sociological science

Supplement to:

Keuchenius, Anna, Petter Törnberg, and Justus Uitermark. 2026. “Echo Chambers Are Defined by Conflict, Not Isolation” Sociological Science 13: 565-588.

Supplementary Material to **Echo chambers are defined by conflict, not isolation**

Section A: Twitter word embedding

To infer the sentiment of a mention (endorsement, opposition or neutral/ambiguous) we used the fastText algorithm with pretrained word vectors constructed from a large Twitter dataset of Dutch tweets in 2018. We use a word embedding based on a Twitter corpus instead of a more general corpus, such as a Wikipedia corpus, because this embedding can encode particularities of the ways users discursively express themselves on Twitter, in particular towards others. Users express sentiments towards other users on Twitter with specific words and expressions that are expected to linguistically differ from the word use and sentences found in more general corpora of Dutch language, such as Wikipedia, news articles or books.

The dataset of Dutch tweets was compiled by the Netherlands eScience Center in the scope of the TwiNL project which has as its ambition to collect all tweets on Twitter from the Netherlands (Sang & van den Bosch, 2013). Selecting the tweets from this dataset published in 2018 (n=179,789,348) and tokenizing with the tokenize package of the python nltk python library gave roughly 6 million unique tokens. Vectors with dimension 300 were created with the use of the skipgram algorithm (Bojanowski et al, 2016).

References

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. “Enriching Word Vectors with Subword Information.” *Transactions of the Association for Computational Linguistics* 5:135–46. doi: 10.1162/tacl_a_00051.

Sang, Erik Tjong Kim, and Antal Van den Bosch. "Dealing with big data: The case of witter." *Computational Linguistics in the Netherlands Journal* 3 (2013): 121-134