



A Transformer-Based Analysis of Tweets in Germany to Investigate the Appearance and Evolution of the 2021 Eifel Flood in Social Media

Nadja Veigel^{1,2,3}, Heidi Kreibich², Jens A. de Bruijn⁴, Jeroen C.J.H. Aerts⁴, and Andrea Cominola^{1,3}

Abstract

Background

- Risk communication and flood-related citizens' behaviors, attitudes, and perceptions before, during, and after the flood are currently understudied.
- Social media data can be used to understand human behavior, as warnings, intensity and impact are reflected in social media topics [1].

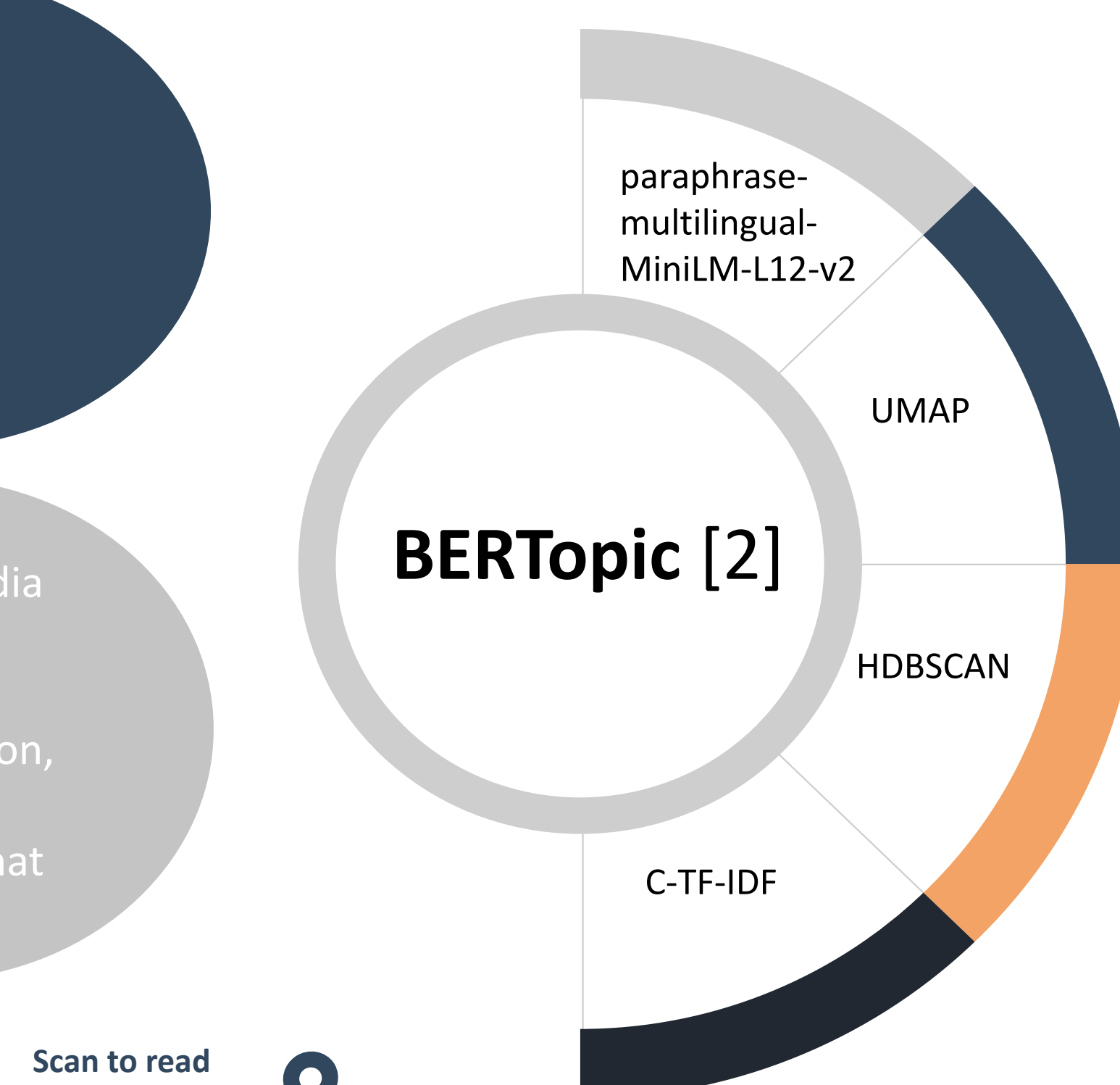
Aim

- We analyze differences between topics arising on social media for events with a high impact compared to less devastating floods that occurred in the past based on 42,000 tweets.
- Our approach can contribute to modelling risk communication, behavioral drivers, and social interactions in relation to different types of floods identifying indirect flood impacts that are not reported in traditional flood documentation

Results and Challenges

- Social media contains a lot of irrelevant information. This requires a lot of filtering and data preprocessing.
- We can see difference between the topics representing flash floods and riverine flooding.

Methods



Scan to read the full abstract



Results – Time Series

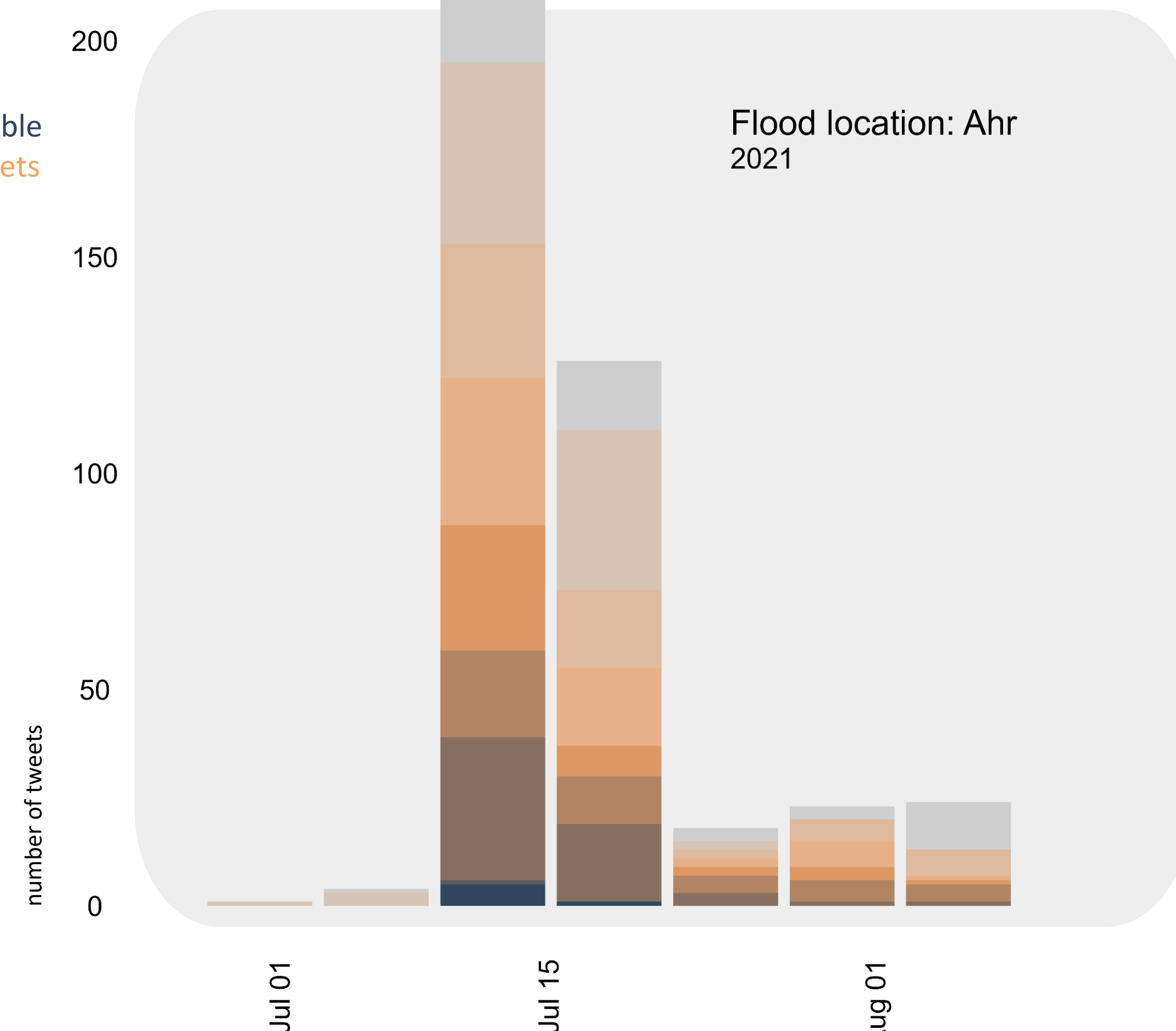


Fig. 3 Occurrence of the 10 most common topics (Fig. 2) over time for 3 specific events marked in Fig. 1. Stacked bar charts show the number of tweets assigned with a topic. Tweets were aggregated per week. The plots represent intermediate results and some topics such as "south germany, Berlin, Frankfurt" may also be considered as noise.

Data

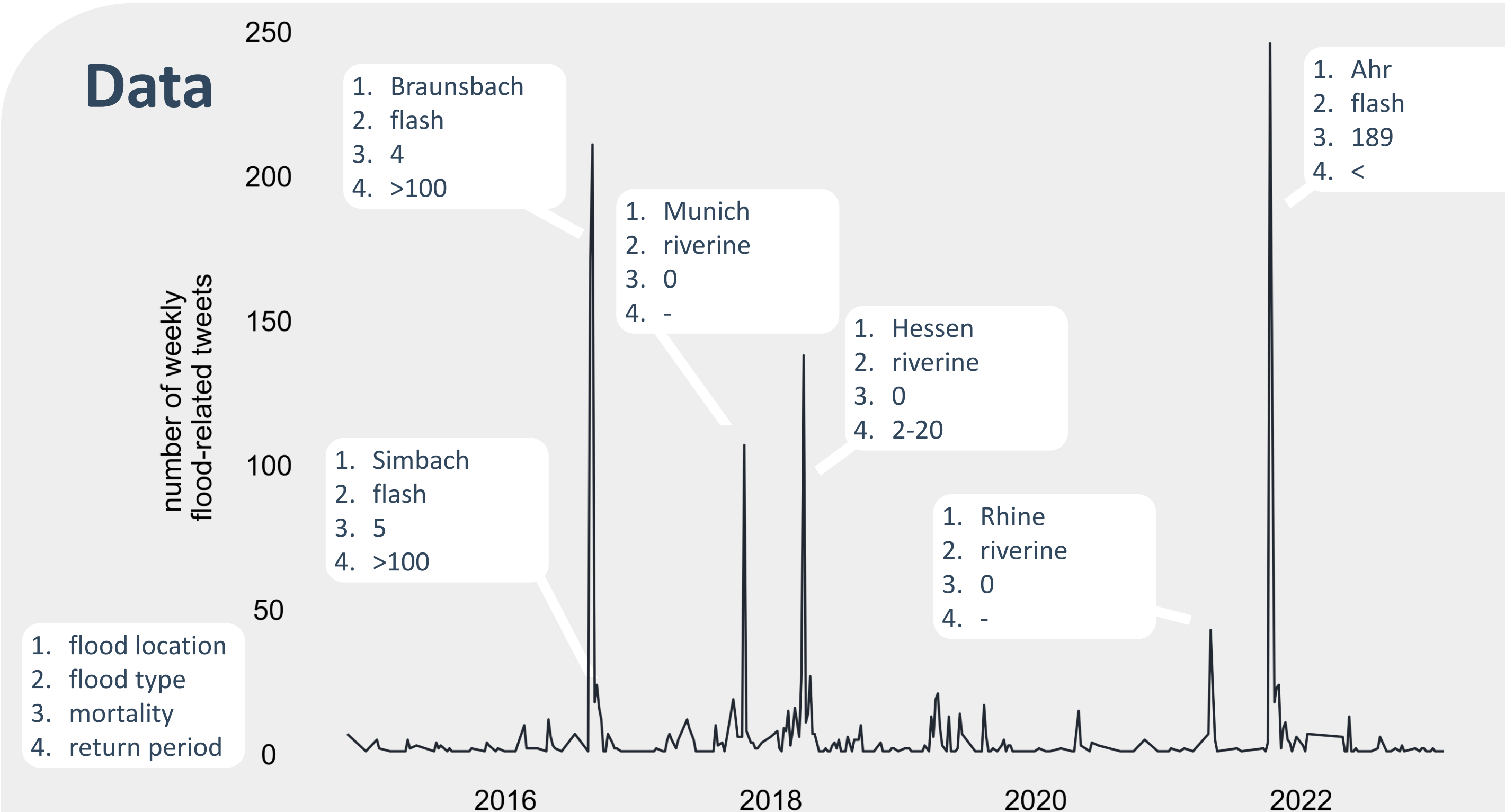


Fig. 1 We aim for homogeneous automatic assessment of Twitter data over time and analyse the content of 42,186 tweets containing selected keywords related to flooding [6] posted in German since 2014

Conclusion & Challenges



Social media data contains a lot of topics that can be considered as noise from the point of view of flood risk assessment (Fig. 2). This makes automatization of topic detection difficult.



Topics of tweets vary between events of different type. During flash floods evacuations and fatalities/missing people are reported, while there are more warnings circulated before riverine floods (Fig. 3).



To refine this analysis we will manually merge and extract topics of interest for flood related human behavior, indirect impacts (e.g. on the traffic) and emergency management.

References

- Reuter, C. & Schröter, J. (2015). Microblogging during the European Floods 2013: What Twitter May Contribute in German Emergencies. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 7(1), 22-41. <http://doi.org/10.4018/IJISCRAM.2015010102>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.
- McInnes, L., Healy, J., UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, *ArXiv e-prints* 1802.03426, 2018
- L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering In: *Journal of Open Source Software*, The Open Journal, volume 2, number 11. 2017
- de Bruijn, J.A., de Moel, H., Jongman, B. et al. A global database of historic and real-time flood events based on social media. *Sci Data* 6, 311 (2019). <https://doi.org/10.1038/s41597-019-0326-9>

Infographics and icons adapted from www.slidesgo.com

Contact

Nadja Veigel

Technische Universität Berlin
Chair of Smart Water Networks
Straße des 17. Juni 135
10623 Berlin

email: nadja.veigel@tu-berlin.de
website: www.tu.berlin/swn,
www.heibrids.berlin