9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Two years of Trending Topics: Measuring issue alignment in the German Twittersphere

Twitter, large-scale social phenomena, issue alignment, network analysis, topic modeling

Extended Abstract

Twitter has been extensively studied by computational social scientists, but analyses that span up several discussion topics on the platform have remained rare. Moreover, with few exceptions, see e.g. Chen et al. (2021), studies of nation-wide Twitter activity have focused on simple following relationships of users (Bruns et al. 2017). The present work fills this research gap by providing a high-level view on public debate on the platform. It investigates the interaction structure as well as the content of trending topics (which we call *trends* in the following) of the German Twittersphere over two years (March 2021 until March 2023). Analyses of trending topics have been restricted to the topic keywords and simple tweet metrics in the past. We collected a volume of 15,225,114 tweets mentioning each day's most salient trending topic(s) on Twitter over the course of two years.

This dataset allows us to analyse tweets and users mentioning each topic and construct retweet networks for each trend. First, we embed the trends in a 2D space based on their user overlap (Fig. 1A). There, we can clearly distinguish clusters of thematically similar topics. For instance, we can observe through the position in the embedding that the discussion around antisemitism attracts a different user base than discussions around sports. Political topics (e.g. federal elections or the Ukraine war) are more tightly knit together but still distinguishable according to their positioning in the projected space. In order to group these trends together thematically, we compute a sentence-embedding-based topic model using BERTopic (Grootendorst 2022). This allows us to categorize the data into larger themes such as COVID, Sports, Ukraine, Floods, Police Violence, Immigration, Holidays, Antisemitism.

It is striking that the majority of trends are ordered along a time axis, while trends related to sports are not (see Fig. 1B). This hints towards a fundamental difference between topics that have volatile user engagement and more stable topics such as sports or gaming. In Fig. 1C, we observe the average user overlap within and between topics, which shows that sports has a low user overlap with all other, more political, topics. Looking at the intra-topic user overlap, we see the highest values for political topics such as police violence, immigration and antisemitism.

With the data set at hand, we are also able to test whether there exists an alignment of user positions across several issues – may they be explicitly political or not. It is an open question whether sorting tendencies across issues exist in Germany, and if so, at what degree. Such tendencies can be problematic for the public debate since they might enhance in-group identities and out-group aversion, and might impede solution-oriented discourse. We approach this question by partitioning the retweet network of each trend into clusters of densely connected users. More precisely, we infer a stochastic block model constrained to max. two blocks, which yields a partitioning of the nodes into one cluster for networks that are unlikely to be generated by a two-block-SBM, and two clusters otherwise.

Next, we examine issue alignment by quantifying the similarity of clusterings across topics. We compute the mutual information between all clusterings of retweet networks associated to each trend and average these values over topics in order to measure an alignment score between

each pair of topics. For example, police violence and immigration debates show comparatively high alignment (Fig. 1D). On the other hand, discussions around natural disasters such as floods or heatwaves, and sports exhibit an overall weaker alignment to trends within the same topic, as well as to trends from other topics.

In an additional step, we shall investigate further the role of different types of users: we observe on one hand that there exist highly active and influential users that are retweeted more than others (*opinion leaders*), and on the other hand, highly active users that retweet a lot more than others (*opinion disseminators*). What role do these different types of users play in shaping the phenomenon of issue alignment on social media? Is issue alignment more prevalent within these groups? Exploring these questions on this data set will improve our understanding of political polarization and ideological alignment on social media.

References

Bruns, Axel et al. (2017). "The Australian Twittersphere in 2016: Mapping the follower/followee network". In: *Social Media+ Society* 3.4, p. 2056305117748162.

Chen, Ted Hsuan Yun et al. (2021). "Polarization of climate politics results from partisan sorting: Evidence from Finnish Twittersphere". In: *Global Environmental Change* 71, p. 102348.

Grootendorst, Maarten (2022). "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". In: *arXiv preprint arXiv:2203.05794*.

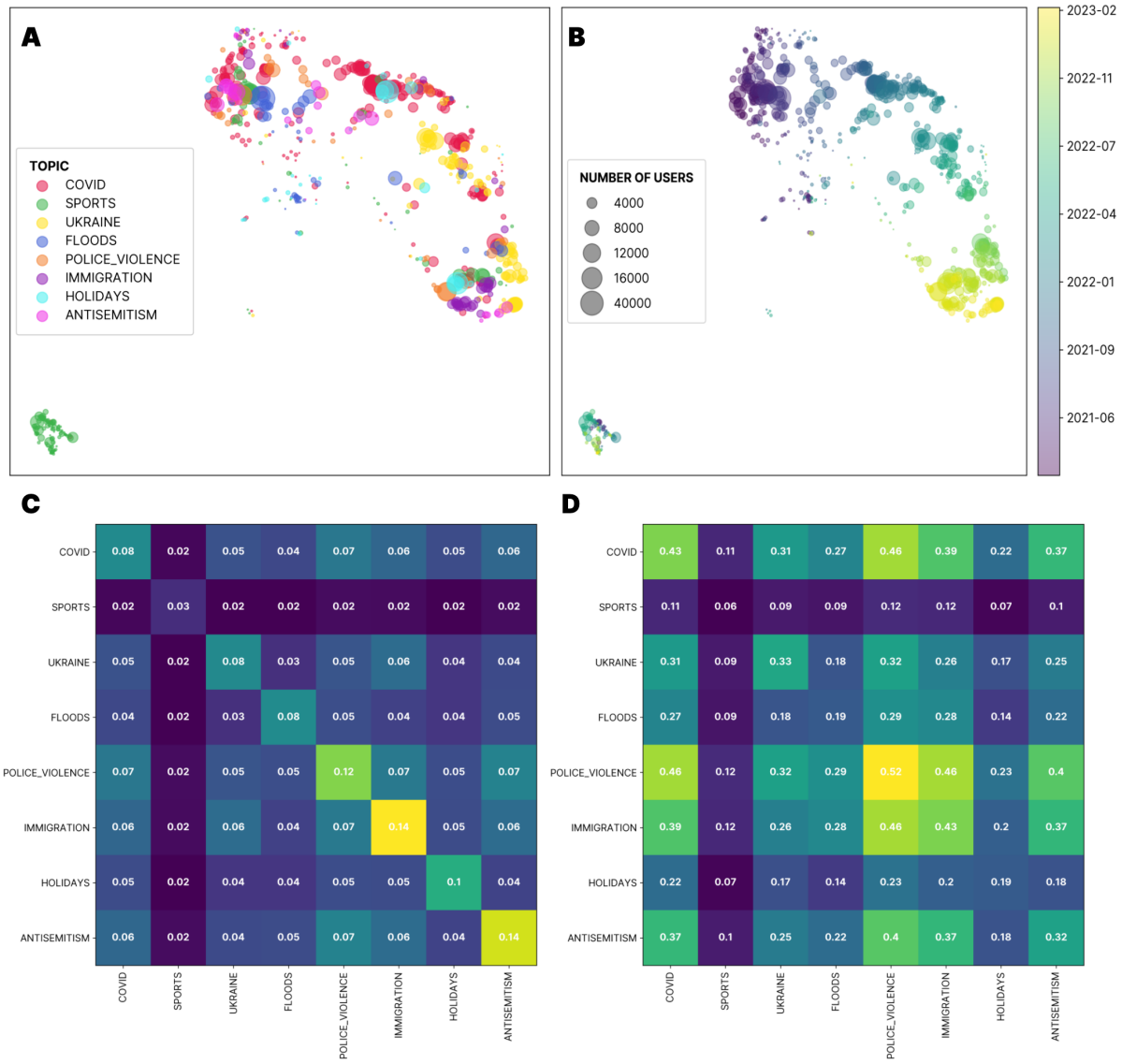

Figure 1: **A** shows a UMAP embedding of the Jaccard distance matrix between trends with respect to users. Each trend is colored according to the topic. We see that the cluster of trends discussing sports (green) is separated from the other topics. **B** shows the same trend embedding colored according to the time they were collected. We see that user overlap correlates with temporal proximity. **C** shows the pairwise Jaccard index averaged over topics. **D** shows the mutual information averaged over topics.