

How global crises compete for our attention: Insights from 13.5 million tweets on climate change during COVID-19

Tim Repke^{a,b,*}, Max Callaghan^{a,b}, William F. Lamb^{a,b,c}, Sarah Lück^{a,b}, Finn Müller-Hansen^{a,b}, Jan C. Minx^{a,b,c}

^aMercator Research Institute on Global Commons and Climate Change (MCC), EUREF-Campus 19, Berlin, 10829, Germany

^bPotsdam Institute for Climate Impact research (PIK), Telegrafenberg A 31, Potsdam, 14473, Germany

^cPriestley Centre, University of Leeds, Leeds, LS2 9JT, UK

Abstract

The COVID-19 pandemic disrupted peoples' daily lives and dominated the public discourse. It thus displaced people's attention to and concerns about climate change. We analyze 13.5 million tweets by 3.2 million distinct users on climate change posted before and after the onset of the pandemic (2018–2021) and show that attention to climate dropped substantially in 2020 with the onset of the pandemic. While research has helped to explain this drop in the context of issue attention theory, our analysis highlights a remarkable recovery in attention in 2021 towards pre-pandemic levels. Moreover, our large-scale, transformer-based text analysis reveals important thematic shifts during this period. In particular, we show a sustained drop in attention to activist movements and subsequently an increased focus on climate causes and climate solutions. Activist movements, such as the school protests that have mobilized millions around the globe in 2019, have measurably lost traction on Twitter. However, in parts due to increased awareness of causes and solutions, the climate change discourse in general recovered from the COVID-19 pandemic.

© 2024 Published by Elsevier Ltd.

Keywords: Climate Change, COVID-19 Pandemic, Public Attention, Social Media, Topic Model

1. Introduction

Public attention plays a key role in driving social change and shaping the political agenda. Researchers have therefore closely tracked attention to climate issues [1], including on social media [2], on the basis that continued public interest and concern provides an important foundation for climate action and policy [3]. Indeed, attention to climate change is constantly growing with an ever increasing amount of advocacy, reporting, and policies.

People feel inclined to voice their opinions on issues and share them with others. Social media platforms provide a simple and widely accessible forum for such exchange. Alternatively, they may demonstrate their support or opposition via street protests, petitions or referenda. In 2019, millions went to the streets and protested for climate action as part of the Fridays for Future protests and were key to bringing climate issue to the top of the political agenda.

The outbreak of the COVID-19 pandemic led to widespread disruptions—spurring public debates about disease control and

dealing with the social and economic repercussions of public health measures. This would seem to confirm issue attention theory: that our capacity to digest information is limited, and new salient issues tend to supersede prior ones [4, 5, 6]. If so, it may be challenging to maintain public attention on climate change, as new national and global issues are emerging all the time. Investigating the past allows us to better understand which topics withstand major shifts and may thus be most important to people, as these issues prevail within a finite pool of worry.

In this article, we analyse trends in public attention to climate on social media. Large-scale surveys are typically perceived as the gold standard to measure attitudes but require significant resources and take a long time. Complementary data-driven or bottom-up analyses of news reports, publications, social media, or protests can provide relevant indicators in a more timely fashion—a route that we take for Twitter.

Climate change on Twitter has been the subject of many studies [7], using methods such as sentiment analysis [8] and network analysis [9]. Recent studies show a decline of climate related tweets in many English and Spanish-speaking countries [10] and that they are largely displaced by COVID-19 is-

*Corresponding author: repke@mcc-berlin.net

sues in Switzerland [11]. This decline has been related to the increasing worry about COVID-19 [2?]. Loureiro et al. [?] provide a more in-depth overview of such worries expressed on Twitter early on during the pandemic. Climate action-related tweets saw a shift in focus from government actions to environmental behaviours [12]. Tweets on climate activism and public protests strongly decreased in Germany [13]. These studies only analyse higher-level quantitative shifts.

Our analysis of 13.5M tweets by 3.2M distinct users extends this literature by covering the entire period between 2018 and 2021 and providing further evidence that COVID-19 occupied public attention at the expense of climate issues on Twitter. To the best of our knowledge, such an in-depth topical analysis at this scale has not been done before. In particular, we improved an existing approach [14] in order to be able to perform this large-scale, transformer-based text analysis.

This enabled us to explore larger themes but also fine-grained topics over time and to show that the overall shift in attention impacted specific climate topics unequally. While the drop in attention associated with the onset of the pandemic was largely sustained for topics around activist movements, such as Fridays for Future, we observe an increased focus on climate causes and climate solutions as attention to climate issues recovers on Twitter.

2. Methods

The Twitter dataset used in this paper was retrieved via the official API (v2) between December 14th and 24th, 2021. We queried for tweets containing “climate change” posted between 2018 and 2021 which resulted in 20,213,783 tweets (excluding retweets). Our analysis is based on 13,506,789 tweets (66.82% of the original dataset) that were left after filtering (see appendix for exclusion rules). In this work we focus on the impact of the COVID-19 pandemic on the climate change discourse. To this end, we explicitly exclude 2022 with the invasion of Ukraine and resulting economic and energy crises. We chose the four year time-frame symmetrically centred around the start of the pandemic.

Although there are several tweets in our dataset that mention COVID-19, these are only the ones that also mention climate change. For a more complete picture, we additionally retrieved the total number of daily tweets on COVID-19 using the Twitter count API. Furthermore, we use this API to estimate the overall volume of daily English-language tweets.

The topic modelling approach for our analysis of climate change related tweets is based on the method described by BERTopic [14]. The general idea is to project high-dimensional document embeddings of the input texts into a lower-dimensional space and apply a clustering algorithm. Contrary to traditional topic models [15], each tweet is associated with exactly one topic (cluster) and is not represented as a distribution of topics. Scalability issues prevented us from using the reference implementation. To this end, we developed our own processing pipeline that enabled us to include all tweets in the topic model. Links to source code and data provided in supplemental material. Our implementation also allows us

to better control the grouping of tweets, which we adjust to be fine-grained enough to pick out small topics but also retain their wider semantic relationships.

For our analysis, we manually screened all 983 topics identified by the algorithm and deliberatively derived six broader themes: *Causes, Impacts, Solutions, Politics, Movements*, and *Contrarianism*. In addition, we group *COVID-19* topics and non-specific topics as *Other*. Note, that topics in the *COVID-19* theme were detected automatically by our model and are not linked to the COVID-19 tweet counts discussed before. The manual curation of topics into themes was performed by five annotators using a custom interface for screening a random sample of tweets per topic with additional statistics, the temporal distribution of tweets, and frequent keyphrases.

We provide further details on the datasets, topic modelling approach, and manual topic curation in the supplementary material.

3. Results

Our analysis of climate change tweets is two-fold. First, we describe the dataset from a quantitative perspective and show how attention to climate change—measured as tweets per day—has grown to an all-time high in 2019, dramatically declined with the onset of the pandemic, and almost fully rebounded in 2021. Secondly, we describe the thematic shifts based on our topic model, showing that activist movements were impacted most by the pandemic and focus shifted towards causes and solutions during the rebound.

Substantial drop in attention to climate change on Twitter

The COVID-19 pandemic struck at the all-time peak of attention to climate issues on Twitter. After a period of stagnant levels (average of 8,716 daily tweets 2016–2018), the number of climate-related tweets grew from an average of 7,980 daily tweets in 2018 to 15,914 tweets per day in 2019 (Figure 1a; Table 1)—higher than ever before in the history of the platform. We refer to the supplemental material for more details.

Table 1: Quarterly (Q1-4) and annual (YR) counts for tweets containing “climate change” as shown in Figure 1a

Year	Q1	Q2	Q3	Q4	YR
2018	6,544	5,439	7,070	12,810	7,980
2019	14,175	15,507	18,367	15,565	15,914
2020	14,780	6,603	10,865	10,274	10,630
2021	10,548	10,463	15,910	17,235	13,564

With the spread of the COVID-19 pandemic, there was a sharp and extended drop in attention to climate change on Twitter. This coincides with a surge of up to 1.5M tweets per day on COVID-19 in the early months of the pandemic (Figure 1c). Compared to 2019, the average daily number of tweets on climate change decreased by 58% to 6,603 in the second quarter

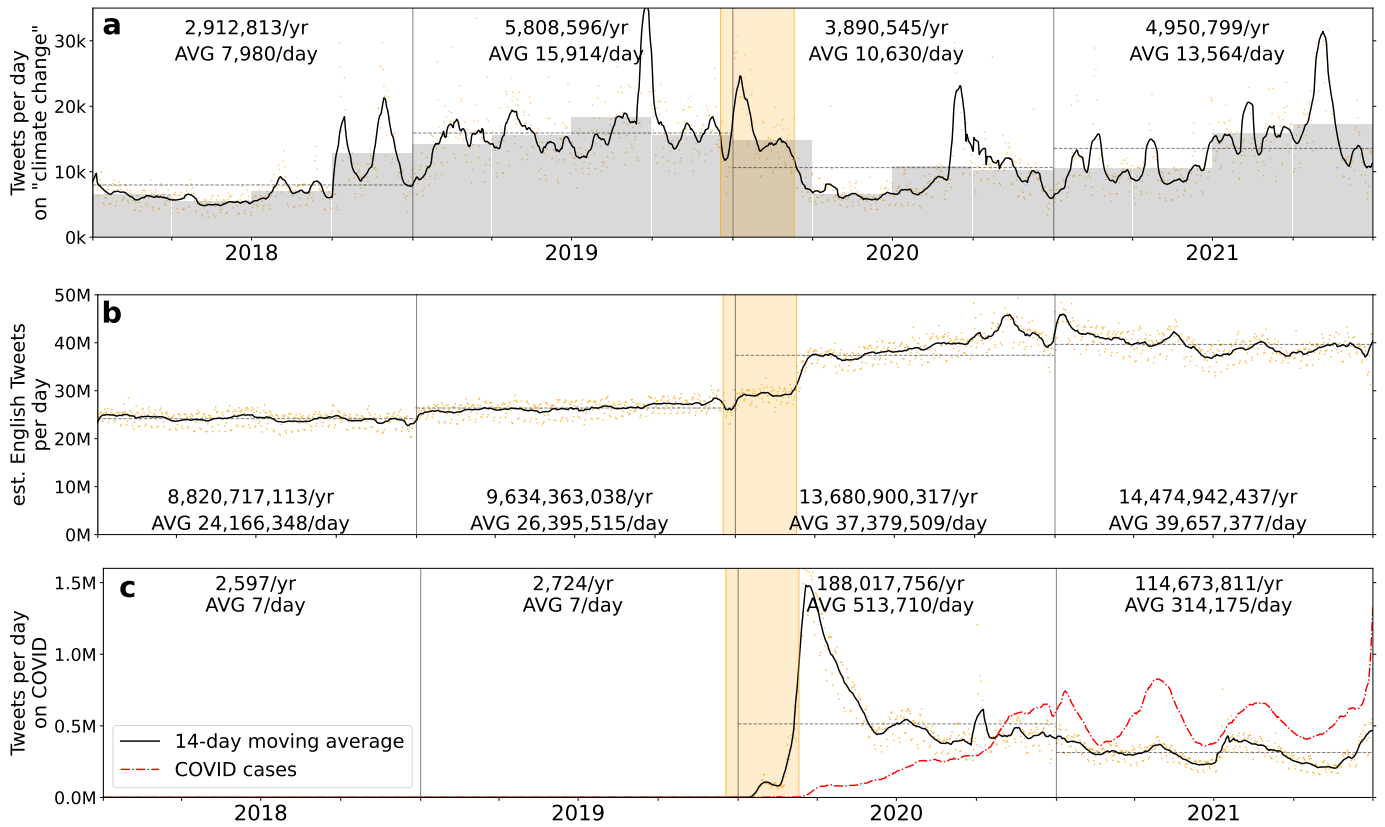


Figure 1: **a**: Number of daily tweets containing “climate change”, **c**: COVID-19, and **b**: estimated total number of English-language tweets. Orange dots show the number of tweets per day and the black line-plots their 14-day moving average. Solid vertical lines separate the years, the orange shaded area marks the time between the first COVID-19 case (2019-12-18) and the declaration as a pandemic by the WHO (2020-03-11). The sum and average of daily tweets is shown in numbers per year, the averages is drawn as horizontal dashed lines. Grey bars in panel **a** additionally show the quarterly averages. The red dash-dot line plot in panel **c** shows the world-wide new COVID cases per day (same axis).

of 2020, coinciding with the start of lockdown policies in several countries. Lower levels of attention remain for the second half of 2020 (daily average in Q3 at 68.3% and Q4 at 64.6% of 2019 average) and continue into the first half of 2021 (Q1 at 66.3% and Q2 at 65.7% of 2019 average). This drop in attention at the beginning of the COVID-19 pandemic takes places despite a simultaneous, sharp increase in the number of overall English-language tweets by 40% of 2019-levels and remains high as shown in Figure 1b. In the second half of 2021, attention rebounds to 2019 levels by matching the 2019 annual average of daily tweets in Q3 2021 (+49% of 2020 average) and surpassing it by 8% in Q4 2021 (+62.1% of 2020 average). Overall, there were an average number of 13,596 tweets per day in 2021. Hence, attention in 2021 was not much lower than in 2019 and considerably higher than in most other years before the pandemic—an important observation that has been largely overlooked in the related literature [2, 13, 12, 11, 10].

Divergence in the recovery for different themes

We explore how the upheavals during COVID-19 have affected the themes in the climate conversation by combining unsupervised text mining—clustering of BERT [14] embeddings reduced in dimensionality—with rigorous manual validation and annotation. We group the 983 topics identified by

the algorithm into six broader themes: *Causes, Impacts, Solutions, Politics, Movements, and Contrarianism*. Two supplemental themes group *COVID-19* topics or non-specific topics as *Other*. Thematic shifts discussed in this section are based on the daily proportion of tweets per theme as shown in Figure 2. The proportions are relative to the daily number of tweets on “climate change”, excluding those assigned to the *Other* theme. Alternative frames of reference are in the appendix.

In the first half of 2020, up to 10% of the attention of climate change tweets rapidly shifts to the newly emerged *COVID-19* theme (2.2% overall in 2020 and 1.3% in 2021). After the initial drop during 2020 Q2, the subsequent trends of each theme diverge: *Causes, Impacts, and Solutions* have tended to increase, while *Politics, Movements, and Contrarian* have tended to decrease.

For the years leading up to the pandemic, attention to the *Movements* theme saw steady growth, reaching an all time high of 5.9-6.2% in Q3/Q4 2019 following climate impact events (e.g. Australian wild fires, July 2020), political debates (e.g. Greta’s speech at UN: “How dare you!”, Sept. 2019), as well as climate protests (e.g. Extinction rebellion in London, Oct. 2019). However, presumably due to protests and meetings in person no longer being possible, attention dropped in Q1/Q2 2020 to a below average share of 3.1%. Even though the situ-

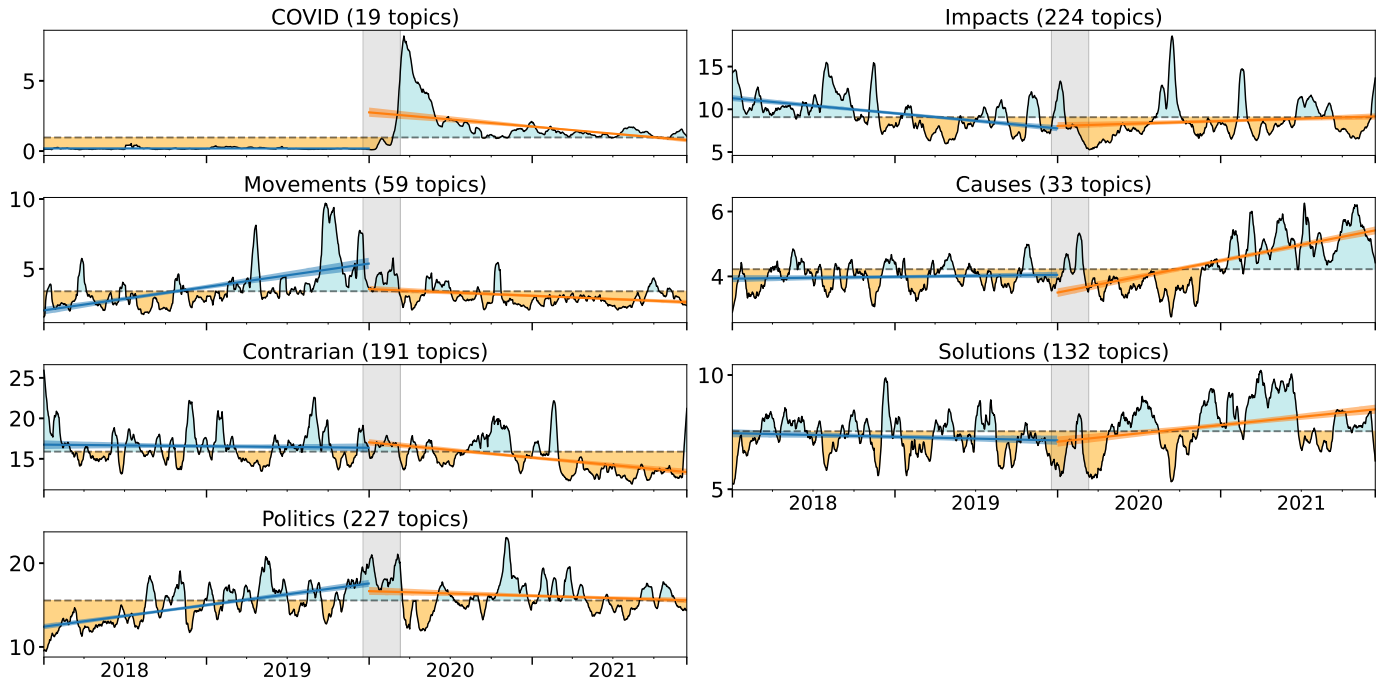


Figure 2: Daily share of tweets on climate change per theme (in %). The dashed horizontal line indicates the overall share of tweets contained per theme and the areas between the curves and that line are shaded to highlight the deviations of attention over time. Larger shifts are shown as linear trend lines with a standard deviation band of the years before (blue) and after the onset of the pandemic (orange). The grey shaded area marks the time between the first COVID-19 case (2019-12-18) and the declaration as a pandemic by the WHO (2020-03-11). Note, that the vertical scale is not shared across panels. All plots are smoothed with a 14-day moving average. Values across panels for a particular day may not add up to 100% as the *Other* theme, to which 35.1% of all tweets are assigned, is not shown here.

ation began to normalise in many countries, we are not able to identify signs of major recovery. Note, that the observed timeframe does not include most recent “Last Generation” protests.

The two largest themes in our dataset are *Contrarian* and *Politics* with each 16% of all tweets. The *Contrarian* theme covers tweets expressing skepticism or denial of the scientific consensus on climate change. Its annual share dropped from 17% 2018-2020 to 14.2% in 2021 and the main topics are the denial of climate change or discourses of delay [16]. This sudden reduction in attention can be attributed to controversies related to Donald Trump, which dropped in 2021 after he left office. In the years before the pandemic, *Politics* has shown a similar growth pattern as *Movements*, with peaks in attention slightly behind. Topics mainly cover policy initiatives, reactions to statements by politicians, but also political aspects of climate-related disasters, which remain a major driver of climate change debate. After a steady growth to a quarterly share of 18.5% in 2020 Q1, attention dropped to the level of 2018. The share temporarily recovered to 18% in 2020 Q4 and 2021 Q1 driven by topics covering the US presidential elections and winter storms in Texas, before declining back to around 15.3%.

The third largest is the *Impacts* theme (9% of all tweets) where we observe substantial shifts in the long-term dynamics of the thematic distribution. Attention already declined before the pandemic from 10.5% to 8.2% in 2019 and stabilizes at an annual average of 9% after the pandemic. Tweets on extreme temperatures (heat and cold) lead to seasonality effects with Q1 and Q3 being the strongest quarters in all four years (1-2 per-

centage points above yearly average). In general, we noticed a shift from “distant” impacts on animals (such as diminishing habitats for ice bears and penguins) to massive natural disasters like floods, wild fires, or droughts. Although 2021 was a record year in climate-related disasters [17], this is not reflected in the attention to the *Impacts* theme, indicating that Twitter users seem to have other priorities in the climate change discourse, or that many regional events are not picked up in our English-language sample.

The *Causes* theme had a share of 3.9% before the pandemic, dropping to an all time low of 3.3% in 2020 Q3 to then continuously grow to 5.6% in 2021 Q4 (5.2% overall in 2021). The highest co-occurring keywords and phrases of topics and manual screening, *Causes* themed tweets tend to blame capitalism and corporate greed as the culprit for climate change or delay of climate action. In 2021, more specific topics appeared or grew, for example fracking, use of private jets, and court cases against Exxon Mobil condemning environmental impacts caused by them.

Attention to *Solutions* grew from 7.1% of annual tweets in prior years to 8% in 2021. *Solutions* themed tweets often discussed specific tree planting projects, philanthropists pledging to donate money, or celebrities raising awareness. The overall trend was slightly negative before the pandemic, although the theme displayed a particularly high share in Q4 2020 to Q2 2021, which has subsequently declined.

4. Discussion

We provide new evidence supporting the hypothesis that the COVID-19 pandemic distracted attention away from climate change. While it is crucial to discuss the attention drop [18, 2, 19, 10? ?], the remarkable recovery towards high pre-pandemic levels in Q4 2021 has been neglected so far. However, these trends on Twitter should not be unconditionally extrapolated to the general public—particularly in a global context, since we only looked at English-language tweets. Also, active Twitter users may not be representative sample of the general public. Our analysis based on 13.5M tweets posted by 3.2M unique Twitter users across four years. Given this scale, and the rigorous validation of our topic model, we are confident in the findings we presented. Furthermore, we carefully chose to only report on strong, large-scale trends we discovered.

The recovery after the COVID-19-related drop was accompanied by important thematic changes of tweets on climate change, such as a stronger focus on climate *Solutions* and an upward trend in *Causes* as a theme. This may indicate increasing public discontent with those seen to be causing the climate crisis and a greater public engagement with the steps governments are taking to mitigate climate change. A renewed focus on climate solutions, particularly in the period during which COVID-19 recovery packages were being discussed, may demonstrate a greater public engagement with the steps governments are taking to mitigate climate change.

Our dataset as a whole is similar to datasets that have been provided previously [20] and we are confident in the general trends observed. Whereas other works mostly focus on quantitative analyses, we employed an advanced, transformer-based topic modelling technique. This allows us to efficiently allocate tweets to topics at a highly granular level, capturing discourses and mini-discourses around isolated events. Future research should adapt these approaches to explore other social media sources and develop a more comprehensive view of public discourses on climate.

The keyword-based search commonly used in the literature has important limitations, as climate relevant discourses can also take place in the absence of these keywords. Extending the list of keywords will likely never be complete, while, at the same time also capturing an increasing number of non-related tweets, which would obfuscate the analysis. Hence we argue, that any social media analysis—given the chosen keywords are specific, yet general enough—can only provide a snapshot of the wider discourse on a particular topic. However, it can still serve as an indicator of the relative volume of posts and reveal relevant themes. Importantly, other research using Twitter data is often using hashtags to query for tweets. This may result in a wider range of languages and broader geographic coverage. At the same time, it also picks up significantly more automated tweets or attention-seeking tweets that use trending hashtags for an improved change of increasing visibility.

In order to provide a clear understanding of the impact of COVID-19 on attention to climate change, we deliberately focus our analysis on the two years before and after the onset of the pandemic. We do not extend our analysis into 2022 in light

of the war in Ukraine as the related energy crisis has strongly affected the climate discourse. The impact of the Ukraine conflict on energy systems around the world has precipitated shifts in new investments in both fossil fuel and green infrastructure which have profound implications for international climate goals. Any signal (e.g. number of tweets on a topic) derived from the analysis of data after 2021 would come with large uncertainties, as user behaviour may have changed due to a multitude of external factors, particularly the many other geopolitical conflicts that have since unfolded and acquisition of the Twitter (now X) by Elon Musk in 2022, which has significantly impacted how and which people use the platform.

Understanding how this subsequent crisis has affected attention and attitudes towards climate change is an important matter for future research.

Data Availability

The source code for all our analyses is available on GitHub via <https://github.com/TimRepke/twitter-climate>. The list of tweet ids, topic assignments, and our topic to theme annotations is available on zenodo via <https://doi.org/10.5281/zenodo.7778199>.

Acknowledgments

This project received funding from the European Union's Horizon 2020 research and innovation programme under the European Research Council (ERC) Grant Agreement No. 951542-GENIE-ERC-2020-SyG, "GeoEngineering and Negative Emissions pathways in Europe" (GENIE) and from the German Federal Ministry of Education and Research under the ARIADNE project (grant no. 03SFK5J0). The content of this deliverable does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the author(s).

Authorship contribution statement

Data curation: F.M.-H., M.C.; Methodology and software: T.R.; Formal analysis of topic model: T.R.; Annotation of topics to themes: T.R., M.C., S.L., F.M.-H.; Visualisation: T.R., F.M.-H.; Writing (original draft): T.R., J.M., F.M.-H., W.L.; Writing (reviewing & editing): all

References

- [1] P. Hart, E. Nisbet, T. Myers, Public attention to science and political news and support for climate change mitigation, *Nature Climate Change* (5) (2015) 541–545. doi:10.1038/nclimate2577.
- [2] O. Smirnov, P.-H. Hsieh, Covid-19, climate change, and the finite pool of worry in 2019 to 2021 twitter discussions, *Proceedings of the National Academy of Sciences* 119 (43) (2022) e2210988119.
- [3] J. D. Farmer, C. Hepburn, M. C. Ives, T. Hale, T. Wetzer, P. Mealy, R. Rafaty, S. Srivastav, R. Way, Sensitive intervention points in the post-carbon transition, *Science* 364 (6436) (2019) 132–134.
- [4] A. Downs, Up and down with ecology—the "issue-attention" cycle, *Public Interest* 28 (1972) 38–50.

- [5] J.-H. Zhu, Issue Competition and Attention Distraction: A Zero-Sum Theory of Agenda-Setting, *Journalism Quarterly* 69 (4) (1992) 825–836, publisher: SAGE Publications. doi:10.1177/107769909206900403.
- [6] H.-B. Brosius, M. H. Kepplinger, KILLER AND VICTIM ISSUES: ISSUE COMPETITION IN THE AGENDA-SETTING PROCESS OF GERMAN TELEVISION, *International Journal of Public Opinion Research* 7 (3) (1995) 211–231. doi:10.1093/ijpor/7.3.211.
- [7] J. R. Fownes, C. Yu, D. B. Margolin, Twitter and climate change, *Sociology Compass* 12 (2018) e12587. doi:10.1111/soc4.12587.
- [8] E. M. Cody, A. J. Reagan, L. Mitchell, P. S. Dodds, C. M. Danforth, Climate change sentiment on Twitter: An unsolicited public opinion poll, *PLoS ONE* 10 (8) (2015) 0–18. arXiv:1505.03804, doi:10.1371/journal.pone.0136092.
- [9] H. T. P. Williams, J. R. McMurray, T. Kurz, F. H. Lambert, Network analysis reveals open forums and echo chambers in social media discussions of climate change, *Global Environmental Change* 32 (2015) 126–138. doi:10.1016/j.gloenvcha.2015.03.006.
- [10] M. L. Loureiro, M. Alló, How has the COVID-19 pandemic affected the climate change debate on Twitter?, *Environmental Science and Policy* 124 (2021) 451–460. doi:10.1016/j.envsci.2021.07.011.
- [11] A. Rauchfleisch, D. Siegen, D. Vogler, How COVID-19 Displaced Climate Change: Mediated Climate Change Activism and Issue Attention in the Swiss Media and Online Sphere, *Environmental Communication* 0 (0) (2021) 1–9. doi:10.1080/17524032.2021.1990978.
- [12] M. Gaytan Camarillo, E. Ferguson, V. Ljevar, A. Spence, Big Changes Start With Small Talk: Twitter and Climate Change in Times of Coronavirus Pandemic, *Frontiers in Psychology* 12 (June) (2021). doi:10.3389/fpsyg.2021.661395.
- [13] J. Haßler, A. K. Wurst, M. Jungblut, K. Schlosser, Influence of the pandemic lockdown on Fridays for Future's hashtag activism, *New Media and Society* (2021). doi:10.1177/14614448211026575.
- [14] M. Grootendorst, BERTopic: Neural topic modeling with a class-based TF-IDF procedure, *CoRR abs/2203.05794* (2022). arXiv:2203.05794, doi:10.48550/arXiv.2203.05794.
- [15] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, *Journal of machine Learning research* 3 (2003) 993–1022.
- [16] W. F. Lamb, G. Mattioli, S. Levi, J. T. Roberts, S. Capstick, F. Creutzig, J. C. Minx, F. Müller-Hansen, T. Culhane, J. K. Steinberger, Discourses of climate delay, *Global Sustainability* 3 (2020) e17. doi:10.1017/sus.2020.13.
- [17] S. A. Mazhin, M. Farrokhi, M. Noroozi, J. Roudini, S. A. Hosseini, M. E. Motlagh, P. Kolivand, H. Khankeh, Worldwide disaster loss and damage databases: A systematic review, *Journal of education and health promotion* 10 (2021). doi:10.4103/jehp.jehp_1525_20.
- [18] D. Evensen, L. Whitmarsh, P. Bartie, P. Devine-Wright, J. Dickie, A. Varley, S. Ryder, A. Mayer, Effect of “finite pool of worry” and covid-19 on uk climate change perceptions, *Proceedings of the National Academy of Sciences* 118 (3) (2021). doi:10.1073/pnas.2018936118.
- [19] M. Boykoff, M. Aoyagi, A. Benham, P. Chandler, M. Daly, K. Doi, R. Fernández-Reyes, E. Hawley, L. McAllister, M. McNatt, et al., World newspaper coverage of climate change or global warming, 2004–2022, *Cooperative Institute for Research in Environmental Sciences, University of Colorado* (2022). doi:10.25810/4c3b-b819.46.
- [20] D. Effrosynidis, A. I. Karasakalidis, G. Sylaios, A. Arampatzis, The climate change twitter dataset, *Expert Systems With Applications (ESWA)* 204 (117541) (2022). doi:10.1016/j.eswa.2022.117541.
- [21] R. Kouzy, J. Abi Jaoude, A. Kraitam, M. B. El Alam, B. Karam, E. Adib, J. Zarka, C. Traboulsi, E. Akl, K. Baddour, Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter, *Cureus* 12 (3) (2020). doi:10.7759/cureus.7255.
- [22] M. O. Lwin, J. Lu, A. Sheldenkar, P. J. Schulz, W. Shin, R. Gupta, Y. Yang, Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends, *JMIR Public Health and Surveillance* 6 (2) (2020) 1–4. doi:10.2196/19447.
- [23] K. Garcia, L. Berton, Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA, *Applied Soft Computing* 101 (2021) 107057. doi:10.1016/j.asoc.2020.107057.
- [24] L. v. d. Maaten, G. Hinton, Visualizing data using t-SNE, *Journal of Machine Learning Research (JMLR)* 9 (2008) 2579–2605.
- [25] D. Kobak, G. Linderman, S. Steinerberger, Y. Kluger, P. Berens, Heavy-tailed kernels reveal a finer cluster structure in t-SNE visualisations, in: U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, C. Robardet (Eds.), *Proceedings of the European Conference on Machine Learning (ECML)*, Springer-Verlag, 2019, pp. 124–139. doi:10.1007/978-3-030-46150-8_8.
- [26] L. McInnes, J. Healy, UMAP: uniform manifold approximation and projection for dimension reduction, *CoRR abs/1802.03426* (2018). arXiv:1802.03426.
- [27] R. González-Márquez, P. Berens, D. Kobak, Two-dimensional visualization of large document libraries using t-SNE, in: A. Cloninger, T. Doster, T. Emerson, M. Kaul, I. Ktena, H. Kvinge, N. Miolane, B. Rieck, S. Tymochko, G. Wolf (Eds.), *Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022*, Vol. 196 of *Proceedings of Machine Learning Research*, PMLR, 2022, pp. 133–141.
- [28] L. McInnes, J. Healy, Accelerated hierarchical density based clustering, in: R. Gottumukkala, X. Ning, G. Dong, V. Raghavan, S. Aluru, G. Karypis, L. Miele, X. Wu (Eds.), *IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, 2017, pp. 33–42. doi:10.1109/ICDMW.2017.12.

Appendix A. The climate change tweets dataset

The Twitter dataset used in this paper was retrieved via the official API (v2) between December 14th and 24th, 2021. Missing tweets for the remainder of 2021 were added in October of 2022, tweet counts for 2010–2022 were retrieved in December of 2022. Even though we focus on 2018–2021 in this paper, we looked at the wider context to verify that the observed time-frame is not an “outlier” in the overall timeline. We queried for “climate change” from 2018 to 2021 which resulted in 20,213,783 tweets (excluding retweets). For our analyses, we filtered this dataset down to 13,506,789 tweets (66.82% of the original dataset) using the following rules:

- 3,633,551 (17.98%) tweets were duplicates, meaning they had the exact same text.
- 2,594,789 (12.84%) tweets did not contain our search term directly, but it was part of a URL.
- 881,409 (4.36%) tweets were not written in English.
- 451,554 (2.23%) tweets were too short, meaning they had less than four tokens.
- 215,860 (1.07%) tweets had more than five hashtags. Based on our preliminary manual analysis, such tweets provide very little value and could mostly be considered as “spam”.

Note, that multiple rules may apply to a single tweets, thus these numbers may not add up to the actual difference. In order to verify we do not introduce biases in the temporal distribution of our dataset, we computed the monthly share of tweets that are affected by these filtering rules. This share is consistent throughout time for each rule and the combined rule-set.

We compared our simple query with more comprehensive queries from the academic literature [?] and non-academic datasets (e.g. <https://www.kaggle.com/leonshangguan/climate-change-tweets-ids-until-aug-2021>) that are comprised of several climate related hashtags and other search terms. This resulted in slightly higher numbers of tweets (28,088,904 tweets compared to our 20,213,783 tweets). However, we found that our simple search strategy led to very similar time series (Pearson correlation coefficient: 0.990). The more recently published climate change twitter dataset [20] only contains 15M tweets spanning 13 years.

Appendix B. Additional tweet counts

Twitter does not officially provide the total number of daily tweets. In order to distinguish larger trends in the usage of the platform from actual changes in attention to climate change, we develop an indicator that estimates the daily number of English-language tweets. To do so, we query “-wrbvjxxuqgekwmvouvfhogx (the OR be OR ...) lang:en -is:retweet -is:quote” the academic Twitter /2/tweets/counts/all API for all English-language tweets that contain one or more of the top 100 most used words

in English, according to the Oxford English Corpus. Since the API requires at least one non-stopword, we add a negative clause for a random string of characters. Furthermore, we exclude retweets and quoted tweets to dismiss duplicates, analogous to our dataset of “climate change” tweets.

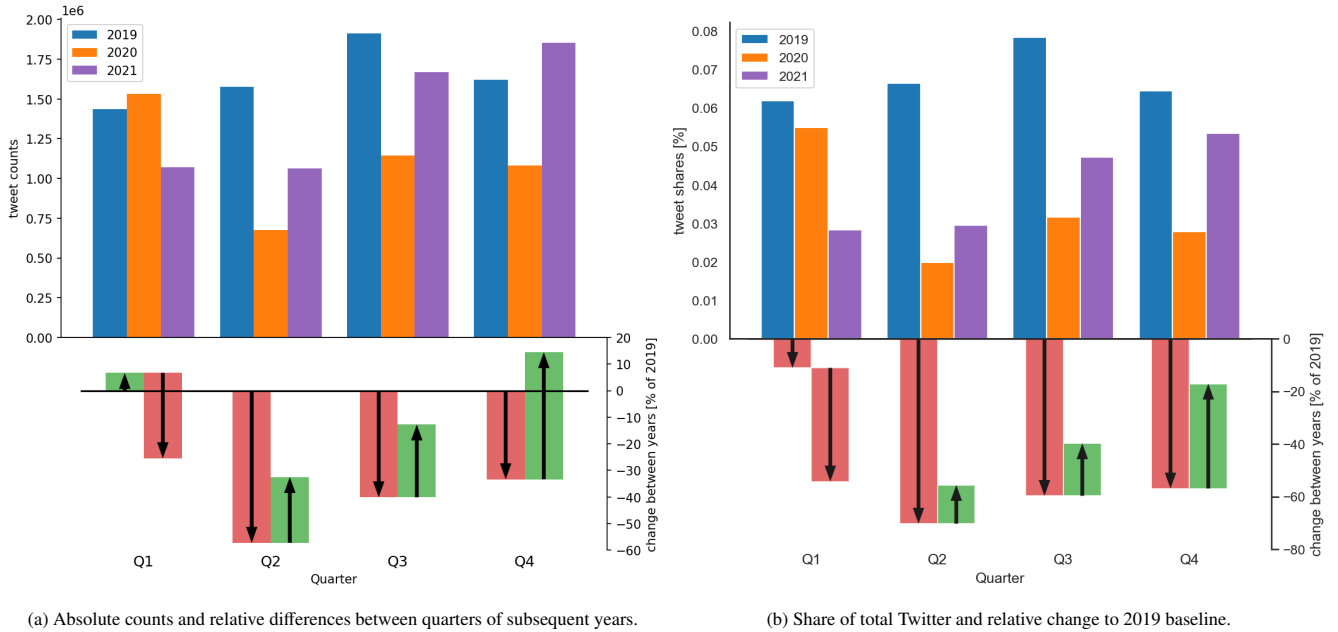
Similarly, we estimate the number of tweets related to COVID-19 by querying the Twitter /counts API for “COVID”, “COVID-19”, “COVID19”, “ncov” (which stands for novel COVID), “ncov19”, “ncov2019”, “sars-cov-2”, “sarscov2”, “wuhanvirus” (used only in [21]), other keywords such as “corona” have been shown to yield too many false positives, also “wuhan” (used by [22]), “quarantine” (used in [23]), “pandemic”, which is too general. The number of news articles on climate change and global warming was derived from the dataset compiled by [19].

Appendix C. Topic modelling based on BERT

The topic modelling approach for our analysis of climate change related tweets is based on the method described by BERTopic [14]. The general idea is to project document embeddings of the input documents into a lower dimensional space and apply a clustering algorithm. Contrary to traditional topic models, each tweet is associated with exactly one topic (cluster) and is not represented as a distribution of topics. Scalability issues prevented us from using the reference implementation. To this end, we developed our own processing pipeline in order to include all tweets in the topic model.

After embedding all 13.5M tweets using a pre-trained transformer model, we selected every second tweet (ordered by time) and applied tSNE [24, 25] in two stages: First, a random sample of 150k tweets from the subsampled 7M tweets is used to initialise the projection two-dimensional space with a perplexity of 300 and a high learning rate. The affinities of the remaining tweets are computed and projected into the initialised space, which is then further fine-tuned with a lower perplexity of 20 and a small learning rate. By setting the degrees of freedom to 0.6, the resulting two-dimensional layout of the data results in more dense clusters with a larger distance between clusters. The dimensionality reduction helps to improve the performance of the clustering algorithm and limit the impact of potentially over-specific document embeddings. Originally, BERTopic uses UMAP [26] to project the embeddings into a lower-dimensional space. We opted to use tSNE for its ability to better represent large document collections [27].

In order to find these clusters, we use HDBSCAN [28]. Each cluster found by HDBSCAN is then interpreted as a topic. We determined the ideal hyper-parameters by performing a parameter sweep with the goal of limiting the number of tweets HDBSCAN identified as outliers and keeping the size of the largest topic as small as possible. We found a cluster selection $\epsilon = 0.009$, minimum samples for cluster seeding 10 and minimum cluster (leaf) size of 200 to work best by performing a parameter sweep with the goal of keeping the number of tweets HDBSCAN identified as outliers small as well as keeping the size of the largest topic as small as possible. The previously excluded second half of the dataset is merged into the



Supplementary Figure S1: Quarterly number of tweets on climate change.

model by assigning tweets to clusters by finding their 30 nearest neighbours in the embedding space and determining the majority class. Alternatively using the nearest cluster centroid did not provide satisfying results.

Appendix D. Manual curation of topics

Our topic model, described in the previous section, identified 983 clusters. In order to draw any higher-level conclusions from that, we manually annotated each topic by assigning it to a theme. The themes were determined by consolidating the candidates proposed in an initial open-domain round of annotations by three annotators. Following that, each topic was labelled by one of six annotators based on reading a sample of tweets from that topic to one of the eight themes: COVID-19, politics, movements, impacts, causes, solutions, contrarian, and others. We also marked 113 topics (1,025,877 tweets, 7.6% of the dataset) as not relevant. Most of those are considered spam as the topics contained virtually identical tweets with only slight variations in wording, clearly indicating bot activity across a few days, months, and sometimes even years. Supplementary Table S2 provides an overview of the number of topics per theme, the average topic sizes, number of tweets, and at what point in the observed time-frame the most tweets were posted. It is important to note, that our dataset does not allow any comparative statements of how the pandemic and the climate crisis are discussed on Twitter. Instead we can only explore tweets in the COVID-19 theme *in the context of* tweets that mention climate change.

Appendix E. User behaviour

Supplementary Figure S7 shows ratios between the cumulative number of tweets and the cumulative number of users over

time per theme. A steep incline of the curve indicates that a fixed group of users is posting regularly, while convergence or flattening of the curve indicates that only few users tweet repeatedly. Note, that users may be more active on the platform in general, but we are limited to the perspective of only including tweets mentioning climate change. Only few users tweet repeatedly about the pandemic in the context of climate change. For most of the themes, the number of tweets per user initially grows but converges over time. Interestingly, this ratio climbs the highest and the longest in the politics and contrarian themes. With the onset of the pandemic, the growth stagnates for over a year until it starts growing again towards the end of 2021. In contrast, the ratio for tweets on movements initially grows fast but then quickly converges to 1.5 tweets per user.

Supplementary Figure S6 shows the overlapping themes users post in. note, that some topics are labelled with more than one theme, thus there's an overlap by double-counting

Supplementary Table S1: Quarterly (Q1–4) and annual (YR) tweet counts for “climate change” as shown in top panel Figure 1

Year	Daily average				
	Q1	Q2	Q3	Q4	YR
2018	6,544	5,439	7,070	12,810	7,980
2019	14,175	15,507	18,367	15,565	15,914
2020	14,780	6,603	10,865	10,274	10,630
2021	10,548	10,463	15,910	17,235	13,564

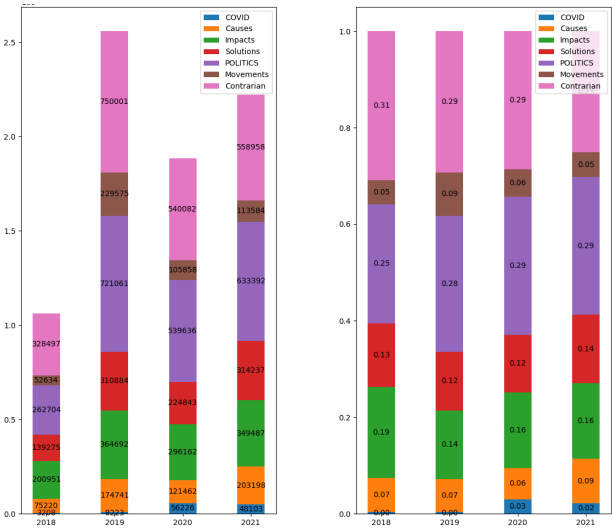
Year	Total count				
	Q1	Q2	Q3	Q4	YR
2018	589,000	494,915	650,401	1,178,497	2,912,813
2019	1,275,775	1,411,105	1,689,744	1,431,972	5,808,596
2020	1,344,968	600,829	999,548	945,200	3,890,545
2021	949,337	952,106	1,463,710	1,585,646	4,950,799

Supplementary Table S2: Statistics for themes showing the number of topic clusters (and the proportion of topics assigned to that theme), the number of tweets in that theme (and the respective overall proportion), the average number of tweets per topic (and standard deviation), the month in the observed time-frame with the most tweets, the number of spot-topics (and proportion of all spot-topics), and the ratio of spot-topics to topics and ratio of tweets in spot-topics to tweets per theme.

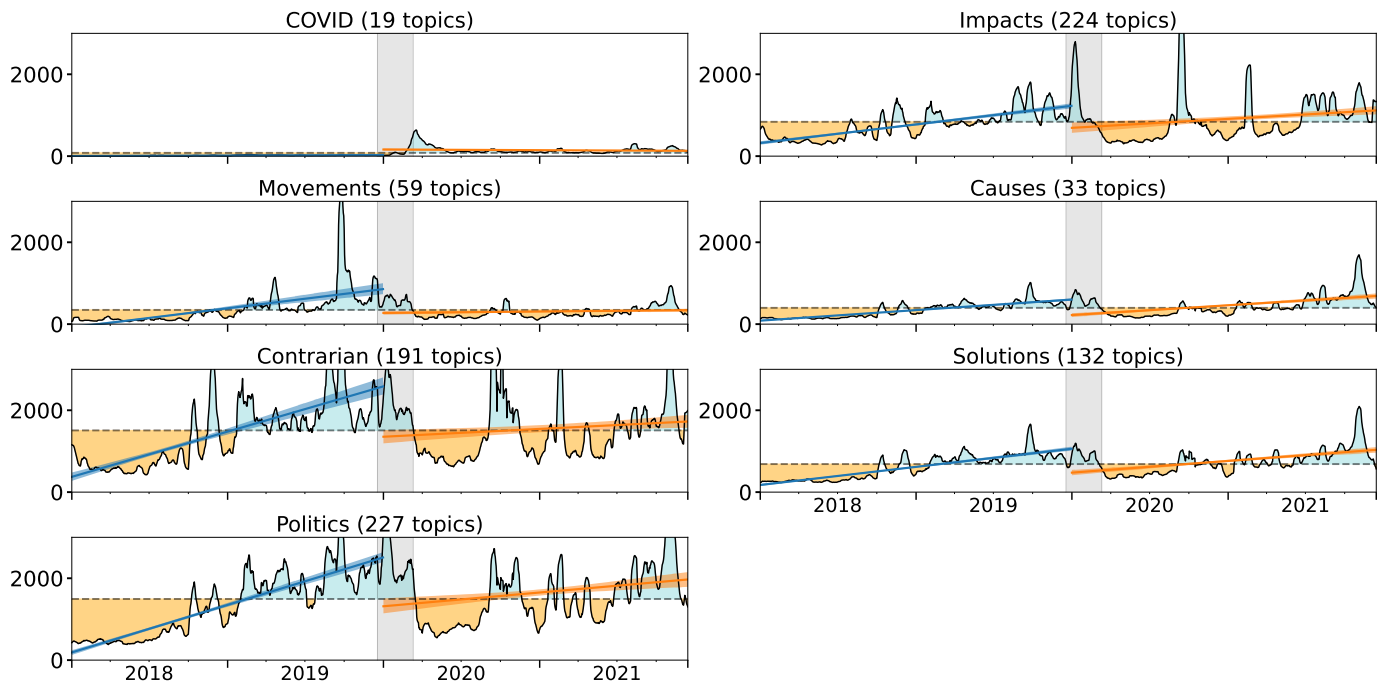
Theme	Topics	Tweets	Tweets/Topic	Peak month	Spot-topics	Ratios
NotRelevant	113 (11.5%)	1,025,877 (7.6%)	189 (1142.5)	2019-09	5 (8.2%)	4.42% / 0.4%
COVID-19	19 (1.9%)	115,760 (0.9%)	127 (531.2)	2020-03	1 (1.6%)	5.26% / 0.5%
Politics	227 (23.1%)	2,156,793 (16.0%)	198 (1211.4)	2020-01	25 (41.0%)	11.01% / 2.6%
Movements	59 (6.0%)	501,651 (3.7%)	177.1 (1106.9)	2019-09	6 (9.8%)	10.17% / 2.5%
Impacts	224 (22.8%)	1,211,292 (9.0%)	112.7 (742.0)	2020-09	14 (23.0%)	6.25% / 4.9%
Causes	33 (3.4%)	574,621 (4.3%)	362.8 (1405.0)	2021-11	2 (3.3%)	6.06% / 0.2%
Solutions	132 (13.4%)	989,239 (7.3%)	156.1 (532.3)	2021-11	6 (9.8%)	4.55% / 1.1%
Contrarian	191 (19.4%)	2,177,538 (16.1%)	237.5 (1279.4)	2019-09	7 (11.5%)	3.66% / 2.2%
Other	198 (20.1%)	4,744,306 (35.1%)	499.2 (6129.0)	2021-11	10 (16.4%)	5.05% / 0.3%
TOTAL	983	13,506,243	286.2 (3147.9)	2019-09	61	— / 0.9%



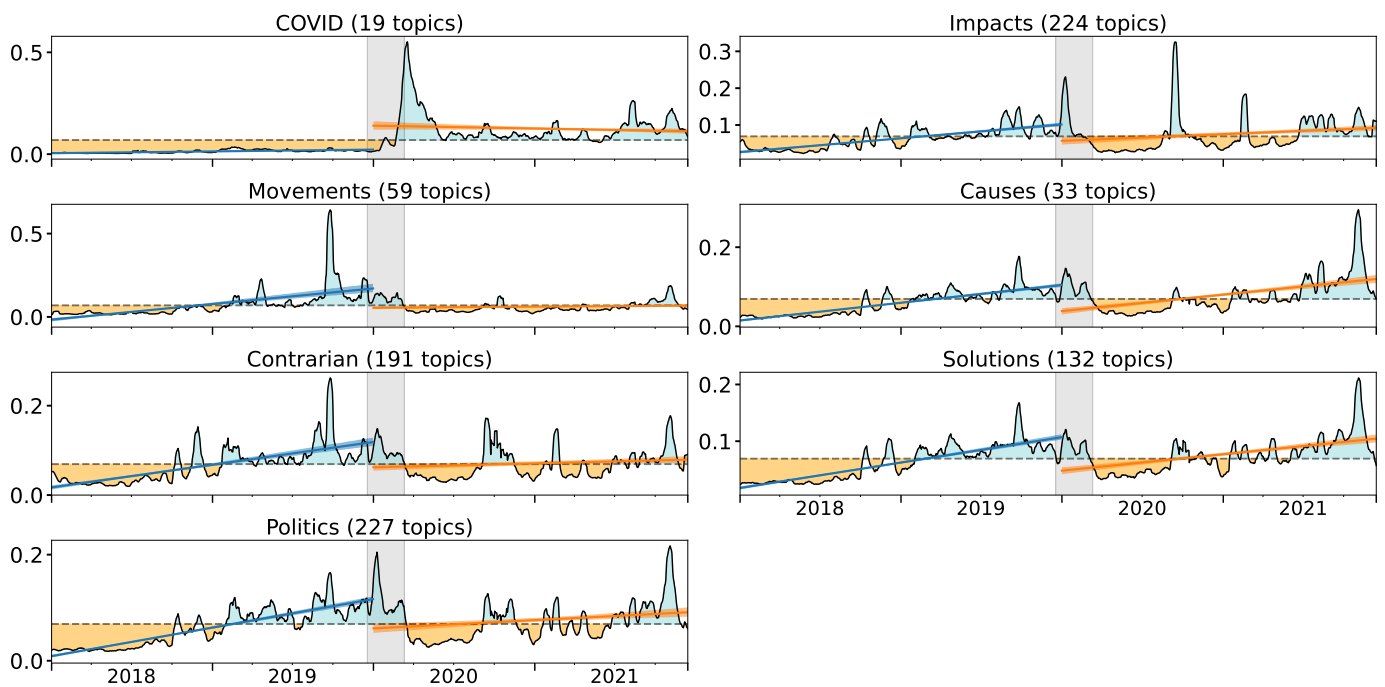
Supplementary Figure S2: Two-dimensional projection of the dataset. Each dot corresponds to a tweet that is placed near its semantically similar neighbours. Colours correspond to the theme assignments. Isolines based on the kernel densities of the distribution of dots per theme.



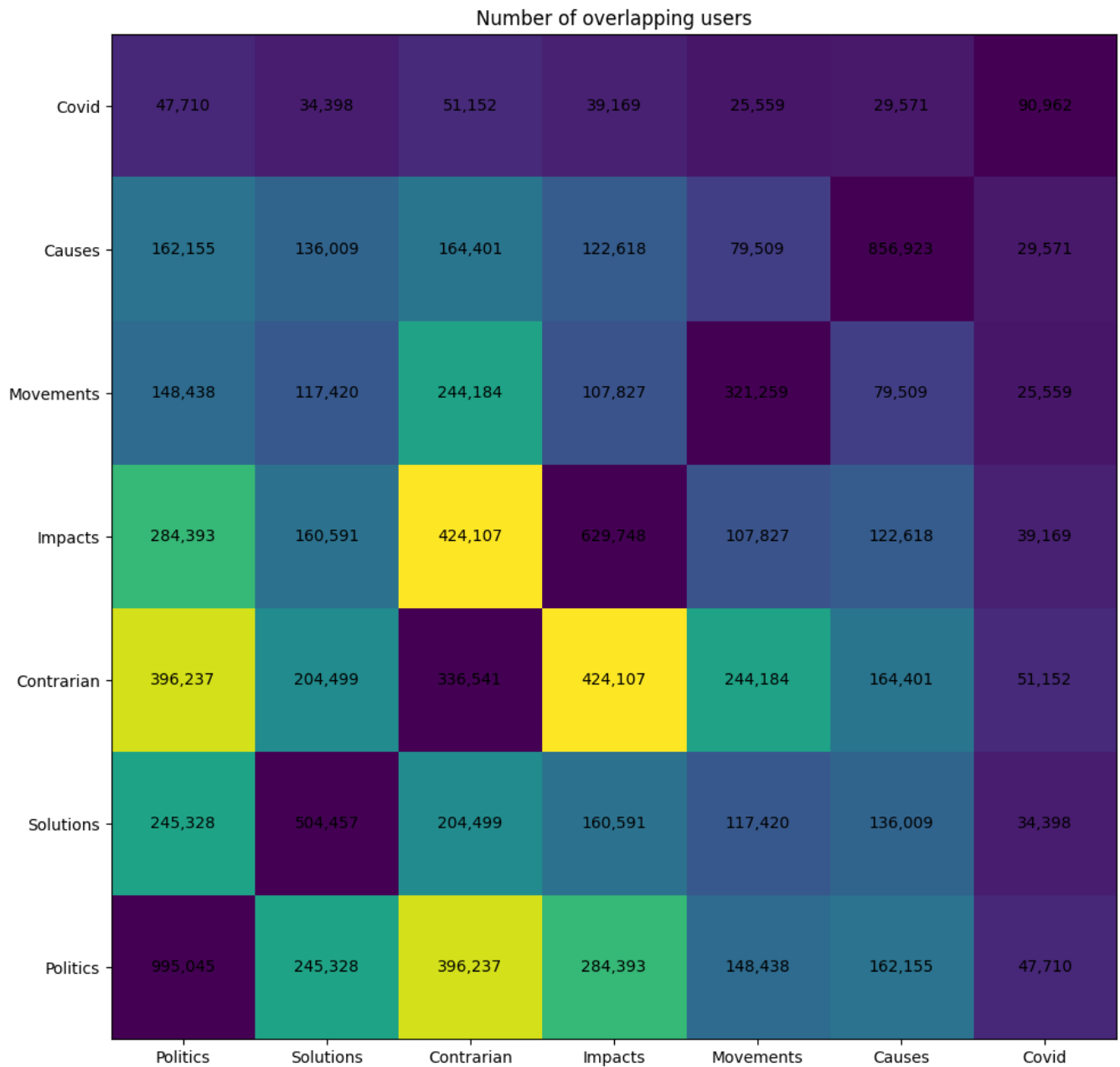
Supplementary Figure S3: Absolute number of tweets per theme per year and distribution of themes per year.



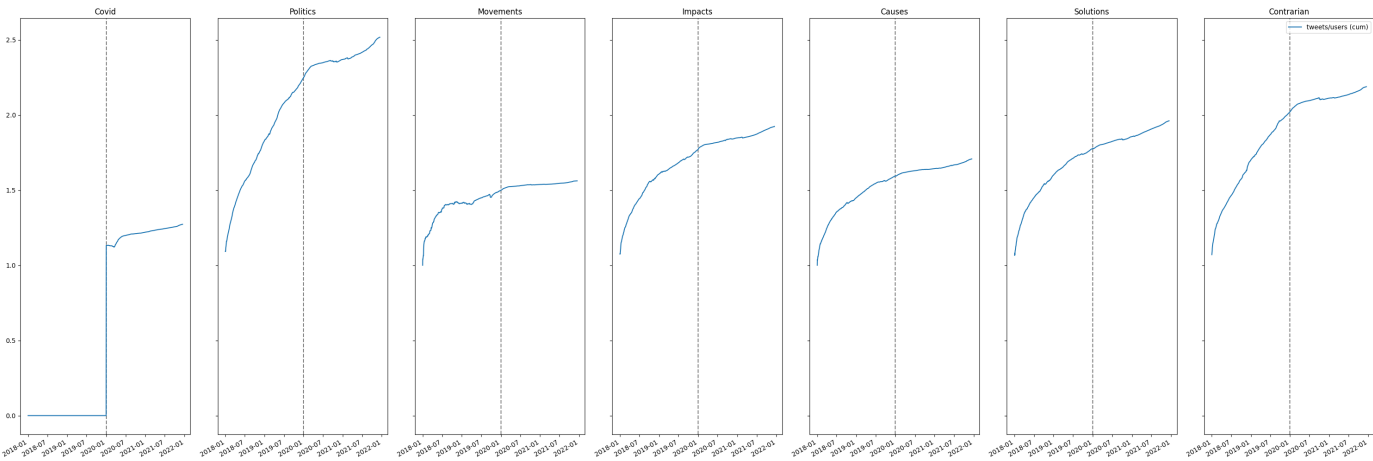
Supplementary Figure S4: Analogous to Figure 2 with absolute numbers and shared y-axis.



Supplementary Figure S5: Analogous to Figure 2 showing the share of tweets per day per theme (in %) in relation to the total number of tweets per theme.



Supplementary Figure S6: Number of users posting in more than one themes.



Supplementary Figure S7: Average number of tweets per user over time.