reaction to the events induced by the death of George Floyde, the official Twitter profile was tweeting pictures of billboards with real tweet texts supporting the movement and the description of the *Twitter* user (representing the Twitter company) on its profile stated during June 2020:

#BlackTransLivesMatter

#BlackLivesMatter

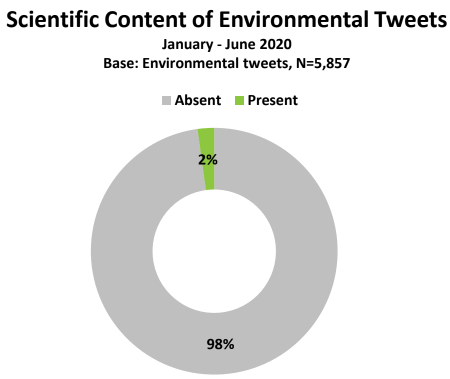
Displaying chosen tweets by ordinary users on billboards and sharing them on the official Twitter profile (https://twitter.com/Twitter) is a common strategy of the company to gain more attention and credits from its users.

Although the testing of the GBM itself went awry, the harvested data is suitable for testing at least a broader assumption coming from the theoretical model. The gained data of Twitter environmental debate still contains valuable information about climate activism.

After gaining environmental tweets featuring the message of scientific consensus on anthropogenic climate change, a faulty assumption that the environmental debate on Twitter has predominantly a denying content could be created. Majority of tweets (91 %) detected as a part of environmental debate are an act of climate activism (see Figure 20). The highest proportion of non-activist environmental tweets was detected in February and May 2020 (13%) on the contrary the least non-activist tweets constituted the detected environmental debate on Twitter in June 2020. The ML classification tended to slightly overstate non-activist character of the environmental tweets as can be seen in Figure 23.

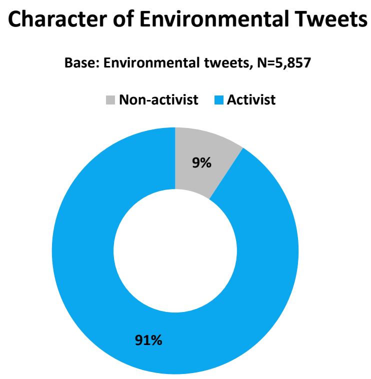
While detecting the message, about 97 % of climatologists reaching the consensus on anthropogenic climate change in the environmentally themed tweet, the hypothesis derived from the main idea of the GBM was rejected straight away. Already an early exploration of the harvested dataset proved that the randomly selected observational Twitter data from the first six month of 2020 is not convenient for any deeper analysis of the scientific consensus message. After half a year of streaming millions of tweets only 94 tweets were directly associated with this particular aspect of the environmental debate. To put it more concrete, only 31 tweets were related to the scientific consensus on anthropogenic climate change, while majority of it (18 tweets) was questioning the message and four tweets shared a link of a video called *"Do 97% of Climate Scientists Really Agree?"* that is clearly a part of a misinformation campaign. (https://t.co/mA0ReQcEVa) and its description states: "*Is it true that 97% of climate scientists agree that climate change is real? Where does the 97% figure come from? And if it is true, do they agree on both the severity of and the solution to climate change? New York Times bestselling author Alex Epstein, founder of the Center for Industrial Progress, reveals the origins of the "97%" figure and explains how to think more clearly about climate change.*" Brief exploration of the dataset suggests that the collected tweets do not contain sufficient number of acts representing perception of the scientific consensus.

Figure 19 Character of Environmental Tweets



Keywords: "scien", "expert", "research", "analys" to detect a acientific content.

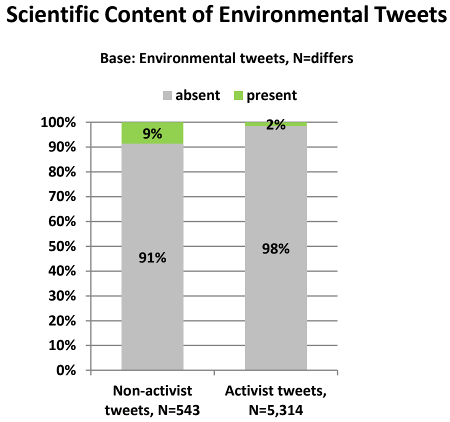
Figure 20 Character of Environmental Tweets



After gaining environmental tweets featuring the message of scientific consensus on anthropogenic climate change, an erroneous assumption that the environmental debate on Twitter has predominantly a denying content could be created. As shown in Figure 20, the majority of tweets belonging to the environmental debate were carrying the act of climate activism (91%).

The highest proportion of non-activist environmental tweets was detected in February and May 2020 (13 %) on the contrary the least non-activist tweets constituted the detected environmental debate on Twitter in June 2020. The ML classification tended to slightly overstate non-activist character of the environmental tweets (see Figure 23).

Figure 21 Scientific Content of Environmental Tweets



(To detect a scientific content environmental tweets filtered by scientific themed keywords ("scien", "expert", "research", "analys", “climatologist”)

Figure 22 Character of Environmental Tweets (Manually Coded vs. Classified by ML)

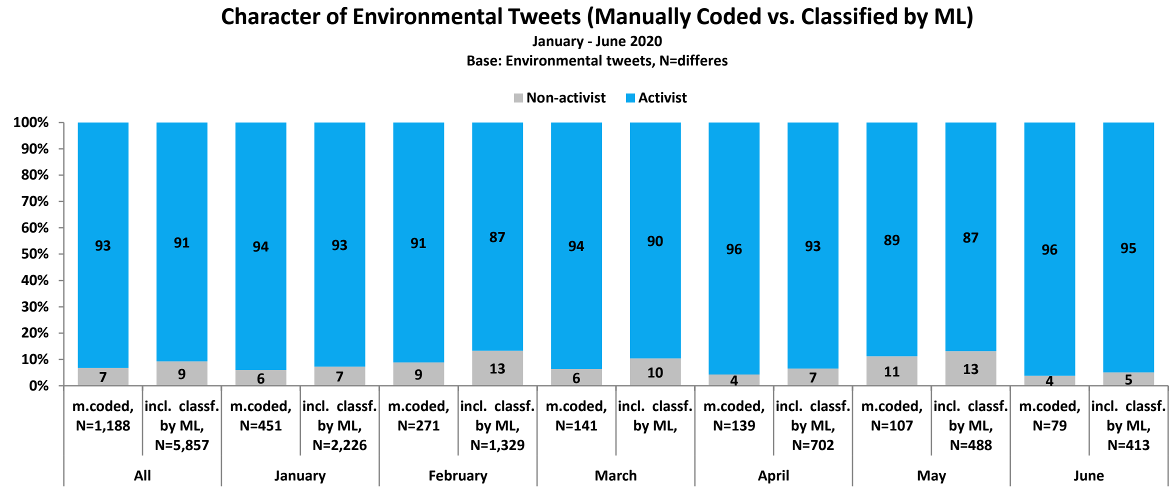


Figure 23 Character of Environmental Tweets

Validation

Manually coded dataset was split into two groups: 75% of it was used to train the classifier while the rest of 25% was manually coded and used to validate the classifier. The results of the classification were compared against the information of the manually coded dataset to obtain the classification rate.

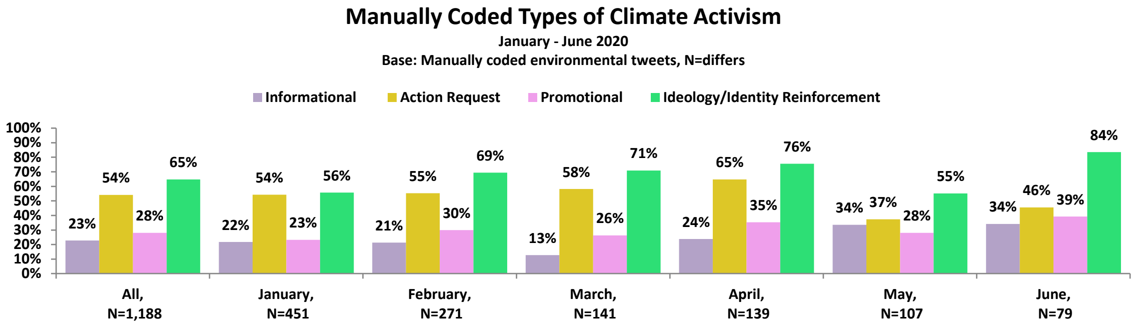
The misclassification rate ranges between 12 % and 40 %. Apart from applying the Naiive Bayes classifier on each of the Twitter activism category (i.e. informational, action request, promotional, ideology reinforcing) separately, the algorithm was used to classify either presence or absence of climate activism (in general) in the environmental tweets as well. The latter approach proved to be more successful (see error rates in Table 6).

Table 6 Validation of the Machine Learning Classification

Validadtion of the Machine Learning Classifier				
	Informational	**Action Request**	**Promo**	**Ideology/Identity Reinforcing**
Incorrectly Classified	12%	35%	40%	40%
	Activism in General			
Incorrectly Classified	8%			

Proportion of incorrectly classified environmental tweets. The validation process was applied at 289 manually coded environmental tweets (20 % of all manually coded tweets).

Figure 24 Manually Coded Types of Climate Activism



(The manually coded part constitutes of 20 % of detected environmental tweets and were used as a training dataset for the supervised machine learning algorithm.)

The detailed information about the category of environmental activism performed on Twitter over the course of the studied period was discovered only on manually coded environmental tweets (1,188 tweets). The most popular types of climate activism performed on Twitter were action request and ideology/identity reinforcement, more than half of all the manually coded environmental tweets contained these two activist strategies (54 % and 65 % respectively). January when the Australian bushfires were an important subject of the environmental debate was the month of lowest

February, March, April and June when the performance of ideology/identity reinforcement was above the overall average. During the manual coding it was noticed that both of the significant events of the studied period (i.e. the coronavirus pandemic and the #BLM movement) were featuring as an important argument used by the Twitter climate activists to interpret the issue of climate change. Different strategies how to reframe the environmental issue in the light of coronavirus pandemic were notice during the manual coding. It suggests that the authors of the studied tweets while focused on the environmental issue tried to incorporate the trending topics in order to be part of the influential debate, get attention and at the same time still support their interests and goals. One of the examples how the coronavirus shaped the environmental debate on Twitter was making parallels between the two issues which can be illustrated by: RT @InspiringU2: Like #climatechange deniers, #Coronavirus deniers are the spreaders that lead to pandemics; RT @YaleE360: The coronavirus pandemic and the slower-moving dangers of climate change parallel one another in important way; RT @bethsawin: Dear epidemiologists, We feel for you. Love, Climate scientists; RT @WCELaw: #Climatechange will require the type of transformative action that we have seen w/#COVID19 - from governments and the private sector; RT @DENROfficial: "We need to flatten both the pandemic and climate change curves." #BeatCOVID19 #ClimateAction; RT @UN: With #COVID19 and the climate emergency demanding collective global action, the UN joined people around the world in celebrating #EarthDay.

Another strategy how to incorporate the coronavirus theme in environmental tweets was to carry the message that because of getting overwhelmed by the pandemic the climate emergency was being omitted: RT @RachelNotley: Jason Kenney wants you to believe he cancelled environmental monitoring because of the pandemic. I'm not buying it.; The coronavirus pandemic is the dress rehearsal for a much costlier threat. Let's focus on #climateaction; RT @wef: Scientists warn worse pandemics are on the way if we don't protect nature #Coronavirus #Environment; RT @GlobalGoalsUN: Ocean action should not come to a halt while we tackle the #COVID19 pandemic. Just like climate change, we need to look at long term solutions for the health of our planet as a whole.; RT @UNICEF: As we deal with #COVID19, the climate crisis continues.

Slightly different strategy how to interpret the pandemic intensified was to claim that the climate emergency is much more serious problem than COVID-19: Covid-19 pandemic is 'fire drill' for effects of climate crisis, says UN official; The coronavirus pandemic is the dress rehearsal for a much costlier threat. Let's focus on #climateaction; Panic about poverty, income inequality, and climate change? Nah. Panic about this freaking virus? YES.

The opposite approach identified in within the manually coded date was framing the pandemic as an opportunity for the climate action: Today marks 50 years of #EarthDay. This year, #COVID19 underscores how important #ClimateAction is to protecting the economy saving lives around the world. Now is the time to #RecoverBetter & drive collective action; from COP14 on biodiversity to #COP26 on #ClimateChange.; RT @drvox: I hope seeing how the US has responded to coronavirus has quashed any remaining notion that climate action depends on facts, reason, persuasion, or even appeal to personal interests. You want change, take power. Everything else is vapor.; @KetanJ0 I wonder if we nomalized the time scale climate change vs inaction we would see similarities to covid spread.

Table 7 Logistic Regression

Variables in the Logistic Regression									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% Confidential Interval for EXP(B)	
								Lower	Upper
Step 1a	Absence of Scientific Argument	1.824	0.190	92.506	1.000	0.000	6.200	4.275	8.991

The computational of logistic regression used standard enter method, Nagelkerko R^2=0.027

According to the harvested dataset, a usage of scientific argument in an environmentally themed tweet suggests that the chance of a tweet with scientific reasoning is higher to not be an act of climate activism. The regression coefficient (B) was 1.824. The chance of an environmentally themed tweet with a scientific reasoning was approximately 6.2 times higher to not be an act of climate activism (with 95% confidential interval between 4.275 and 8.991).

Discussion

This research works with not such a conventional dataset which brings some advantageous features such as its nonreactivity and bigness, but on the other hand the research faced new challenges. Since the Twitter data was originally created for other purposes than research, the aims of the Twitter and its designers are likely to be imprinted in the dataset. Thus, it is not ensured that this data is somehow a direct reflection of Twitter users´ behavior or attitudes. In order to avoid systematic errors and derive data capable of offering information about naturally occurring behavior, it is necessary to understand as much as possible about the people and processes that formed the investigated dataset. Analysis without any understanding of how Twitter and its policy work could undoubtedly generate wrong conclusions. Without knowledge of possible impacts of Twitter platform on the content of tweets and behavior of Twitter users, the output of analysis is likely to say more about Twitter itself than about human behavior.

That is why a lot of effort was devoted to properly pre-processing the *found* dataset and explore its dynamics.

Another debatable feature of Twitter data is its non-representativeness. Nevertheless, there are some scientific questions for which nonrepresentative data can be quite effective and the task of this research belongs to this sort of questions. This research carries out a testing theory based on a within-sample comparison, which can still provide evidence that scientific arguments cause/increase probability of performing climate activism. Although there are questions about the extent to which a relationship that holds within a Twitter-users sample also holds within some another subpopulation. The Twitter data can be powerful for testing the theory, if the presentation of the research makes it clear about the characteristics of the tested sample. (Salganik 2018, 29-33) Moreover, the expectation of the patterns to be transportable is in this case supported with theoretical and empirical evidence, since the GBM has been successfully replicated on a large national quota sample (N=6301) of the US population. (van der Linden et al., 2019) This approach also has the potential to solve forthcoming difficulties for social scientific research caused by a constant lowering response rate. Matthew J. Salganik for example recons that: "(…) *estimates from many different groups will do more to advance social research than a single estimate from a probabilistic random sample*". (Salganik 2018, 33)

In this study it has turned out that the dynamics in twitterverse during the studied period significantly differed from experimental conditions where the Gateway Bedlief Model was validated. Since the tested theoretical model did not fit the Twitter data, an alternative explanation in relation to perception of a descriptive norm (in the form of scientific consensus) can be based on psychological reactance. Psychological reactance (Brehm, 1966) refers to the notion that attempts to limit people´s freedom, for example, via perceived pressure to adopt a certain view or attitude, can cause resistance to persuasion or even backfire. In other words: "*People do not like being told what to do, how to act, or what to believe*." (Ma et al. 2019; 73).

This result does not mean the testing of the social scientific theory on big data itself was completely unsuccessful. The presented study successfully applied machine learning algorithms in order to generate environmental debate and detect whether the environmentally themed tweet contained an act of climate activism or not. Although the Naïve Bayes Classifier performed poorly while predicting the concrete types of climate activism on Twitter (i.e. informational, action request, promotional/supportive and ideology/identity reinforcement), the supervised ML method proved to be a convenient strategy on how to retrieve the general information whether an environmentally themed tweet contain an act of climate activism or not. This result fits in line with previous research by Hodges and Stosking (2016) which successfully applied supervised machine learning algorithms to substitute the conventional social scientific procedure of manual coding. The previous research worked with a single-environmental-issue sample, to put it more concrete with sample about the Keystone XL pipeline (Hodges, Stosking 2016) while my research dealt with an environmental debate in general.

The application of machine learning algorithms is controversial because the algorithm is anticipated to make mistakes. It is evaluated according to the proportion of wrongly classified cases (in other words the model does make mistakes and it is applied anyway) and from the very beginning it let a violation of its assumptions happen. The Naïve Bayes algorithm assumes that words entering the model are equally important and independent features which is definitely not guaranteed in the analysed tweets. All in all coding data by applying machine learning algorithms seems like the research gave up ambitions to work with ´correct´ data. It must be confessed that the applied method is not fault-free. It is a result of a trade-off which on the other hand gives information about a huge amount of twitter users. The volume of the dataset places some traditional research approaches such as manual coding beyond feasibility within the scope of this study. The bigness of the studied dataset is capable to deliver information about important trends and patterns even at a cost of some noise.

Another difficulty apart from the bigness of the data needed to be process was on the contrary lack of harvested tweets that represented perceived scientific consensus on anthropogenic global warming. The exploration of the harvested dataset even revealed that it can be said that the scientific-consensus message has been hijacked by misinformation campaigns in the twitterverse during the studied period. The predominance of climate-change deniers present in the collected dataset does not have to be a generally valid rule. It is most likely caused by the specifics of the public debate during the studied period when the global pandemic of coronavirus and issue of racism dominated. Previous research proves how real-life events invoke discussions in cyberspace. Although the public discussion of the first six months of 2020 was significantly loaded with non-environmental topics and no game-changing environmental incentive was delivered, research by Leas and colleagues justifies why the original research aim was set appropriately. It propounds that if, for example, any public persona had used the argument based on scientific consensus during the studied period, the dataset would contain tweets about it. The researchers provide the evidence coming from Leonardo DiCaprio´s 2016 Oscar acceptance speech. Tweets including the terms “climate change” or “global warming” reached record hights, increasing 636% with more than 250,000 tweets the day DiCaprio spoke. In relation to *the “DiCaprio effect”* the researchers examined words closely linked to content from DiCaprio´s speech such as “hottest year” and compared them with unmentioned content also associated with climate change. (Leas et al. 2016)

The particularities of the studied period did not bring only obstacles. On the other hand, some extra unanticipated information was retrieved because the environmental debate contained a lot of references and interpretations of the coronavirus pandemic as a problem closely linked to the environmental issue. Based on some observations made during the manual coding, I suggest that the coronavirus crisis might bring new strategies of how the climate change is framed by climate activists. Therefore the strategies to utilize the coronavirus-linked development by climate activists as a new argument, might be worthy of research attention in the future.

Since the studied months were very particular, a study with similar aims might bring new findings in the future once the coronavirus is overcome, something significant happens in the environmental discourse and climatic issue gets stronger and more active again. Thus I would suggest to not give up another try to just stream the random sample of Twitter data and wait until the message about scientific consensus is remobilized by some public persona is one option but to design an experiment containing the research intervention of the scientific consensus messaging in order to make sure the message is present in the data is another option.

It needs to be admitted that there are important limitations coming from the character of the Twitter data. In some aspects I worked with a black box because the way how Twitter is designed is unknown. Especially the impact of its algorithms that might favourable the spread a curtain types of content as suggested by Vosoughi, Roy and Aral 2018 regarding the difference of easier and faster spread of misinformation compared to verified information. It is beyond the scope of this study to find out whether the predominance of denying content in the tweets containing the message about the scientific consensus was caused/significantly encouraged by the Twitter algorithm or there was some other reason.

There were also technical limits given by my social scientific background, lacking more sophisticated computer scientific knowledge. Thus for example the effort do detect and eliminate bots from the dataset was most likely sufficient enough. Therefore next time I would encourage to run an interdisciplinary research studying the Twitter data because despite the technical challenges the Twitter data once again proved to be a valuable data source for social scientific research – the construction of environmental debate within the whole gathered dataset was successful. The presented descriptive statistics provided evidence how real-life evens significantly shaped the twitterverse and got imprinted in the data.

Conclusion

The theoretical model of the GBM turned out to not be suitable for the testing on the random sample of tweets from the first six months of 2020 because the chosen data harvesting method did not gain a convenient number of tweets containing the message that 97 % of climate scientists reached consensus on anthropogenic climate change. Additionally, the majority (18 tweets) of tweets explicitly discussing the message about scientific consensus was questioning it and four tweets shared a link of a video called *Do 97% of Climate Scientists Really Agree?* that was clearly a part of a misinformation campaign.

Although the original research question had to be modified and rephrased on a more general level, the advantageous aspect of harvesting 1 % of the whole twitterverse proved to be the possibility to capture unanticipated developments which was exceptionally beneficial during the studied period. It was the time when non-environmental issues (coronavirus pandemic and #BLM movement responding to the death of George Floyd) significantly featured in the twitterverse including the environmentally themed debate. The harvested dataset held information about the dynamics of significant non-environmental events in the twitterverse, thus the activeness of coronavirus-themed tweets and #BLM-themed tweets could provide a possible explanation to the steady decreasing dynamics of the environmental debate over the studied six months. It is aluable information also because the form of climate activism itself incorporated and explicitly reacted to the significant non-environmental social development.

The presented study successfully applied machine learning algorithms in order to generate environmental debate and detect whether the environmentally themed tweet contained an act of climate activism or not. Although the Naïve Bayes Classifier performed poorly while predicting the concrete types of climate activism on Twitter (i.e. informational, action request, promotional/supportive and ideology/identity reinforcement), the supervised ML method proved to be a convenient strategy on how to retrieve the general information whether an environmentally themed tweet contain an act of climate activism or not.

The analysis suggests that the main idea of the GBM does not hold in the Twitter environment. On the contrary, the results revealed that the chance of environmentally themed argument are six times higher to not be an expression of climate activism.

List of Figures, Tables and Pictures

References

Aljets, J & Ebinger, K. (2016) Youth environmental movements: implementing socio-ecological transformation with enthusiasm. *Degrowth in Bewegung (en)*. Konzeptwek Neue Ökonomie, Leipzig, Germany. Available at: < https://www.degrowth.info/en/dim/degrowth-in-movements/youth-environmental-movement/ >

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–236.

Akiyama, Mitsuaki & Yagi, Takeshi & Yada, Takeshi & Mori, Tatsuya & Kadobayashi, Youki. (2017). Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots. Computers & Security.

Back, Mitja D., Albrecht C.P. Küfner, and Boris Egloff. 2010. "The Emotional Timeline of September 11, 2001." Psychological Science 21 (10): 1417–19.

Bimber, B & Gil de Zúñiga, H. (2020). The unedited public sphere. *New Media & Society*. 22. 700-715.

Boczkowski, P. & Mitchelstein, E. & Matassi, M. (2018). "News comes across when I'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New Media & Society*. 20.

Boukes, M. (2019). Social network sites and acquiring current affairs knowledge: The impact of Twitter and Facebook usage on learning about the news. *Journal of Information Technology & Politics*. 1-16.

Bradshaw, S. &Howard, P. (2019). The Global Disinformation Order - Global Inventory of Organised Social Media Manipulation. Computational Propaganda Research Project. University of Oxford. [online].
Available at:
< https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf (Accessed: 15 May 2020) >

Brehm, Jack W. (1966). *A theory of psychological reactance*. Academic Press.

Budak, Ceren, and Duncan Watts. 2015. "Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement." Sociological Science 2: 370–97.

Chaney, Allison & Stewart, Brandon & Engelhardt, Barbara. (2018). How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility. In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18). Association for Computing Machinery, New York, NY, USA, 224–232.

Dahal, Biraj & Kumar, Sathish & Li, Zhenlong. (2019). Spatiotemporal Topic Modeling and Sentiment Analysis of Global Climate Change Tweets.

Denning, D. (2001). Activism, Hacktivism, and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy. In Arquilla J. & Ronfeldt D. (Eds.), *Networks and Netwars: The Future of Terror, Crime, and Militancy* (pp. 239-288). Santa Monica, CA; Arlington, VA; Pittsburgh, PA: RAND Corporation. Retrieved July 11, 2020, from www.jstor.org/stable/10.7249/mr1382osd.13

Denning, Dorothy. (2000). Activism, Hacktivism, and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy. Networks and Netwars: The Future of Terror, Crime, and Militancy. 248.

Habib, Ammara & Asghar, Dr. Muhammad & Khan, Adil & Habib, Anam & Khan, Aurangzeb. (2019). False information detection in online content and its role in decision making: a systematic literature review. Social Network Analysis and Mining.

Harris, A. (2008). Young women, late modern politics, and the participatory possibilities of online cultures. *Journal of Youth Studies*. Volume 11.

Hodges, Heather & Stocking, Galen. (2016). A pipeline of tweets: environmental movements' use of Twitter in response to the Keystone XL pipeline. *Environmental Politics*. 25. 1-25.

Howard, P. & Kollanyi, B.. (2016). Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum. *SSRN Electronic Journal*. 10.

Hunt, E. (2016) What is fake news? How to spot it and what you can do to stop it. Guardian.

Kirilenko, Andrei & Stepchenkova, Svetlana. (2014). Public Microblogging on Climate Change: One Year of Twitter Worldwide. Global Environmental Change.

Lantz, Brett. (2019) *Machine learning with R: expert techniques for predictive modeling*. Packt Publishing Ltd.

Leas, Eric & Althouse, Benjamin & Dredze, Mark & Obradovich, Nick & Fowler, James & Noar, Seth & Allem, Jon-Patrick & Ayers, John. (2016). Big Data Sensors of Organic Advocacy: The Case of Leonardo DiCaprio and Climate Change. PLOS ONE.

Lee, Sangho & Kim, Jong. (2014). Early filtering of ephemeral malicious accounts on Twitter. Computer Communications. 54. 10.1016/j.comcom.2014.08.006.

Liu, Fei Tony & Ting, Kai & Zhou, Zhi-Hua. (2012). Isolation-Based Anomaly Detection. ACM Transactions on Knowledge Discovery From Data - TKDD.

Ma, Yanni & Dixon, Graham & Hmielowski, Jay. (2019). Psychological Reactance From Reading Basic Facts on Climate Change: The Role of Prior Views and Political Identification. Environmental Communication.

Manning, N. (2012). 'I mainly look at things on an issue by issue basis': Reflexivity and phronêsis in young people's political engagements. *Journal of Youth Studies* - J YOUTH STUD. Volume 16, pages 1-17.

Murphy, D.. (2018) *Twitter: social communication in the twitter age*, Cambridge, UK; Medford, MA : Polity Press.

Nigam, Kamal & Mccallum, Andrew & Thrun, Sebastian & Mitchell, Tom. (2000). Text Classification from Labeled and Unlabeled Documents using EM. Machine Learning. 39. 103-134.

O'Loughlin, B. & Gillespie, M. (2011). Dissenting Citizenship? Young People and Political Participation in the Media-security Nexus. *Parliamentary Affairs*. Volume 65, pages 115-137.

Patridge, E. (2008). From ambivalence to activism: Young people's environmental views and actions. *Youth Studies Australia*. Volume 27, pages 18-25.

Petrov, N. P. (2019) 'How Putin Tries to Depoliticize Russia's Youth' Available at:

https://www.chathamhouse.org/expert/comment/how-putin-tries-depoliticize-russia-s-youth?utm_source=Chatham%20House&utm_medium=email&utm_campaign=11203329_Publication%20alert%20Nikolai%20and%20Ekaterina%20EC%2020200107&dm_i=1S3M,6O4JL,V4DFLV,QLCLM,1#(Accessed: 23 May 2020)

Poblete, Barbara & Garcia, Ruth & Mendoza, Marcelo & Jaimes, Alejandro. (2011). Do all birds tweet the same?: characterizing twitter around the world.

Pury, Cynthia. (2011). Automation Can Lead to Confounds in Text Analysis: Back, Kufner, and Egloff (2011) and the Not-So-Angry Americans. Psychological science. 22. 835-6.

Ratkiewicz, Jacob, Michael D. Conover, Mark Meiss, Bruno Goncalves, Alessandro Flammini, and Filippo Menczer Menczer. 2011. "Detecting and Tracking Political Abuse in Social Media." In Fifth International AAAI Conference on Weblogs and Social Media.

Salganik, Matthew J. (2018) *Bit by bit: social research in the digital age*. Princeton, NJ, Princeton University Press.

Sharma, Karishma & Qian, Feng & Jiang, He & Ruchansky, Natali & Zhang, Ming & Liu, Yan. (2019). Combating Fake News: A Survey on Identification and Mitigation Techniques. ACM Transactions on Intelligent Systems and Technology.

driver's license across all age groups. Transportation Research Institute Report, UMTRI-2016-4. University of Michigan, Ann Arbor, Michigan, USA

Steinskog A, Therkelsen J, Gambäck B (2017) Twitter topic modeling by tweet aggregation. In: Proceedings of the 21st nordic conference on computational linguistics, pp 77–86.

Stolle, Dietlind and Marc Hooghe. 2011. "Shifting Inequalities: Patterns of Exclusion and Inclusion in Emerging Forms of Political Participation." *European Societies*. Volume 13, Issue 1, pages 119–142.

van der Linden, S.L., Leiserowitz, A. & Maibach, E.W. (2019) The gateway belief model: A large-scale replication. *Journal of Environmental Psychology*, 62, 49-58

van der Linden, S.L., Leiserowitz, A., Feinberg, G.D., & Maibach, E.W. (2015). The Scientific Consensus on Climate Change as a Gateway Belief: Experimental Evidence. *PloS one*.

Vosoughi, S., Roy, D, Aral, S.. The spread of true and false news online. *Science*. Available at: https://science.sciencemag.org/content/359/6380/1146 (Accessed: 15 May 2020)

Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest. (1966). Unobtrusive Measures. Chicago: Rand McNally.

Wu K, Yang S, Zhu KQ (2015) False rumors detection on Sina Weibo by propagation structures. In: IEEE international conference on data engineering, ICDE.

Xu, Shuo & Li, Yan & Zheng, Wang. (2017). Bayesian Multinomial Naïve Bayes Classifier to Text Classification.

Project Brief (Projekt diplomové práce)

Charles University (Univerzita Karlova)

Faculty of Social Sciences (Fakulta sociálních věd)

Bc. et Bc. Jana Bicanová

bicanovajana@gmail.com

SS (LS) 2019/2020

1st Year (1. ročník)

Proposed Title of Thesis (Předpokládaný název práce):

Perceived scientific consensus on global warming as a factor of climate activism on Twitter

Zaznamenání vědeckého konsensu o lidstvem zapříčiněném globálním oteplování jako prvek ovlivňující klimatický aktivismus na Twitteru

Key words: scientific consensus, global warming, climate change, climate crisis, social activism, big data, Twitter

Klíčová slova: vědecký konsensus, globální oteplování, klimatická změna, klimatická krize, sociální aktivismus, big data, Twitter

Expected date of Submission (Předpokládaný termín dokončení):

SS (LS) 2019/2020

Supervisor (Vedoucí práce)

Mgr. Jan Urban, Ph.D.

Brief Characteristics of the Theme and Existing Literature about the Issue (Stručná charakteristika tématu a stávající literatura)

Big data from social media has become an important source of information about many contemporary social phenomena such as social movement #BlackLivesMatter (Keib, Himelboim, Han, 2018), protest movement #Occupywallstreet (Wang, Caskey, 2016), spread of fear caused by epidemics (Lent et al., 2017), political polarization (Morales et al., 2015) and mental health (Tsugawa et al., 2015; Park, Cha, Cha, 2012). Social scientists have already tried to model public belief about possibility that the H1N1 (Swine Flu) virus would become a pandemic. (Ritterman, Osborne, Klein, 2009) And recently, big data from Twitter has been used to study also opinions about global climate change and their change in time (Cody et al., 2015), prediction of polar opinions (Amelkin et al., 2017), how information is transacted on social network (Ardon et al., 2013) and environmental movements´ use of Twitter to mobilize networked publics (Hodges, Stocking, 2016). However, the use of big data for testing and development of existing social scientific theories is limited by the fact that these theories use operational definitions, measures, and analytical procedures that are not easily applicable to big data from social networks given sheer volume, highly-structured content, and incompleteness of such data.

In this work I will explore how big Twitter data can be used to gain insight into factors of climate activism. More specifically, I will use the theoretical framework of the Gateway Belief Model (van der Linden et al., 2015; van der Linden et al., 2019)) to understand whether and how beliefs about scientific consensus on global climate change facilitate or hinder Twitter activism. According to GBM, the perception of scientific agreement is crucial for the change of attitudes and leads to support for action. The formation of such opinion is described as a two-step sequential process. It deals with the scientific consensus that humankind plays a significant role in climate change/crisis. The perception of such a consensus gives way to the following: 1) belief in climate change; 2) worry about climate change and/or 3) belief in human causation, which in turn determines the support for action. (van der Linden et al., 2015, 6-7)

Value of the proposed research lies in (a) probably the first attempt to test a GBM on big data, additional validation of this theory; (b) methodological contribution: this work will show that big Twitter data can be use not only for descriptive purposes but also for testing of existing social scientific theories.

Proposed Methodology (Předpokládané metody zpracování)

I will use Twitter data harvesting using Twitter packages in *R* (e.g., packages *rwteet*, *tritteR*) and Twitter API to collect a sample of data on Twitter interactions for a period spanning several months. Criteria for the tweets chosen for analysis, will be based on the topic of tweets related to climate change/crisis. In the next step, I will assess social activity and position of a sample of Twitter users based on their interactions and shared content of their Tweets and Retweets.

Finally, using machine-learning algorithms as well as manual coding by external raters, I will explore the content of Tweets produced by these Twitter users; specifically, I will try to operationalize and measure key constructs that appear in the Gateway-Belief-Model, such as perception of the scientific consensus and worry about impacts of global climate change (van der Linden et al., 2015). For example sentimental analysis (applied e.g. by Cody et al., 2015; Kouloumpis et al., 2011) might detect an occurrence of worry about climate change. The indication of a perceived scientific consensus would be considered a response to the fact that about 97 % of climate scientists have concluded that human-caused climate change/crisis is happening. (Cook et al., 2016) Based on these inputs, I should be able to explore the association between some of the key variables of the GBM and climate activism on Twitter.

Methodologically, analysis of Twitter data poses some challenges. For instance, the sheer volume of data (there are about 6,000 Twitter interactions per second) means that the data from Twitter spanning even a short period of time cannot be stored in RAM of ordinary computers and therefore cannot be analyzed in many statistical softwares that use RAM to store the data. Fortunately, statistical environment *R*, which I will be using for my analyses, offers several packages that facilitate analysis of extremely large datasets, scarce data matrices, as well as distributed computing (e.g., Lantz, 2013). Another potential problem of my work lies in the fact that conventional methods of content analysis are simply not feasible for big data. One way to overcome this limitation is to use machine learning as a workhorse for content analysis of Twitter interactions and use predictive tests (e.g., Lantz, 2013) or external raters (i.e., manual coders) to validate predictions of these algorithm.

Proposed Structure of Thesis (Předběžná struktura práce)

1. Introduction

- topic and aims of the work

2. Climate activism and its manifestation of Twitter

- sub chapter: GBM

3. Twitter interactions - what we know about interactions on Twitter, what has been done on Twitter so far

- sub chapter: use of Twitter data for the study of social activism

- sub chapter: use of Twitter data for the study of climate-related behavior (including activism)

4. Method

- Hypotheses - as derived from GBM/ from elsewhere

- Data - describe data and data harvesting (or whatever I shall use)

- Analysis - describe analytical methods

5. Results

- describe the findings; ideally structured by the research questions/ hypotheses

6. Discussion

- summarize results

- put my results in the context of existing literature

- limitations

- ideas for further research

7. Conclusions

- summarize main findings and draw very general implications (of near-cosmic significance) from my work

8. References

9. Appendix

Proposed References (Orientační seznam literatury)

Amelkin, V., Singh, A.K., & Bogdanov, P. (2017). A Distance Measure for the Analysis of Polar Opinion Dynamics in Social Networks. *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 159-162.

An, J., Cha, M., Gummadi, K.P., & Crowcroft, J.A. (2011). Media Landscape in Twitter: A World of New Conventions and Political Diversity. *ICWSM*.

Ardon, S., Bagchi, A., Mahanti, A., Ruhela, A., Seth, A., Tripathy, R.M., & Triukose, S. (2013). Spatio-temporal and events based analysis of topic popularity in twitter. *CIKM*.

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*.

Cook, J., Oreskes, N., Doran, P.T., Anderegg, W.R., Verheggen, B., Maibach, E., Carlton, J.S., Lewandowsky, S., Skuce, A.G., Green, S.A., Nuccitelli, D., Jacobs, P., Richardson, M.I., Winkler, B., Painting, R., & Rice, K. (2016). Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*. 11.

Hodges, H.E., & Stocking, G. (2016). A pipeline of tweets: environmental movements' use of Twitter in response to the Keystone XL pipeline. *Environmental Politics*, *25*, 1-25.

Keib, K., Himelboim, I., & Han, J. (2018). Important tweets matter: Predicting retweets in the #BlackLivesMatter talk on twitter. *Computers in Human Behavior, 85*, 106-115.

Kouloumpis, E., Wilson, T., & Moore, J.D. (2011). Twitter Sentiment Analysis: The Good the Bad and the OMG! *ICWSM*.

Lantz, B. (2013). *Machine learning with R: learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications*. Birmingham, UK: Packt Publ.

Lent, L.G., Sungur, H., Kunneman, F.A., Velde, B.V., & Das, E. (2017). Too Far to Care? Measuring Public Attention and Fear for Ebola Using Twitter. *Journal of medical Internet research*.

Morales, A.J., Borondo, J., Losada, J.C., & Benito, R.M. (2015). Measuring Political Polarization: Twitter shows the two sides of Venezuela. *Chaos, 25, 3*.

Park, M., Cha, C., & Cha, M. (2012). Depressive Moods of Users Portrayed in Twitter.

Riquelme, F. (2016). Measuring user influence on Twitter: A survey. *Inf. Process. Manage., 52*, 949-975.

Ritterman, J., Osborne, M., & Klein, E. (2009). Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic.

Sameh, A. (2013), A Twitter Analytic Tool to Measure Opinion, Influence and Trust. *Journal of Industrial and Intelligent Information*, 1, 37-45.

Segerberg, A., & Bennett, W.L. (2011). Social media and the organization of collective action : using Twitter to explore the ecology of two climate change protests.

Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., & Ohsaki, H. (2015). Recognizing Depression from Twitter Activity. *CHI*.

van der Linden, S.L., Leiserowitz, A. & Maibach, E.W. (2019)The gateway belief model: A large-scale replication. *Journal of Environmental Psychology*, 62, 49-58

van der Linden, S.V., Leiserowitz, A., & Maibach, E.W. (2017). Scientific agreement can neutralize politicization of facts. *Nature Human Behaviour, 2*, 2-3.

van der Linden, S.L., Leiserowitz, A., Feinberg, G.D., & Maibach, E.W. (2015). The Scientific Consensus on Climate Change as a Gateway Belief: Experimental Evidence. *PloS one*.

Wang, Z., & Caskey, K. (2016). #Occupywallstreet: An Analysis of Twitter Usage during a Protest Movement.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S.S. (2012). A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. *ACL*.

Ye, S., & Wu, S.F. (2010). Measuring Message Propagation and Social Influence on Twitter.com. *SocInfo*.

Table 8 Original format of collected data

id_str	user_followers	text	user_description	created	user_location	user_created
1224801284819050000	670	RT @ConservationOrg: Scientific communities overwhelmingly agree that the climate is in crisis and we've got 10 years to drastically cut carbon emissions	#Australia #Enterprise + #Sustainability in #Business + #Community #abOriginal #REALMix #GlobalGoals #DecadeofAction #Australia-wide Circulation.	04/02/2020 21:05:56	Australia	10/06/2016 02:24:08
1228380162783410000	29110	97% of climate cultists imagine that they can win a scientific debate by saying stuff like this.	Climate change is the religion of people who think they're too smart for religion.	14/02/2020 18:07:07		12/12/2008 17:01:46
1229662806087220000	904	The latest The Science Daily! https://t.co/N1M91pQ5Je Thanks to @veltecnetworks @leontodd @DavidSkyBrody #cybersecurity #climatechange	Computer Prof., EMT-P, wx forecaster, Amateur Radio Op. KC9AVZ, Skywarn Spotter, Space Program/NASA Enthusiast, Astrophotography/Astronomy, #STS135 Attendee :))	18/02/2020 07:03:53	Jackson WI USA	02/09/2008 18:57:55
1229714853213820000	463	RT @BeachMilk: This is how they brainwash an entire population into believing the #ClimateChangeHoax is real. SCIENTISTS ARE PAID to "prove" #ClimateChange theories	Geologist. Lives south of Calgary in country. Solar panels on home. Believe Canadian economy is not given enough attention. Retweets aren't an endorsement.	18/02/2020 10:30:42	Alberta, Canada	16/02/2009 18:00:03
1229887427839280000	34535	NASA Shows Us How Climate Change Will Drastically Change the #Ocean: Applied Sciences https://t.co/KzHX0x7Zqk https://t.co/QE0MAtW0Wb	#Environment #Climate 24/7 by Dr. Glen Barry with #DataScience - #Forest #Ocean #Water #Science #Ecology #Indigenous #HumanRights #Trafficking #Coronavirus	18/02/2020 21:56:27	New York, NY	11/01/2013 14:43:30
1230088406341960000	1705	Climate change: Fertiliser could be used to power ocean-going ships #science #environment #cleanenergy https://t.co/K0oY6yE2uo	I'm a Bohemian. Iâ€™m a Traveler.	19/02/2020 11:15:04		14/09/2010 00:18:50
1231843131147960000	88	RT @FriendsOScience: @VassyKapelos @MariekeWalsh These #carbon climateers would have killed it https://t.co/sShKJFePEn	Christian Parental rights Free Speech MAD MAX 2019	24/02/2020 07:27:43		06/03/2019 18:01:44
1231898953152650000	1224	Trump budget calls for slashing funds to climate science centers - https://t.co/pn4QCzTGyP #GoogleAlerts	Focuses on the issues of climate, sea level rise, animal's welfare, humanity challenges. shearing ideas, research, trends and topics facing our planet	24/02/2020 11:09:32	Florida, USA	05/12/2017 21:32:20

for a limited amount of information 1) tweet ID (the unique code given to the tweet and collected as a variable labelled *id_str*); 2) text of the tweet (collected as a variable labelled *raw_text*); 3) time when the tweet was shared (collected as a variable labelled *created*); 4) author of the tweet´s user name (collected as a variable labelled *user_name*); 5) number of followers the user has (collected as a variable labelled *user_followers*); 6) description of the user written by the user and made available on his or her profile, which is meant to introduce the user (collected as a variable labelled *user_descripton*); 7) user location in the case that the user allowed this piece of information to be shared (collected as a variable labelled *user_location*); 8) time when the profile of the user was created (collected as a variable labelled *user_created*) and number of likes the tweet had by the time of collection which is always zero because the streaming is at the very moment of the sharing act.

Appendix

Appendix 1. Statistics of collected tweets featuring Covid19 event

Table A1a. Statistics of collected tweets featuring Covid19 during January

Date	All collected data (in thousands)	Coronavirus pandemic related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Jan/1	792	16	0,00202%	0,23572%	1,13107%	-99,14%	-99,82%
Jan/2	429	30	0,00699%	0,23572%	1,13107%	-97,03%	-99,38%
Jan/3	432	11	0,00255%	0,23572%	1,13107%	-98,92%	-99,77%
Jan/4	516	9	0,00174%	0,23572%	1,13107%	-99,26%	-99,85%
Jan/5	754	26	0,00345%	0,23572%	1,13107%	-98,54%	-99,70%
Jan/6	761	41	0,00539%	0,23572%	1,13107%	-97,71%	-99,52%
Jan/7	731	30	0,00410%	0,23572%	1,13107%	-98,26%	-99,64%
Jan/8	771	18	0,00234%	0,23572%	1,13107%	-99,01%	-99,79%
Jan/9	647	44	0,00680%	0,23572%	1,13107%	-97,11%	-99,40%
Jan/10	428	16	0,00373%	0,23572%	1,13107%	-98,42%	-99,67%
Jan/11	754	54	0,00716%	0,23572%	1,13107%	-96,96%	-99,37%
Jan/12	740	24	0,00324%	0,23572%	1,13107%	-98,62%	-99,71%
Jan/13	728	48	0,00660%	0,23572%	1,13107%	-97,20%	-99,42%
Jan/14	676	21	0,00311%	0,23572%	1,13107%	-98,68%	-99,73%
Jan/15	743	25	0,00336%	0,23572%	1,13107%	-98,57%	-99,70%
Jan/16	316	32	0,01012%	0,23572%	1,13107%	-95,71%	-99,11%
Jan/17	770	88	0,01142%	0,23572%	1,13107%	-95,15%	-98,99%
Jan/18	791	70	0,00885%	0,23572%	1,13107%	-96,25%	-99,22%
Jan/19	436	57	0,01307%	0,23572%	1,13107%	-94,46%	-98,84%
Jan/20	599	240	0,04008%	0,23572%	1,13107%	-83,00%	-96,46%
Jan/21	750	998	0,13310%	0,23572%	1,13107%	-43,53%	-88,23%
Jan/22	466	1 006	0,21577%	0,23572%	1,13107%	-8,46%	-80,92%
Jan/23	621	2 500	0,40266%	0,23572%	1,13107%	70,82%	-64,40%
Jan/24	751	4 675	0,62287%	0,23572%	1,13107%	164,24%	-44,93%
Jan/25	781	6 854	0,87805%	0,23572%	1,13107%	272,50%	-22,37%
Jan/26	769	7 681	0,99891%	0,23572%	1,13107%	323,77%	-11,68%
Jan/27	749	4 417	0,58936%	0,23572%	1,13107%	150,03%	-47,89%
Jan/28	657	4 069	0,61906%	0,23572%	1,13107%	162,63%	-45,27%
Jan/29	758	5 066	0,66859%	0,23572%	1,13107%	183,64%	-40,89%
Jan/30	598	5 847	0,97737%	0,23572%	1,13107%	314,64%	-13,59%
Jan/31	381	3 352	0,88055%	0,23572%	1,13107%	273,56%	-22,15%

Table A1b. Statistics of collected tweets featuring Covid19 during February

Date	All collected data (in thousands)	Coronavirus pandemic related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Feb/1	743	3 983	0,53571%	0,60455%	1,13107%	-11,39%	-52,64%
Feb/2	651	7 437	1,14321%	0,60455%	1,13107%	89,10%	1,07%
Feb/3	390	1 840	0,47146%	0,60455%	1,13107%	-22,02%	-58,32%
Feb/4	747	3 518	0,47064%	0,60455%	1,13107%	-22,15%	-58,39%
Feb/5	587	2 122	0,36155%	0,60455%	1,13107%	-40,20%	-68,03%
Feb/6	388	1 681	0,43365%	0,60455%	1,13107%	-28,27%	-61,66%
Feb/7	612	3 100	0,50637%	0,60455%	1,13107%	-16,24%	-55,23%
Feb/8	386	1 691	0,43837%	0,60455%	1,13107%	-27,49%	-61,24%
Feb/9	335	1 438	0,42864%	0,60455%	1,13107%	-29,10%	-62,10%
Feb/10	679	2 712	0,39925%	0,60455%	1,13107%	-33,96%	-64,70%
Feb/11	768	3 039	0,39589%	0,60455%	1,13107%	-34,52%	-65,00%
Feb/12	727	2 553	0,35096%	0,60455%	1,13107%	-41,95%	-68,97%
Feb/13	739	2 935	0,39728%	0,60455%	1,13107%	-34,28%	-64,88%
Feb/14	738	2 028	0,27479%	0,60455%	1,13107%	-54,55%	-75,71%
Feb/15	518	1 449	0,27995%	0,60455%	1,13107%	-53,69%	-75,25%
Feb/16	653	1 581	0,24218%	0,60455%	1,13107%	-59,94%	-78,59%
Feb/17	751	2 039	0,27155%	0,60455%	1,13107%	-55,08%	-75,99%
Feb/18	692	2 138	0,30885%	0,60455%	1,13107%	-48,91%	-72,69%
Feb/19	748	2 367	0,31661%	0,60455%	1,13107%	-47,63%	-72,01%
Feb/20	745	1 787	0,23980%	0,60455%	1,13107%	-60,33%	-78,80%
Feb/21	568	1 577	0,27781%	0,60455%	1,13107%	-54,05%	-75,44%
Feb/22	550	1 699	0,30875%	0,60455%	1,13107%	-48,93%	-72,70%
Feb/23	761	2 906	0,38177%	0,60455%	1,13107%	-36,85%	-66,25%
Feb/24	763	4 149	0,54351%	0,60455%	1,13107%	-10,10%	-51,95%
Feb/25	739	5 918	0,80115%	0,60455%	1,13107%	32,52%	-29,17%
Feb/26	736	7 963	1,08262%	0,60455%	1,13107%	79,08%	-4,28%
Feb/27	635	11 458	1,80458%	0,60455%	1,13107%	198,50%	59,55%
Feb/28	443	10 327	2,33314%	0,60455%	1,13107%	285,93%	106,28%
Feb/29	754	14 682	1,94759%	0,60455%	1,13107%	222,15%	72,19%

Table A1c. Statistics of collected tweets featuring Covid19 during March

Date	All collected data (in thousands)	Coronavirus pandemic related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Mar/1	571	8 055	1,41133%	3,25936%	1,13107%	-56,70%	24,78%
Mar/2	403	7 744	1,92113%	3,25936%	1,13107%	-41,06%	69,85%
Mar/3	770	12 827	1,66511%	3,25936%	1,13107%	-48,91%	47,21%
Mar/4	681	11 608	1,70333%	3,25936%	1,13107%	-47,74%	50,59%
Mar/5	399	7 995	2,00156%	3,25936%	1,13107%	-38,59%	76,96%
Mar/6	743	14 732	1,98367%	3,25936%	1,13107%	-39,14%	75,38%
Mar/7	754	15 280	2,02738%	3,25936%	1,13107%	-37,80%	79,24%
Mar/8	764	14 274	1,86868%	3,25936%	1,13107%	-42,67%	65,21%
Mar/9	744	20 059	2,69529%	3,25936%	1,13107%	-17,31%	138,30%
Mar/10	783	25 810	3,29633%	3,25936%	1,13107%	1,13%	191,43%
Mar/11	332	11 989	3,60727%	3,25936%	1,13107%	10,67%	218,92%
Mar/12	661	40 014	6,04906%	3,25936%	1,13107%	85,59%	434,81%
Mar/13	766	50 978	6,65691%	3,25936%	1,13107%	104,24%	488,55%
Mar/14	786	41 931	5,33719%	3,25936%	1,13107%	63,75%	371,87%
Mar/15	761	32 791	4,30904%	3,25936%	1,13107%	32,21%	280,97%
Mar/16	732	32 345	4,42123%	3,25936%	1,13107%	35,65%	290,89%
Mar/17	529	23 759	4,49231%	3,25936%	1,13107%	37,83%	297,17%
Mar/18	589	26 399	4,48172%	3,25936%	1,13107%	37,50%	296,24%
Mar/19	760	32 608	4,29320%	3,25936%	1,13107%	31,72%	279,57%
Mar/20	741	27 654	3,73036%	3,25936%	1,13107%	14,45%	229,81%
Mar/21	759	24 639	3,24527%	3,25936%	1,13107%	-0,43%	186,92%
Mar/22	749	25 089	3,35181%	3,25936%	1,13107%	2,84%	196,34%
Mar/23	693	24 019	3,46617%	3,25936%	1,13107%	6,35%	206,45%
Mar/24	323	10 308	3,18885%	3,25936%	1,13107%	-2,16%	181,93%
Mar/25	807	25 110	3,11286%	3,25936%	1,13107%	-4,49%	175,21%
Mar/26	787	22 016	2,79741%	3,25936%	1,13107%	-14,17%	147,32%
Mar/27	770	22 543	2,92907%	3,25936%	1,13107%	-10,13%	158,96%
Mar/28	790	20 873	2,64220%	3,25936%	1,13107%	-18,94%	133,60%
Mar/29	744	17 806	2,39171%	3,25936%	1,13107%	-26,62%	111,45%
Mar/30	425	10 118	2,38138%	3,25936%	1,13107%	-26,94%	110,54%
Mar/31	663	15 894	2,39579%	3,25936%	1,13107%	-26,49%	111,82%

Table A1d. Statistics of collected tweets featuring Covid19 during April

Date	All collected data (in thousands)	Coronavirus pandemic related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Apr/1	740	17 297	2,33721%	1,47861%	1,13107%	58,07%	106,64%
Apr/2	773	17 299	2,23933%	1,47861%	1,13107%	51,45%	97,98%
Apr/3	762	15 841	2,07757%	1,47861%	1,13107%	40,51%	83,68%
Apr/4	772	14 263	1,84779%	1,47861%	1,13107%	24,97%	63,37%
Apr/5	537	9 317	1,73505%	1,47861%	1,13107%	17,34%	53,40%
Apr/6	192	3 324	1,73329%	1,47861%	1,13107%	17,22%	53,24%
Apr/7	661	12 027	1,81870%	1,47861%	1,13107%	23,00%	60,79%
Apr/8	653	11 138	1,70655%	1,47861%	1,13107%	15,42%	50,88%
Apr/9	763	13 149	1,72253%	1,47861%	1,13107%	16,50%	52,29%
Apr/10	769	12 746	1,65677%	1,47861%	1,13107%	12,05%	46,48%
Apr/11	215	3 110	1,44519%	1,47861%	1,13107%	-2,26%	27,77%
Apr/12	433	6 181	1,42890%	1,47861%	1,13107%	-3,36%	26,33%
Apr/13	408	6 340	1,55362%	1,47861%	1,13107%	5,07%	37,36%
Apr/14	780	11 390	1,46100%	1,47861%	1,13107%	-1,19%	29,17%
Apr/15	781	11 495	1,47179%	1,47861%	1,13107%	-0,46%	30,12%
Apr/16	798	12 032	1,50868%	1,47861%	1,13107%	2,03%	33,38%
Apr/17	639	8 897	1,39238%	1,47861%	1,13107%	-5,83%	23,10%
Apr/18	518	7 064	1,36267%	1,47861%	1,13107%	-7,84%	20,48%
Apr/19	283	3 585	1,26530%	1,47861%	1,13107%	-14,43%	11,87%
Apr/20	802	10 266	1,28016%	1,47861%	1,13107%	-13,42%	13,18%
Apr/21	782	10 148	1,29816%	1,47861%	1,13107%	-12,20%	14,77%
Apr/22	798	10 293	1,29029%	1,47861%	1,13107%	-12,74%	14,08%
Apr/23	795	9 436	1,18628%	1,47861%	1,13107%	-19,77%	4,88%
Apr/24	108	1 175	1,08703%	1,47861%	1,13107%	-26,48%	-3,89%
Apr/25	491	5 443	1,10788%	1,47861%	1,13107%	-25,07%	-2,05%
Apr/26	804	7 675	0,95411%	1,47861%	1,13107%	-35,47%	-15,65%
Apr/27	792	8 062	1,01798%	1,47861%	1,13107%	-31,15%	-10,00%
Apr/28	803	8 295	1,03337%	1,47861%	1,13107%	-30,11%	-8,64%
Apr/29	796	8 409	1,05676%	1,47861%	1,13107%	-28,53%	-6,57%
Apr/30	478	4 151	0,86769%	1,47861%	1,13107%	-41,32%	-23,29%

Table A1e. Statistics of collected tweets featuring Covid19 during May

Date	All collected data (in thousands)	Coronavirus pandemic related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
May/1	611	4 877	0,79869%	0,67137%	1,13107%	18,96%	-29,39%
May/2	772	5 878	0,76136%	0,67137%	1,13107%	13,40%	-32,69%
May/3	782	6 330	0,80993%	0,67137%	1,13107%	20,64%	-28,39%
May/4	783	7 517	0,95954%	0,67137%	1,13107%	42,92%	-15,17%
May/5	768	8 058	1,04892%	0,67137%	1,13107%	56,24%	-7,26%
May/6	598	5 633	0,94146%	0,67137%	1,13107%	40,23%	-16,76%
May/7	583	4 780	0,82036%	0,67137%	1,13107%	22,19%	-27,47%
May/8	761	5 512	0,72407%	0,67137%	1,13107%	7,85%	-35,98%
May/9	759	5 191	0,68386%	0,67137%	1,13107%	1,86%	-39,54%
May/10	787	5 564	0,70695%	0,67137%	1,13107%	5,30%	-37,50%
May/11	785	6 671	0,84927%	0,67137%	1,13107%	26,50%	-24,91%
May/12	624	4 999	0,80082%	0,67137%	1,13107%	19,28%	-29,20%
May/13	532	3 934	0,73920%	0,67137%	1,13107%	10,10%	-34,65%
May/14	787	5 532	0,70287%	0,67137%	1,13107%	4,69%	-37,86%
May/15	766	5 293	0,69109%	0,67137%	1,13107%	2,94%	-38,90%
May/16	885	5 355	0,60477%	0,67137%	1,13107%	-9,92%	-46,53%
May/17	818	4 588	0,56115%	0,67137%	1,13107%	-16,42%	-50,39%
May/18	674	4 721	0,70035%	0,67137%	1,13107%	4,32%	-38,08%
May/19	595	3 522	0,59174%	0,67137%	1,13107%	-11,86%	-47,68%
May/20	651	3 637	0,55846%	0,67137%	1,13107%	-16,82%	-50,63%
May/21	455	2 923	0,64281%	0,67137%	1,13107%	-4,25%	-43,17%
May/22	783	4 239	0,54124%	0,67137%	1,13107%	-19,38%	-52,15%
May/23	816	4 708	0,57696%	0,67137%	1,13107%	-14,06%	-48,99%
May/24	638	3 114	0,48829%	0,67137%	1,13107%	-27,27%	-56,83%
May/25	621	3 250	0,52349%	0,67137%	1,13107%	-22,03%	-53,72%
May/26	761	4 133	0,54342%	0,67137%	1,13107%	-19,06%	-51,96%
May/27	782	3 559	0,45540%	0,67137%	1,13107%	-32,17%	-59,74%
May/28	288	1 340	0,46539%	0,67137%	1,13107%	-30,68%	-58,85%
May/29	759	3 648	0,48042%	0,67137%	1,13107%	-28,44%	-57,53%
May/30	563	2 438	0,43315%	0,67137%	1,13107%	-35,48%	-61,70%
May/31	341	903	0,26505%	0,67137%	1,13107%	-60,52%	-76,57%

Table A1f. Statistics of collected tweets featuring Covid19 during June

Date	All collected data (in thousands)	Coronavirus pandemic related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Jun/1	404	1 207	0,29878%	0,32141%	1,13107%	-7,04%	-73,58%
Jun/2	737	1 897	0,25734%	0,32141%	1,13107%	-19,93%	-77,25%
Jun/3	751	2 212	0,29468%	0,32141%	1,13107%	-8,32%	-73,95%
Jun/4	730	2 103	0,28794%	0,32141%	1,13107%	-10,41%	-74,54%
Jun/5	721	2 108	0,29241%	0,32141%	1,13107%	-9,02%	-74,15%
Jun/6	706	1 827	0,25895%	0,32141%	1,13107%	-19,44%	-77,11%
Jun/7	353	869	0,24643%	0,32141%	1,13107%	-23,33%	-78,21%
Jun/8	513	1 445	0,28157%	0,32141%	1,13107%	-12,39%	-75,11%
Jun/9	689	2 112	0,30664%	0,32141%	1,13107%	-4,60%	-72,89%
Jun/10	706	1 985	0,28112%	0,32141%	1,13107%	-12,54%	-75,15%
Jun/11	660	1 972	0,29901%	0,32141%	1,13107%	-6,97%	-73,56%
Jun/12	708	2 251	0,31791%	0,32141%	1,13107%	-1,09%	-71,89%
Jun/13	492	1 328	0,27004%	0,32141%	1,13107%	-15,98%	-76,13%
Jun/14	502	1 611	0,32055%	0,32141%	1,13107%	-0,27%	-71,66%
Jun/15	513	1 893	0,36897%	0,32141%	1,13107%	14,80%	-67,38%
Jun/16	631	2 347	0,37170%	0,32141%	1,13107%	15,64%	-67,14%
Jun/17	614	2 184	0,35579%	0,32141%	1,13107%	10,69%	-68,54%
Jun/18	611	1 639	0,26829%	0,32141%	1,13107%	-16,53%	-76,28%
Jun/19	369	1 051	0,28460%	0,32141%	1,13107%	-11,45%	-74,84%
Jun/20	541	1 981	0,36623%	0,32141%	1,13107%	13,94%	-67,62%
Jun/21	601	2 012	0,33487%	0,32141%	1,13107%	4,19%	-70,39%
Jun/22	134	395	0,29430%	0,32141%	1,13107%	-8,43%	-73,98%
Jun/23	610	2 726	0,44654%	0,32141%	1,13107%	38,93%	-60,52%
Jun/24	615	2 560	0,41627%	0,32141%	1,13107%	29,51%	-63,20%
Jun/25	343	1 475	0,42983%	0,32141%	1,13107%	33,73%	-62,00%
Jun/26	336	1 451	0,43123%	0,32141%	1,13107%	34,17%	-61,87%
Jun/27	593	2 061	0,34745%	0,32141%	1,13107%	8,10%	-69,28%
Jun/28	479	1 204	0,25122%	0,32141%	1,13107%	-21,84%	-77,79%
Jun/29	465	1 561	0,33572%	0,32141%	1,13107%	4,45%	-70,32%
Jun/30	340	1 463	0,43053%	0,32141%	1,13107%	33,95%	-61,94%

Appendix 2. Statistics of collected tweets featuring #BLM

Table A2a. Statistics of collected tweets featuring #BLM during January

Date	All collected data (in thousands)	Coronavirus pandemic related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Jan/1	792	16	0,00202%	0,23572%	1,13107%	-99,14%	-99,82%
Jan/2	429	30	0,00699%	0,23572%	1,13107%	-97,03%	-99,38%
Jan/3	432	11	0,00255%	0,23572%	1,13107%	-98,92%	-99,77%
Jan/4	516	9	0,00174%	0,23572%	1,13107%	-99,26%	-99,85%
Jan/5	754	26	0,00345%	0,23572%	1,13107%	-98,54%	-99,70%
Jan/6	761	41	0,00539%	0,23572%	1,13107%	-97,71%	-99,52%
Jan/7	731	30	0,00410%	0,23572%	1,13107%	-98,26%	-99,64%
Jan/8	771	18	0,00234%	0,23572%	1,13107%	-99,01%	-99,79%
Jan/9	647	44	0,00680%	0,23572%	1,13107%	-97,11%	-99,40%
Jan/10	428	16	0,00373%	0,23572%	1,13107%	-98,42%	-99,67%
Jan/11	754	54	0,00716%	0,23572%	1,13107%	-96,96%	-99,37%
Jan/12	740	24	0,00324%	0,23572%	1,13107%	-98,62%	-99,71%
Jan/13	728	48	0,00660%	0,23572%	1,13107%	-97,20%	-99,42%
Jan/14	676	21	0,00311%	0,23572%	1,13107%	-98,68%	-99,73%
Jan/15	743	25	0,00336%	0,23572%	1,13107%	-98,57%	-99,70%
Jan/16	316	32	0,01012%	0,23572%	1,13107%	-95,71%	-99,11%
Jan/17	770	88	0,01142%	0,23572%	1,13107%	-95,15%	-98,99%
Jan/18	791	70	0,00885%	0,23572%	1,13107%	-96,25%	-99,22%
Jan/19	436	57	0,01307%	0,23572%	1,13107%	-94,46%	-98,84%
Jan/20	599	240	0,04008%	0,23572%	1,13107%	-83,00%	-96,46%
Jan/21	750	998	0,13310%	0,23572%	1,13107%	-43,53%	-88,23%
Jan/22	466	1 006	0,21577%	0,23572%	1,13107%	-8,46%	-80,92%
Jan/23	621	2 500	0,40266%	0,23572%	1,13107%	70,82%	-64,40%
Jan/24	751	4 675	0,62287%	0,23572%	1,13107%	164,24%	-44,93%
Jan/25	781	6 854	0,87805%	0,23572%	1,13107%	272,50%	-22,37%
Jan/26	769	7 681	0,99891%	0,23572%	1,13107%	323,77%	-11,68%
Jan/27	749	4 417	0,58936%	0,23572%	1,13107%	150,03%	-47,89%
Jan/28	657	4 069	0,61906%	0,23572%	1,13107%	162,63%	-45,27%
Jan/29	758	5 066	0,66859%	0,23572%	1,13107%	183,64%	-40,89%
Jan/30	598	5 847	0,97737%	0,23572%	1,13107%	314,64%	-13,59%
Jan/31	381	3 352	0,88055%	0,23572%	1,13107%	273,56%	-22,15%

Table A1b. Statistics of collected tweets featuring #BLM during February

Date	All collected data (in thousands)	BLM related tweets	Daily share	average share (this month)	average share (half year)	daily difference compated to monthly average	daily difference compated to half year average
Feb/1	743	8	0,00108%	0,00165%	0,13053%	-34,79%	-99,18%
Feb/2	651	9	0,00138%	0,00165%	0,13053%	-16,15%	-98,94%
Feb/3	390	10	0,00256%	0,00165%	0,13053%	55,29%	-98,04%
Feb/4	747	13	0,00174%	0,00165%	0,13053%	5,40%	-98,67%
Feb/5	587	14	0,00239%	0,00165%	0,13053%	44,57%	-98,17%
Feb/6	388	7	0,00181%	0,00165%	0,13053%	9,44%	-98,62%
Feb/7	612	12	0,00196%	0,00165%	0,13053%	18,79%	-98,50%
Feb/8	386	7	0,00181%	0,00165%	0,13053%	9,98%	-98,61%
Feb/9	335	4	0,00119%	0,00165%	0,13053%	-27,74%	-99,09%
Feb/10	679	19	0,00280%	0,00165%	0,13053%	69,52%	-97,86%
Feb/11	768	21	0,00274%	0,00165%	0,13053%	65,80%	-97,90%
Feb/12	727	5	0,00069%	0,00165%	0,13053%	-58,34%	-99,47%
Feb/13	739	18	0,00244%	0,00165%	0,13053%	47,67%	-98,13%
Feb/14	738	8	0,00108%	0,00165%	0,13053%	-34,30%	-99,17%
Feb/15	518	3	0,00058%	0,00165%	0,13053%	-64,87%	-99,56%
Feb/16	653	4	0,00061%	0,00165%	0,13053%	-62,86%	-99,53%
Feb/17	751	7	0,00093%	0,00165%	0,13053%	-43,50%	-99,29%
Feb/18	692	5	0,00072%	0,00165%	0,13053%	-56,23%	-99,45%
Feb/19	748	8	0,00107%	0,00165%	0,13053%	-35,15%	-99,18%
Feb/20	745	6	0,00081%	0,00165%	0,13053%	-51,20%	-99,38%
Feb/21	568	5	0,00088%	0,00165%	0,13053%	-46,62%	-99,33%
Feb/22	550	6	0,00109%	0,00165%	0,13053%	-33,92%	-99,16%
Feb/23	761	16	0,00210%	0,00165%	0,13053%	27,39%	-98,39%
Feb/24	763	11	0,00144%	0,00165%	0,13053%	-12,67%	-98,90%
Feb/25	739	28	0,00379%	0,00165%	0,13053%	129,73%	-97,10%
Feb/26	736	21	0,00286%	0,00165%	0,13053%	73,03%	-97,81%
Feb/27	635	13	0,00205%	0,00165%	0,13053%	24,09%	-98,43%
Feb/28	443	8	0,00181%	0,00165%	0,13053%	9,54%	-98,62%
Feb/29	754	10	0,00133%	0,00165%	0,13053%	-19,61%	-98,98%

Table A2c. Statistics of collected tweets featuring #BLM during March

Date	All collected data (in thousands)	BLM related tweets	Daily share	average share (this month)	average share (half year)	daily difference compated to monthly average	daily difference compated to half year average
Mar/1	571	3	0,00053%	0,00119%	0,13053%	-55,78%	-99,60%
Mar/2	403	35	0,00868%	0,00119%	0,13053%	630,45%	-93,35%
Mar/3	770	39	0,00506%	0,00119%	0,13053%	325,90%	-96,12%
Mar/4	681	15	0,00220%	0,00119%	0,13053%	85,17%	-98,31%
Mar/5	399	6	0,00150%	0,00119%	0,13053%	26,37%	-98,85%
Mar/6	743	6	0,00081%	0,00119%	0,13053%	-32,03%	-99,38%
Mar/7	754	11	0,00146%	0,00119%	0,13053%	22,78%	-98,88%
Mar/8	764	7	0,00092%	0,00119%	0,13053%	-22,91%	-99,30%
Mar/9	744	5	0,00067%	0,00119%	0,13053%	-43,48%	-99,49%
Mar/10	783	8	0,00102%	0,00119%	0,13053%	-14,05%	-99,22%
Mar/11	332	1	0,00030%	0,00119%	0,13053%	-74,69%	-99,77%
Mar/12	661	4	0,00060%	0,00119%	0,13053%	-49,13%	-99,54%
Mar/13	766	5	0,00065%	0,00119%	0,13053%	-45,07%	-99,50%
Mar/14	786	1	0,00013%	0,00119%	0,13053%	-89,29%	-99,90%
Mar/15	761	4	0,00053%	0,00119%	0,13053%	-55,78%	-99,60%
Mar/16	732	13	0,00178%	0,00119%	0,13053%	49,49%	-98,64%
Mar/17	529	34	0,00643%	0,00119%	0,13053%	440,82%	-95,08%
Mar/18	589	6	0,00102%	0,00119%	0,13053%	-14,31%	-99,22%
Mar/19	760	1	0,00013%	0,00119%	0,13053%	-88,92%	-99,90%
Mar/20	741	2	0,00027%	0,00119%	0,13053%	-77,30%	-99,79%
Mar/21	759	5	0,00066%	0,00119%	0,13053%	-44,60%	-99,50%
Mar/22	749	7	0,00094%	0,00119%	0,13053%	-21,33%	-99,28%
Mar/23	693	1	0,00014%	0,00119%	0,13053%	-87,86%	-99,89%
Mar/24	323	2	0,00062%	0,00119%	0,13053%	-47,95%	-99,53%
Mar/25	807	5	0,00062%	0,00119%	0,13053%	-47,85%	-99,53%
Mar/26	787	1	0,00013%	0,00119%	0,13053%	-89,31%	-99,90%
Mar/27	770	5	0,00065%	0,00119%	0,13053%	-45,35%	-99,50%
Mar/28	790	7	0,00089%	0,00119%	0,13053%	-25,46%	-99,32%
Mar/29	744	3	0,00040%	0,00119%	0,13053%	-66,10%	-99,69%
Mar/30	425	2	0,00047%	0,00119%	0,13053%	-60,40%	-99,64%
Mar/31	663	3	0,00045%	0,00119%	0,13053%	-61,96%	-99,65%

Table A1d. Statistics of collected tweets featuring #BLM during April

Date	All collected data (in thousands)	BLM related tweets	Daily share	average share (this month)	average share (half year)	daily difference compated to monthly average	daily difference compated to half year average
Apr/1	740	3	0,00041%	0,00202%	0,13053%	-79,96%	-99,69%
Apr/2	773	3	0,00039%	0,00202%	0,13053%	-80,80%	-99,70%
Apr/3	762	1	0,00013%	0,00202%	0,13053%	-93,51%	-99,90%
Apr/4	772	5	0,00065%	0,00202%	0,13053%	-67,97%	-99,50%
Apr/5	537	3	0,00056%	0,00202%	0,13053%	-72,37%	-99,57%
Apr/6	192	4	0,00209%	0,00202%	0,13053%	3,14%	-98,40%
Apr/7	661	6	0,00091%	0,00202%	0,13053%	-55,14%	-99,30%
Apr/8	653	4	0,00061%	0,00202%	0,13053%	-69,69%	-99,53%
Apr/9	763	5	0,00066%	0,00202%	0,13053%	-67,61%	-99,50%
Apr/10	769	1	0,00013%	0,00202%	0,13053%	-93,57%	-99,90%
Apr/11	215	56	0,02602%	0,00202%	0,13053%	1186,76%	-80,06%
Apr/12	433	17	0,00393%	0,00202%	0,13053%	94,33%	-96,99%
Apr/13	408	9	0,00221%	0,00202%	0,13053%	9,05%	-98,31%
Apr/14	780	11	0,00141%	0,00202%	0,13053%	-30,23%	-98,92%
Apr/15	781	10	0,00128%	0,00202%	0,13053%	-36,69%	-99,02%
Apr/16	798	21	0,00263%	0,00202%	0,13053%	30,20%	-97,98%
Apr/17	639	29	0,00454%	0,00202%	0,13053%	124,42%	-96,52%
Apr/18	518	4	0,00077%	0,00202%	0,13053%	-61,85%	-99,41%
Apr/19	283	12	0,00424%	0,00202%	0,13053%	109,43%	-96,76%
Apr/20	802	9	0,00112%	0,00202%	0,13053%	-44,51%	-99,14%
Apr/21	782	21	0,00269%	0,00202%	0,13053%	32,83%	-97,94%
Apr/22	798	12	0,00150%	0,00202%	0,13053%	-25,62%	-98,85%
Apr/23	795	2	0,00025%	0,00202%	0,13053%	-87,57%	-99,81%
Apr/24	108	2	0,00185%	0,00202%	0,13053%	-8,51%	-98,58%
Apr/25	491	61	0,01242%	0,00202%	0,13053%	513,95%	-90,49%
Apr/26	804	29	0,00361%	0,00202%	0,13053%	78,26%	-97,24%
Apr/27	792	17	0,00215%	0,00202%	0,13053%	6,14%	-98,36%
Apr/28	803	7	0,00087%	0,00202%	0,13053%	-56,88%	-99,33%
Apr/29	796	6	0,00075%	0,00202%	0,13053%	-62,72%	-99,42%
Apr/30	478		0,00000%	0,00202%	0,13053%	-100,00%	-100,00%

Table A2e. Statistics of collected tweets featuring #BLM during May

Date	All collected data (in thousands)	BLM related tweets	Daily share	average share (this month)	average share (half year)	daily difference compated to monthly average	daily difference compated to half year average
May/1	611	26	0,00426%	0,12766%	0,13053%	-96,66%	-96,74%
May/2	772	71	0,00920%	0,12766%	0,13053%	-92,80%	-92,95%
May/3	782	215	0,02751%	0,12766%	0,13053%	-78,45%	-78,93%
May/4	783	50	0,00638%	0,12766%	0,13053%	-95,00%	-95,11%
May/5	768	11	0,00143%	0,12766%	0,13053%	-98,88%	-98,90%
May/6	598	4	0,00067%	0,12766%	0,13053%	-99,48%	-99,49%
May/7	583	20	0,00343%	0,12766%	0,13053%	-97,31%	-97,37%
May/8	761	27	0,00355%	0,12766%	0,13053%	-97,22%	-97,28%
May/9	759	11	0,00145%	0,12766%	0,13053%	-98,86%	-98,89%
May/10	787	13	0,00165%	0,12766%	0,13053%	-98,71%	-98,73%
May/11	785	12	0,00153%	0,12766%	0,13053%	-98,80%	-98,83%
May/12	624	4	0,00064%	0,12766%	0,13053%	-99,50%	-99,51%
May/13	532	8	0,00150%	0,12766%	0,13053%	-98,82%	-98,85%
May/14	787	11	0,00140%	0,12766%	0,13053%	-98,91%	-98,93%
May/15	766	4	0,00052%	0,12766%	0,13053%	-99,59%	-99,60%
May/16	885	9	0,00102%	0,12766%	0,13053%	-99,20%	-99,22%
May/17	818	5	0,00061%	0,12766%	0,13053%	-99,52%	-99,53%
May/18	674	3	0,00045%	0,12766%	0,13053%	-99,65%	-99,66%
May/19	595	5	0,00084%	0,12766%	0,13053%	-99,34%	-99,36%
May/20	651	6	0,00092%	0,12766%	0,13053%	-99,28%	-99,29%
May/21	455	3	0,00066%	0,12766%	0,13053%	-99,48%	-99,49%
May/22	783	12	0,00153%	0,12766%	0,13053%	-98,80%	-98,83%
May/23	816	12	0,00147%	0,12766%	0,13053%	-98,85%	-98,87%
May/24	638	3	0,00047%	0,12766%	0,13053%	-99,63%	-99,64%
May/25	621	5	0,00081%	0,12766%	0,13053%	-99,37%	-99,38%
May/26	761	447	0,05877%	0,12766%	0,13053%	-53,96%	-54,98%
May/27	782	2 898	0,37082%	0,12766%	0,13053%	190,46%	184,08%
May/28	288	3 651	1,26802%	0,12766%	0,13053%	893,24%	871,41%
May/29	759	9 845	1,29653%	0,12766%	0,13053%	915,58%	893,25%
May/30	563	6 132	1,08946%	0,12766%	0,13053%	753,38%	734,61%
May/31	341	3 450	1,01265%	0,12766%	0,13053%	693,21%	675,77%

Table A1f. Statistics of collected tweets featuring #BLM during June

Date	All collected data (in thousands)	BLM related tweets	Daily share	average share (this month)	average share (half year)	daily difference compated to monthly average	daily difference compated to half year average
Jun/1	404	5 619	1,39094%	0,74719%	0,13053%	86,16%	965,58%
Jun/2	737	13 723	1,86163%	0,74719%	0,13053%	149,15%	1326,16%
Jun/3	751	11 401	1,51885%	0,74719%	0,13053%	103,27%	1063,56%
Jun/4	730	10 410	1,42535%	0,74719%	0,13053%	90,76%	991,93%
Jun/5	721	8 166	1,13274%	0,74719%	0,13053%	51,60%	767,78%
Jun/6	706	9 263	1,31287%	0,74719%	0,13053%	75,71%	905,77%
Jun/7	353	5 307	1,50496%	0,74719%	0,13053%	101,42%	1052,93%
Jun/8	513	4 800	0,93534%	0,74719%	0,13053%	25,18%	616,54%
Jun/9	689	4 947	0,71825%	0,74719%	0,13053%	-3,87%	450,24%
Jun/10	706	5 567	0,78842%	0,74719%	0,13053%	5,52%	503,99%
Jun/11	660	5 382	0,81606%	0,74719%	0,13053%	9,22%	525,17%
Jun/12	708	5 229	0,73851%	0,74719%	0,13053%	-1,16%	465,76%
Jun/13	492	3 149	0,64033%	0,74719%	0,13053%	-14,30%	390,55%
Jun/14	502	2 740	0,54527%	0,74719%	0,13053%	-27,02%	317,72%
Jun/15	513	2 330	0,45414%	0,74719%	0,13053%	-39,22%	247,91%
Jun/16	631	2 915	0,46165%	0,74719%	0,13053%	-38,22%	253,66%
Jun/17	614	1 766	0,28769%	0,74719%	0,13053%	-61,50%	120,40%
Jun/18	611	2 010	0,32902%	0,74719%	0,13053%	-55,97%	152,06%
Jun/19	369	1 003	0,27160%	0,74719%	0,13053%	-63,65%	108,07%
Jun/20	541	1 777	0,32852%	0,74719%	0,13053%	-56,03%	151,67%
Jun/21	601	1 824	0,30358%	0,74719%	0,13053%	-59,37%	132,56%
Jun/22	134	398	0,29654%	0,74719%	0,13053%	-60,31%	127,17%
Jun/23	610	2 197	0,35989%	0,74719%	0,13053%	-51,83%	175,71%
Jun/24	615	2 224	0,36164%	0,74719%	0,13053%	-51,60%	177,04%
Jun/25	343	1 053	0,30686%	0,74719%	0,13053%	-58,93%	135,08%
Jun/26	336	1 149	0,34148%	0,74719%	0,13053%	-54,30%	161,60%
Jun/27	593	2 394	0,40359%	0,74719%	0,13053%	-45,99%	209,18%
Jun/28	479	1 390	0,29015%	0,74719%	0,13053%	-61,17%	122,28%
Jun/29	465	1 801	0,38735%	0,74719%	0,13053%	-48,16%	196,74%
Jun/30	340	1 111	0,32694%	0,74719%	0,13053%	-56,24%	150,46%

Appendix 3. Statistics of collected tweets featuring the Australian Bushfires

Table A3a. Statistics of collected tweets featuring the Australian bushfires during January

Date	All collected data (in thousands)	Australian Bushfires related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Jan/1	792	819	0,10340%	0,19967%	0,03955%	-48,21%	161,46%
Jan/2	429	917	0,21372%	0,19967%	0,03955%	7,04%	440,43%
Jan/3	432	1 925	0,44605%	0,19967%	0,03955%	123,40%	1027,90%
Jan/4	516	3 894	0,75471%	0,19967%	0,03955%	277,98%	1808,36%
Jan/5	754	7 553	1,00140%	0,19967%	0,03955%	401,53%	2432,14%
Jan/6	761	5 985	0,78673%	0,19967%	0,03955%	294,02%	1889,33%
Jan/7	731	3 719	0,50873%	0,19967%	0,03955%	154,79%	1186,38%
Jan/8	771	2 571	0,33361%	0,19967%	0,03955%	67,08%	743,57%
Jan/9	647	1 655	0,25582%	0,19967%	0,03955%	28,12%	546,86%
Jan/10	428	1 065	0,24857%	0,19967%	0,03955%	24,49%	528,53%
Jan/11	754	1 143	0,15152%	0,19967%	0,03955%	-24,11%	283,13%
Jan/12	740	1 484	0,20060%	0,19967%	0,03955%	0,47%	407,24%
Jan/13	728	956	0,13136%	0,19967%	0,03955%	-34,21%	232,15%
Jan/14	676	788	0,11665%	0,19967%	0,03955%	-41,58%	194,96%
Jan/15	743	818	0,11010%	0,19967%	0,03955%	-44,86%	178,39%
Jan/16	316	344	0,10877%	0,19967%	0,03955%	-45,53%	175,03%
Jan/17	770	592	0,07686%	0,19967%	0,03955%	-61,51%	94,34%
Jan/18	791	345	0,04362%	0,19967%	0,03955%	-78,15%	10,31%
Jan/19	436	151	0,03462%	0,19967%	0,03955%	-82,66%	-12,46%
Jan/20	599	212	0,03540%	0,19967%	0,03955%	-82,27%	-10,49%
Jan/21	750	229	0,03054%	0,19967%	0,03955%	-84,70%	-22,77%
Jan/22	466	176	0,03775%	0,19967%	0,03955%	-81,09%	-4,55%
Jan/23	621	385	0,06201%	0,19967%	0,03955%	-68,94%	56,80%
Jan/24	751	282	0,03757%	0,19967%	0,03955%	-81,18%	-5,00%
Jan/25	781	237	0,03036%	0,19967%	0,03955%	-84,79%	-23,23%
Jan/26	769	649	0,08440%	0,19967%	0,03955%	-57,73%	113,42%
Jan/27	749	608	0,08113%	0,19967%	0,03955%	-59,37%	105,14%
Jan/28	657	237	0,03606%	0,19967%	0,03955%	-81,94%	-8,82%
Jan/29	758	182	0,02402%	0,19967%	0,03955%	-87,97%	-39,26%
Jan/30	598	123	0,02056%	0,19967%	0,03955%	-89,70%	-48,01%
Jan/31	381	77	0,02023%	0,19967%	0,03955%	-89,87%	-48,85%

Table A3b. Statistics of collected tweets featuring the Australian bushfires during February

Date	All collected data (in thousands)	Australian Bushfires related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Feb/1	743	145	0,01950%	0,01527%	0,00925%	27,76%	110,92%
Feb/2	651	164	0,02521%	0,01527%	0,00925%	65,15%	172,64%
Feb/3	390	74	0,01896%	0,01527%	0,00925%	24,21%	105,06%
Feb/4	747	138	0,01846%	0,01527%	0,00925%	20,94%	99,66%
Feb/5	587	100	0,01704%	0,01527%	0,00925%	11,62%	84,27%
Feb/6	388	40	0,01032%	0,01527%	0,00925%	-32,40%	11,60%
Feb/7	612	102	0,01666%	0,01527%	0,00925%	9,14%	80,19%
Feb/8	386	86	0,02229%	0,01527%	0,00925%	46,05%	141,11%
Feb/9	335	67	0,01997%	0,01527%	0,00925%	30,83%	115,99%
Feb/10	679	83	0,01222%	0,01527%	0,00925%	-19,96%	32,15%
Feb/11	768	74	0,00964%	0,01527%	0,00925%	-36,85%	4,25%
Feb/12	727	81	0,01114%	0,01527%	0,00925%	-27,06%	20,42%
Feb/13	739	102	0,01381%	0,01527%	0,00925%	-9,55%	49,32%
Feb/14	738	116	0,01572%	0,01527%	0,00925%	2,97%	69,99%
Feb/15	518	64	0,01236%	0,01527%	0,00925%	-19,00%	33,72%
Feb/16	653	544	0,08333%	0,01527%	0,00925%	445,90%	801,23%
Feb/17	751	160	0,02131%	0,01527%	0,00925%	39,59%	130,45%
Feb/18	692	75	0,01083%	0,01527%	0,00925%	-29,03%	17,17%
Feb/19	748	53	0,00709%	0,01527%	0,00925%	-53,56%	-23,33%
Feb/20	745	42	0,00564%	0,01527%	0,00925%	-63,08%	-39,05%
Feb/21	568	35	0,00617%	0,01527%	0,00925%	-59,61%	-33,32%
Feb/22	550	51	0,00927%	0,01527%	0,00925%	-39,29%	0,23%
Feb/23	761	113	0,01485%	0,01527%	0,00925%	-2,75%	60,55%
Feb/24	763	57	0,00747%	0,01527%	0,00925%	-51,09%	-19,25%
Feb/25	739	34	0,00460%	0,01527%	0,00925%	-69,85%	-50,22%
Feb/26	736	36	0,00489%	0,01527%	0,00925%	-67,94%	-47,07%
Feb/27	635	111	0,01748%	0,01527%	0,00925%	14,52%	89,06%
Feb/28	443	57	0,01288%	0,01527%	0,00925%	-15,64%	39,27%
Feb/29	754	27	0,00358%	0,01527%	0,00925%	-76,54%	-61,27%

Table A3c. Statistics of collected tweets featuring the Australian bushfires during March

Date	All collected data (in thousands)	Australian Bushfires related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Mar/1	571	26	0,00456%	0,00587%	0,00925%	-22,35%	-50,73%
Mar/2	403	180	0,04465%	0,00587%	0,00925%	661,18%	382,93%
Mar/3	770	109	0,01415%	0,00587%	0,00925%	141,19%	53,03%
Mar/4	681	62	0,00910%	0,00587%	0,00925%	55,08%	-1,61%
Mar/5	399	31	0,00776%	0,00587%	0,00925%	32,29%	-16,07%
Mar/6	743	51	0,00687%	0,00587%	0,00925%	17,06%	-25,73%
Mar/7	754	31	0,00411%	0,00587%	0,00925%	-29,89%	-55,52%
Mar/8	764	28	0,00367%	0,00587%	0,00925%	-37,52%	-60,36%
Mar/9	744	35	0,00470%	0,00587%	0,00925%	-19,83%	-49,14%
Mar/10	783	23	0,00294%	0,00587%	0,00925%	-49,93%	-68,23%
Mar/11	332	16	0,00481%	0,00587%	0,00925%	-17,94%	-47,94%
Mar/12	661	116	0,01754%	0,00587%	0,00925%	198,92%	89,65%
Mar/13	766	31	0,00405%	0,00587%	0,00925%	-31,00%	-56,22%
Mar/14	786	44	0,00560%	0,00587%	0,00925%	-4,53%	-39,43%
Mar/15	761	30	0,00394%	0,00587%	0,00925%	-32,80%	-57,36%
Mar/16	732	19	0,00260%	0,00587%	0,00925%	-55,73%	-71,91%
Mar/17	529	20	0,00378%	0,00587%	0,00925%	-35,54%	-59,10%
Mar/18	589	76	0,01290%	0,00587%	0,00925%	119,94%	39,54%
Mar/19	760	38	0,00500%	0,00587%	0,00925%	-14,72%	-45,89%
Mar/20	741	27	0,00364%	0,00587%	0,00925%	-37,92%	-60,61%
Mar/21	759	28	0,00369%	0,00587%	0,00925%	-37,13%	-60,12%
Mar/22	749	42	0,00561%	0,00587%	0,00925%	-4,35%	-39,32%
Mar/23	693	10	0,00144%	0,00587%	0,00925%	-75,40%	-84,39%
Mar/24	323	11	0,00340%	0,00587%	0,00925%	-41,99%	-63,20%
Mar/25	807	14	0,00174%	0,00587%	0,00925%	-70,42%	-81,23%
Mar/26	787	20	0,00254%	0,00587%	0,00925%	-56,68%	-72,52%
Mar/27	770	15	0,00195%	0,00587%	0,00925%	-66,78%	-78,92%
Mar/28	790	31	0,00392%	0,00587%	0,00925%	-33,11%	-57,56%
Mar/29	744	21	0,00282%	0,00587%	0,00925%	-51,92%	-69,49%
Mar/30	425	16	0,00377%	0,00587%	0,00925%	-35,81%	-59,27%
Mar/31	663	18	0,00271%	0,00587%	0,00925%	-53,75%	-70,66%

Table A3d. Statistics of collected tweets featuring the Australian bushfires during April

Date	All collected data (in thousands)	Australian Bushfires related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Apr/1	740	22	0,00297%	0,00285%	0,00925%	4,38%	-67,85%
Apr/2	773	25	0,00324%	0,00285%	0,00925%	13,64%	-65,00%
Apr/3	762	31	0,00407%	0,00285%	0,00925%	42,76%	-56,03%
Apr/4	772	35	0,00453%	0,00285%	0,00925%	59,22%	-50,96%
Apr/5	537	16	0,00298%	0,00285%	0,00925%	4,63%	-67,78%
Apr/6	192	12	0,00626%	0,00285%	0,00925%	119,72%	-32,33%
Apr/7	661	23	0,00348%	0,00285%	0,00925%	22,13%	-62,39%
Apr/8	653	25	0,00383%	0,00285%	0,00925%	34,50%	-58,57%
Apr/9	763	20	0,00262%	0,00285%	0,00925%	-8,00%	-71,67%
Apr/10	769	20	0,00260%	0,00285%	0,00925%	-8,72%	-71,89%
Apr/11	215	4	0,00186%	0,00285%	0,00925%	-34,73%	-79,90%
Apr/12	433	6	0,00139%	0,00285%	0,00925%	-51,29%	-85,00%
Apr/13	408	9	0,00221%	0,00285%	0,00925%	-22,56%	-76,15%
Apr/14	780	21	0,00269%	0,00285%	0,00925%	-5,41%	-70,87%
Apr/15	781	10	0,00128%	0,00285%	0,00925%	-55,04%	-86,15%
Apr/16	798	23	0,00288%	0,00285%	0,00925%	1,27%	-68,81%
Apr/17	639	21	0,00329%	0,00285%	0,00925%	15,40%	-64,46%
Apr/18	518	11	0,00212%	0,00285%	0,00925%	-25,49%	-77,05%
Apr/19	283	3	0,00106%	0,00285%	0,00925%	-62,82%	-88,55%
Apr/20	802	21	0,00262%	0,00285%	0,00925%	-8,05%	-71,68%
Apr/21	782	10	0,00128%	0,00285%	0,00925%	-55,08%	-86,17%
Apr/22	798	18	0,00226%	0,00285%	0,00925%	-20,77%	-75,60%
Apr/23	795	15	0,00189%	0,00285%	0,00925%	-33,78%	-79,61%
Apr/24	108	1	0,00093%	0,00285%	0,00925%	-67,51%	-89,99%
Apr/25	491	38	0,00773%	0,00285%	0,00925%	171,59%	-16,35%
Apr/26	804	21	0,00261%	0,00285%	0,00925%	-8,33%	-71,77%
Apr/27	792	12	0,00152%	0,00285%	0,00925%	-46,79%	-83,61%
Apr/28	803	45	0,00561%	0,00285%	0,00925%	96,85%	-39,37%
Apr/29	796	15	0,00189%	0,00285%	0,00925%	-33,81%	-79,61%
Apr/30	478	6	0,00125%	0,00285%	0,00925%	-55,96%	-86,44%

Table A3e. Statistics of collected tweets featuring the Australian bushfires during May

Date	All collected data (in thousands)	Australian Bushfires related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
May/1	611	5	0,00082%	0,00262%	0,00925%	-68,77%	-91,14%
May/2	772	10	0,00130%	0,00262%	0,00925%	-50,60%	-85,99%
May/3	782	14	0,00179%	0,00262%	0,00925%	-31,68%	-80,63%
May/4	783	14	0,00179%	0,00262%	0,00925%	-31,85%	-80,67%
May/5	768	18	0,00234%	0,00262%	0,00925%	-10,64%	-74,66%
May/6	598	17	0,00284%	0,00262%	0,00925%	8,36%	-69,27%
May/7	583	9	0,00154%	0,00262%	0,00925%	-41,09%	-83,30%
May/8	761	4	0,00053%	0,00262%	0,00925%	-79,96%	-94,32%
May/9	759	7	0,00092%	0,00262%	0,00925%	-64,83%	-90,03%
May/10	787	11	0,00140%	0,00262%	0,00925%	-46,70%	-84,88%
May/11	785	8	0,00102%	0,00262%	0,00925%	-61,16%	-88,99%
May/12	624	2	0,00032%	0,00262%	0,00925%	-87,78%	-96,54%
May/13	532	7	0,00132%	0,00262%	0,00925%	-49,84%	-85,78%
May/14	787	3	0,00038%	0,00262%	0,00925%	-85,46%	-95,88%
May/15	766	14	0,00183%	0,00262%	0,00925%	-30,29%	-80,23%
May/16	885	6	0,00068%	0,00262%	0,00925%	-74,16%	-92,67%
May/17	818	6	0,00073%	0,00262%	0,00925%	-72,01%	-92,06%
May/18	674	7	0,00104%	0,00262%	0,00925%	-60,40%	-88,77%
May/19	595	5	0,00084%	0,00262%	0,00925%	-67,96%	-90,91%
May/20	651	9	0,00138%	0,00262%	0,00925%	-47,30%	-85,05%
May/21	455	49	0,01078%	0,00262%	0,00925%	310,96%	16,54%
May/22	783	31	0,00396%	0,00262%	0,00925%	50,95%	-57,19%
May/23	816	14	0,00172%	0,00262%	0,00925%	-34,57%	-81,45%
May/24	638	17	0,00267%	0,00262%	0,00925%	1,66%	-71,17%
May/25	621	20	0,00322%	0,00262%	0,00925%	22,86%	-65,16%
May/26	761	31	0,00408%	0,00262%	0,00925%	55,45%	-55,92%
May/27	782	26	0,00333%	0,00262%	0,00925%	26,88%	-64,02%
May/28	288	9	0,00313%	0,00262%	0,00925%	19,21%	-66,20%
May/29	759	22	0,00290%	0,00262%	0,00925%	10,49%	-68,67%
May/30	563	14	0,00249%	0,00262%	0,00925%	-5,14%	-73,10%
May/31	341	145	0,04256%	0,00262%	0,00925%	1523,14%	360,29%

Table A3f. Statistics of collected tweets featuring the Australian bushfires during June

Date	All collected data (in thousands)	Australian Bushfires related tweets	Daily share	Average share (this month)	Average share (half year)	Daily difference compated to monthly average	Daily difference compated to half year average
Jun/1	404	118	0,02921%	0,00302%	0,00925%	866,83%	215,90%
Jun/2	737	71	0,00963%	0,00302%	0,00925%	218,80%	4,16%
Jun/3	751	35	0,00466%	0,00302%	0,00925%	54,33%	-49,57%
Jun/4	730	27	0,00370%	0,00302%	0,00925%	22,36%	-60,02%
Jun/5	721	7	0,00097%	0,00302%	0,00925%	-67,86%	-89,50%
Jun/6	706	9	0,00128%	0,00302%	0,00925%	-57,78%	-86,20%
Jun/7	353	2	0,00057%	0,00302%	0,00925%	-81,23%	-93,87%
Jun/8	513	9	0,00175%	0,00302%	0,00925%	-41,95%	-81,03%
Jun/9	689	9	0,00131%	0,00302%	0,00925%	-56,75%	-85,87%
Jun/10	706	7	0,00099%	0,00302%	0,00925%	-67,19%	-89,28%
Jun/11	660	13	0,00197%	0,00302%	0,00925%	-34,76%	-78,68%
Jun/12	708	9	0,00127%	0,00302%	0,00925%	-57,93%	-86,25%
Jun/13	492	6	0,00122%	0,00302%	0,00925%	-59,62%	-86,81%
Jun/14	502	9	0,00169%	0,00302%	0,00925%	-44,00%	-81,70%
Jun/15	513	11	0,00214%	0,00302%	0,00925%	-29,03%	-76,81%
Jun/16	631	49	0,00776%	0,00302%	0,00925%	156,86%	-16,07%
Jun/17	614	79	0,01287%	0,00302%	0,00925%	325,97%	39,18%
Jun/18	611	18	0,00295%	0,00302%	0,00925%	-2,48%	-68,13%
Jun/19	369	5	0,00135%	0,00302%	0,00925%	-55,19%	-85,36%
Jun/20	541	13	0,00240%	0,00302%	0,00925%	-20,45%	-74,01%
Jun/21	601	5	0,00083%	0,00302%	0,00925%	-72,46%	-91,00%
Jun/22	134	1	0,00075%	0,00302%	0,00925%	-75,34%	-91,94%
Jun/23	610	14	0,00229%	0,00302%	0,00925%	-24,09%	-75,20%
Jun/24	615	12	0,00195%	0,00302%	0,00925%	-35,41%	-78,90%
Jun/25	343	7	0,00204%	0,00302%	0,00925%	-32,48%	-77,94%
Jun/26	336	11	0,00327%	0,00302%	0,00925%	8,21%	-64,64%
Jun/27	593	2	0,00034%	0,00302%	0,00925%	-88,84%	-96,35%
Jun/28	479	8	0,00157%	0,00302%	0,00925%	-48,18%	-83,07%
Jun/29	465	13	0,00280%	0,00302%	0,00925%	-7,43%	-69,75%
Jun/30	340	10	0,00280%	0,00302%	0,00925%	-7,47%	-69,77%