Supportign Information for
Content Analysis of Multi-Annual Time Series of Flood-Related Twitter (X) Data

Nadja Veigel[1,2,3], Heidi Kreibich[2], Jens de Bruijn[4,5], Jeroen C.J.H. Aerts[5, 6], and Andrea Cominola[1,3]

[1]Chair of Smart Water Networks, Technische Universität Berlin, Straße des 17. Juni 135, Berlin, 10623, Germany
[2]Section 4.4 Hydrology, GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam, 14473, Germany
[3]Einstein Center Digital Future, Robert-Koch-Forum Wilhelmstraße 67, Berlin, 10117, Germany
[4]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[5]Institute for Environmental Studies, VU University, Amsterdam, the Netherlands
[5]Deltares Institute, Delft, the Netherlands

Correspondence: Nadja Veigel (nadja.veigel@tu-berlin.de)

Contents

1 Supporting Information

1.1 Software

- Research Software in this study was developed using Python (version 3.8) and R (version 4.2.3) language.

- Experimental Setup:

  RAM: 256 GB

  CPU: 2 x 16-Core Epyc 7313

- Bidirectional Encoder Representations from Transformers are developed based on Reimers and Gurevych (2019). We develop UMAP dimensionality reduction based on McInnes et al. (2017). To implement Unsupervised clustering HDB-SCAN we rely on McInnes et al. (2018). The BERTopic library Grootendorst (2022) facilitated the BERTopic integration, accessible at https://maartengr.github.io/BERTopic. The scripts used for the analysis are available at: https://github.com/SWN-group-at-TU-Berlin/SocialMediaNLP_FloodTopics.

- Tweet location identification is based on an algorithm developed by de Bruijn et al. (2017). The dataset of flood related Twitter posts from de Bruijn et al. (2019). The Data is available at https://www.globalfloodmonitor.org/ upon request.

1.2 Supporting Information on Method Development and Data

1.2.1 Data cleaning

The tweets containing the following strings were removed prior to modelling, because they indicated that the tweets are not flood related: $*flüchtl*-refugee, *asyl*-, asyl, *migran*-migration, *ezb-geldflut*-ezbmoneyflood, *fachkräfte-flut*-floodofskilledworkers, *app-flut*-appflood, *pegida*-pegida, *islam*-islam, *nazi*-nazi, *#nowplaying*-#nowplaying, *#sex*-sex, *#arsch*-ass, *#obamacare*-#obamacare$

1.2.2 Hyperparameter Tuning

The HDBSCAN clustering results varies with different hyperparameter settings. We iteratively changed the hyperparameters, since our unsupervised approach does not allow for automatic hyperparameter optimization. Limiting the number of clusters to a small number causes many different tweets to be aggregated in one big cluster, even though their topics are very diverse. This will result in fuzzy clusters. Too many topics however, will result in topics that revolve around very specific themes that have little relevance in a broader context. Here we allow for very cluster sizes of minimum 20 tweets to capture a wide range of themes. Additional hyperparameter settings are listed below:

- min_samples = None

- cluster_selection_epsilon = 0.0

- max_cluster_size = 20

- metric = 'euclidean'

- alpha = 1.0

- p = None

- algorithm = 'best'

- leaf_size = 40

- memory = Memory(location = None)

45
- approx_min_span_tree = True

- gen_min_span_tree = False

- core_dist_n_jobs = 4

- cluster_selection_method = 'eom'

- allow_single_cluster = False

50
- prediction_data = False

- match_reference_implementation = False

1.2.3 Twitter (X) Users

Among the organizational users, @*hochwasserportal_de*, an account associated with a portal providing information or updates related to floods or high-water events, has the highest frequency of posting with 2344 tweets throughout the entire time series.

55 There were four other regionally focused analogous accounts that posted moderately between 300-400 tweets. Although not being substantial considering the total number of tweets, this shows a discrepancy in posting activity among these entities and private individuals during the observed period. The tweets from these organizations were semantically very similar and therefore they were successfully separated from the tweets of private users with BERTopic.

2 Figures
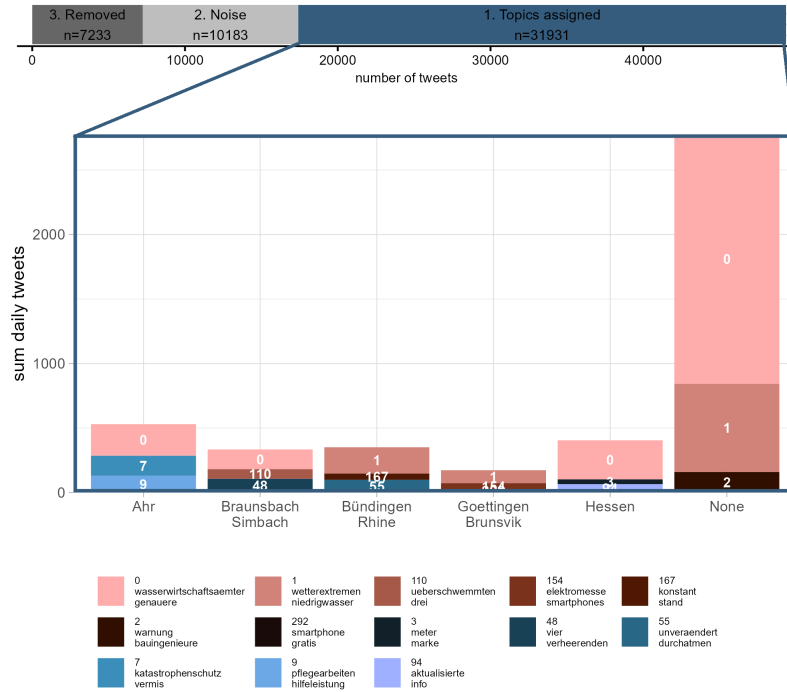
Figure S1. Stacked bar chart showing the three topics representing the highest daily sum of tweets during the event periods. The top chart represents tweets removed by the manual ("Removed") and automatic ("Noise") filtering.

Figure S2. Occurrence of tweets related to the noise topic (top panel, excluded from further analysis) and development of monthly tweets for the four topics with the most occurring instances over the whole time period.
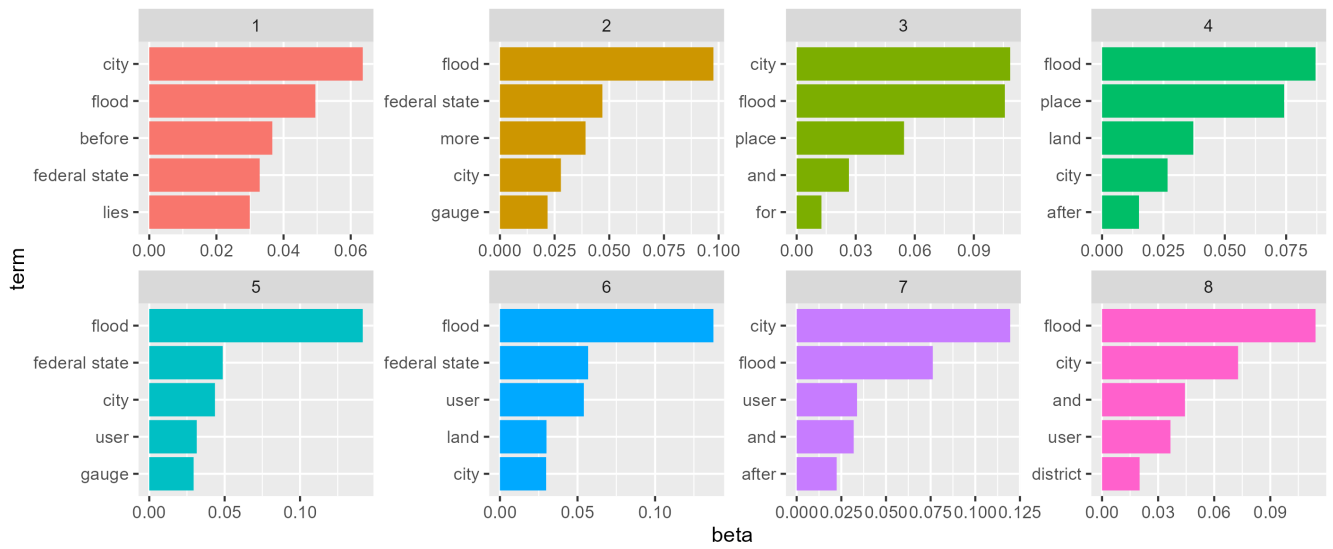
Figure S3. Top terms for 8 LDA topics and beta values.

Table S1: Topic descriptions for all topics represented in Figure 1 in the main manuscript

Topic	Group	Top_n_words	freq	representative text
0	Signal and detect disasters	wasserwirtschaftsaemter - genauere - unteren - mittlerweile - bayerische - informationen - bundesland - sinken - schlimmer - bilanz	2653	pegel bundesland aktuelle hochwasser info liegt vor mehr unter
1	Signal and detect disasters	wetterextremen - niedrigwasser - besser - betroffen - thema - neue - infos - starkregen - heute - hochwasser	1010	pegel bundesland aktuelle hochwasser info liegt vor mehr unter
3	Provide and receive disaster warnings	meter - marke - normalerweise - zurueckgeht - stunde - zentimeter - tauwetter - geregnet - acht - regnet	286	der hochwasser scheitelpunkt scheint erreicht der koelner pegel stagniert nach meter in stadt wurden meter erreicht noch heute sollen die pegel sinken
7	Discuss socio-political and scientific causes and implications of and responsibility for events	katastrophenschutz - vermis - reserve - blackouts - verbesserungsbedarf - provinzial - bahnfamilie - krisenvorsorge - tankstel - wetterkatastrophe	252	katastrophenschutz vor monaten probte land den ernstfall und scheiterte hochwasser
24	Discuss socio-political and scientific causes and implications of and responsibility for events	bundeskanzlerin - angela - unbuerokratische - tausend - merkelmussweg - apokalypseundfilterkaffee - verzweifelten - invasionen - schweigt - headline	147	neuer beitrag nach hochwasser merkel in stadt tagesschau
48	Inform others about ones own condition and location and learn about a disaster-affected individuals condition and location	vier - verheerenden - zerstoerungen - staudamm - erschossen - gekomme - regionalnachrichten - starben - android - telmi	110	vier tote bei hochwasser katastrophe in bundesland
55	Provide and receive disaster preparedness information	unveraendert - durchatmen - hoellentalklamm - nosferatu - totkriegen - leuchtet - bangen - vermisste - stadt - hochwasser	102	der rhein pegel in stadt steht um unveraendert bei zentimetern hochwasser
69	Signal and detect disasters	konstant - stand - gt - cm - rhein - min - pegel - benutzer - stadt - hochwasser	92	hochwasser stadt benutzer rhein gt pegel cm seit h konstant stand

110	Document and learn what is happening in the disaster Event	ueberschwemmten - drei - sommergewittern - break - spiegelonline - nationalekatastrophe - gerrithiemstra - panorama - gefunden - verheerenden	76	drei tote bei hochwasser in bezirk
114	Inform others about ones own condition and location and learn about a disaster-affected individuals condition and location	jahrhundert - kraftwerke - informationen - steigt - todeszahlen - thailandtip - abends - rwe - ministerpraesidentin - verletzt	75	hochwasser zahl der toten steigt auf opfer im raum stadt
122	Raise and develop awareness of an event; donate and receive donations; identify and list ways to help or volunteer	betroffene - finanzminister - euro - bankengruppe - spendenstream - ausbezahlt - deutschenstaat - psd - fraktionschef - finanzhilfe	73	soeder kuendigt euro soforthilfe fuer hochwasser betroffene an bundesland zahlt den betroffenen der
152	Document and learn what is happening in the disaster Event	dammwachen - villafohrde - desserts - entspannte - twitterlaufgruppe - ashura - fieroovend - sonnentag - shoobamusic - wegenmorgen	64	hochwasser in land die nacht verlief ruhig hochwasser in bundesland entspannt sich etwas
154	Provide and receive disaster warnings	elektromesse - smartphones - wenigen - verteidigungsministerium - katastrophenalarm - vorsorgestrategie - evakuierungsmaÇ?nahmen - kreiszeitung - breaking - warten	64	hochwasser landkreis ruft katastrophenalarm aus
167	Signal and detect disasters	konstant - stand - gt - cm - min - rhein - pegel - benutzer - stadt - hochwasser	61	hochwasser stadt benutzer rhein gt pegel cm seit h konstant stand
178	Document and learn what is happening in the disaster Event	langsam - fallenden - glueckgehabt - jahresauftakt - krabbenbroetchen - orteinsatzkraefte - wellblech - schiffmann - sandstrand - flussgebiet	58	hochwasser in bundesland pegelstaende sinken nur langsam

Table S2: Topic descriptions for all topics represented in Figure 2 in the main manuscript

Topic	group	Top_n_words	freq	Document
168	Document and learn what is happening in the disaster Event	niederbayerischen - todesopfern - katastrophalem - onlinewetter - dpa - gmx - verheerenden - perso - gekostet - atempause	61	schwere unwetter mindestens vier tote bei hochwasser in bezirk
167	Signal and detect disasters	konstant - stand - gt - cm - min - rhein - pegel - benutzer - stadt - hochwasser	61	hochwasser stadt benutzer rhein gt pegel cm seit h konstant stand
110	Document and learn what is happening in the disaster Event	ueberschwemmten - drei - sommergewittern - break - spiegelonline - nationalekatastrophe - gerrithiemstra - panorama - gefunden - verheerenden	76	drei tote bei hochwasser in bezirk
94	Signal and detect disasters	aktualisierte - info - liegt - mehr - pegel - bundesland - hochwasser - - -	79	pegel bundesland aktualisierte hochwasser info liegt vor mehr unter
76	Discuss socio-political and scientific causes and implications of and responsibility for events	ortufern - schneeschmelze - ungerecht - festgefroren - crashs - gluehwein - dauerfrost - skifahrer - sibirischen - dachgeschosswohnung	86	nach viel schnee zuvor bescherte hochwasser den freiwilligen feuerwehren stadt und stadt am letzten w
55	Provide and receive disaster preparedness information	unveraendert - durchatmen - hoellentalklamm - nosferatu - totkriegen - leuchtet - bangen - vermisste - stadt - hochwasser	102	der rhein pegel in stadt steht um unveraendert bei zentimetern hochwasser
48	Inform others about ones own condition and location and learn about a disaster-affected individuals condition and location	vier - verheerenden - zerstoerungen - staudamm - erschossen - gekomme - regionalnachrichten - starben - android - telmi	110	vier tote bei hochwasser katastrophe in bundesland
12	Deliver and consume news coverage of the disaster Event	nachrichtenagentur - tennessee - hashtag - satellitenbilder - parteivorsitzende - museumsperspektive - altenburger - floodmaps - bildern - newsdestages	186	hat eine headline geaendert hochwasser in land steinmeier besucht
9	Send and receive requests for help or assistance	pflegearbeiten - hilfeleistung - pizzakarton - obdachlosen - kriesengebiet - obdachloser - khuddam - abschlussmeldung - hilfeleistungskontingent - hilfsangebote	216	betrifft kreis stadt gezielte hilfsangebote werden unter benutzer stadt de gesammelt und koo

7	Discuss socio-political and scientific causes and implications of and responsibility for events	katastrophenschutz - vermis - reserve - blackouts - verbesserungsbedarf - provinzial - bahnfamilie - krisenvorsorge - tankstel - wetterkatastrophe	252	katastrophenschutz vor monaten probte land den ernstfall und scheiterte hochwasser
0	Signal and detect disasters	wasserwirtschaftsaemter - genauere - unteren - mittlerweile - bayerische - informationen - bundesland - sinken - schlimmer - bilanz	2653	pegel bundesland aktuelle hochwasser info liegt vor mehr unter

References

de Bruijn, J. A., de Moel, H., Jongman, B., Wagemaker, J., and Aerts, J. C. J. H.: TAGGS: Grouping Tweets to Improve Global Geoparsing for Disaster Response, Journal of Geovisualization and Spatial Analysis, 2, https://doi.org/10.1007/s41651-017-0010-6, 2017.

de Bruijn, J. A., de Moel, H., Jongman, B., de Ruiter, M. C., Wagemaker, J., and Aerts, J. C.: A global database of historic and real-time flood events based on social media, Scientific data, 6, 311, 2019.

Grootendorst, M.: BERTopic: Neural topic modeling with a class-based TF-IDF procedure, arXiv preprint arXiv:2203.05794, 2022.

McInnes, L., Healy, J., and Astels, S.: hdbscan: Hierarchical density based clustering., J. Open Source Softw., 2, 205, 2017.

McInnes, L., Healy, J., and Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction, arXiv preprint arXiv:1802.03426, 2018.

Reimers, N. and Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, http://arxiv.org/abs/1908.10084, 2019.