

Discovering Pandemic Topics on Twitter

Erik Tjong Kim Sang¹, Marijn Schraagen²,
Mehdi Dastani², Shihan Wang²

¹Netherlands eScience Center, Amsterdam

²Intelligent Software Systems, Utrecht University



Utrecht University

Introduction

We search for topics in the pandemic discussion on Dutch Twitter. We use three methods for extracting topics from tweets: **hashtag counts** of tweets containing one of the strings *covid* or *corona*, a **statistical t-test comparison** of the vocabulary used in these tweets and the vocabulary of the other tweets, and **Latent Dirichlet Allocation**, a method for automatically extracting topics from texts

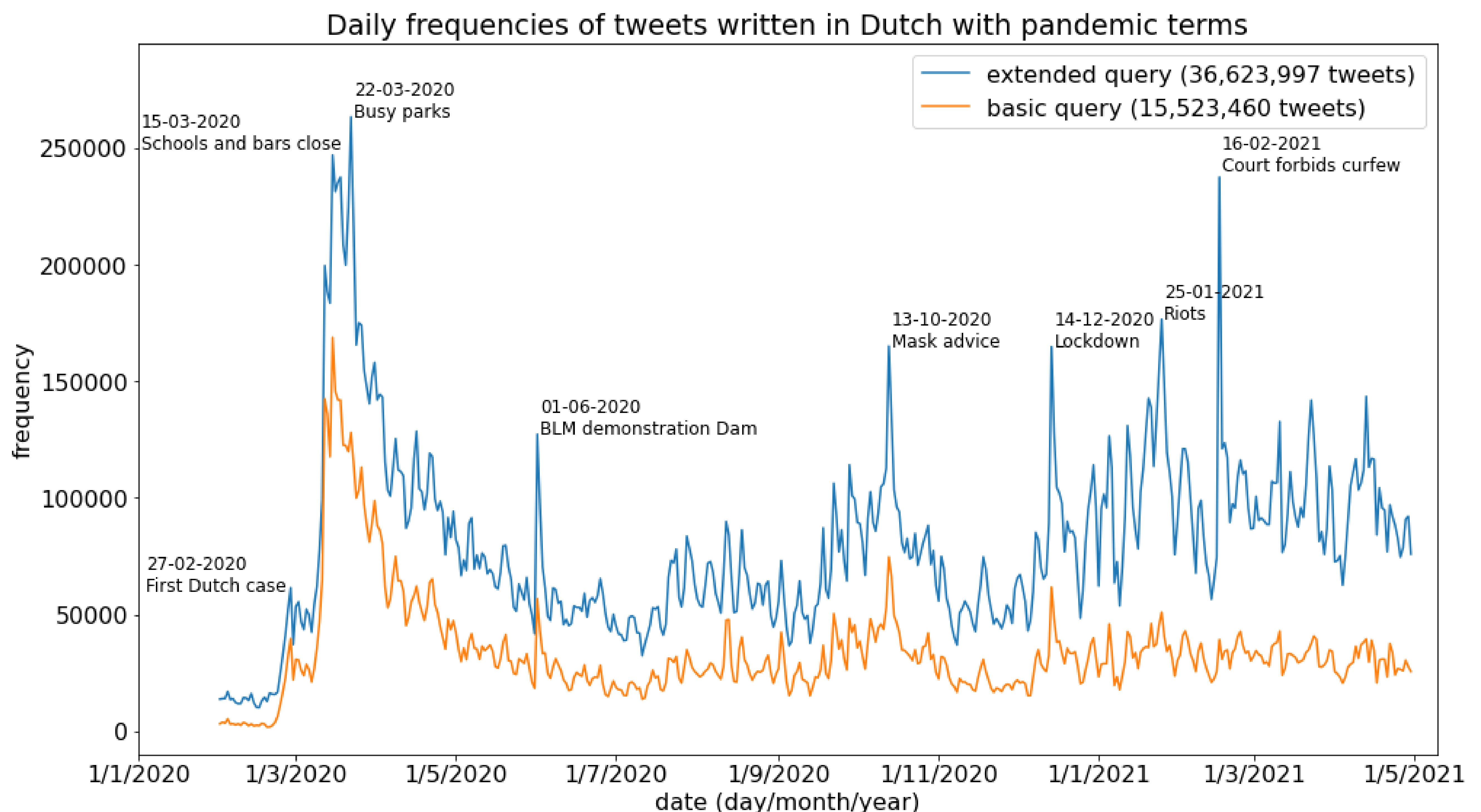
Results

Each of three methods found relevant

pandemic topics in the data. An analysis over time revealed that new topics kept coming up during the course of the pandemic. While the initial keywords *covid* or *corona* were very useful for finding relevant tweets early in the pandemic they used less often in the relevant later discussions.

Application

The detected topic words have been used to extend our search query for pandemic tweets from two to 67 words, resulting in finding 136% more relevant tweets (see graph below).



More information

Paper: <https://ifarm.nl/erikt/papers/dhbenelux-20210509.pdf>

Data: <http://145.100.59.103/downloads/>

