NORDIS – NORdic observatory for digital media and information DISorders

Predicting COVID-19 related collective anxiety on social media

Report on scientific paper with executive summary

Date: 31-08-2022

Final version

NORDIS

Action No:	2020-EU-IA-0189
Project Acronym:	NORDIS
Project title:	NORdic observatory for digital media and information DISorders
Start date of the project:	01/09/2021
Duration of the project:	24
Project website address:	https://datalab.au.dk/nordis
Authors of the deliverable	Marina Charquero-Ballester Jessica Gabriele Walter Astrid Sletten Rybner Ida A. Nissen Kenneth Christian Enevoldsen Anja Bechmann
Activity number	Activity 2
Task	Task 2.2

Reviewers: Minna Aslama Horowitz, Matteo Magnani

Funding:
The research was funded by EU CEF grant number 2394203.

NORDIS

1.0 Executive Summary

The Covid-19 pandemic - with all its many challenges - has allowed for unprecedented studies of public emotions in times of crisis using social media data to take the pulse of the population's emotions. Many studies have highlighted the presence of fear and anxiety during the most uncertain times of the pandemic, especially in those countries most badly affected at the time (e.g., Italy, UK, China). But what happens when we turn our heads towards more resilient countries, such as some of the Nordic countries?

Do uncertainty and health concerns also translate into fear?
This is the first question we asked in our study, in which we bring the focus to four of the Nordic countries (i.e., Denmark, Norway, Finland and Sweden) that share a high level of trust in their democracies, the media, comprehensive healthcare, high levels of education and similar welfare state values. We implemented a multilingual transformer model to detect emotions in a large sample of the Nordic Twitter, containing over 57 million tweets in Danish, Norwegian, Finish and Swedish, without limiting those tweets to be Covid19-related. The data was collected during the second wave of the pandemic, including a period before hospitalisation numbers started to rise up and until numbers had normalised again after reaching the peak in all countries (i.e., from August 2020 to March 2021). Strikingly, our model did not detect any tweets containing fear for this time period in any of the languages present in our data. Of note, the model had been able to correctly detect most of the fear-related tweets in the validation data we used (SemEval 2018 dataset, subtask 5), annotated through a crowdsourcing project in which participants inferred the affectual state of a person from their tweet.

Does high trust equal less fear?
A possible explanation for the lack of fear-related tweets are the high levels of trust in the government, healthcare system and the media in the Nordic countries, which make citizens feel like the crisis is under control. In other words, at least to some degree, the high resilience that the population from the Nordic countries expect from their governments, health care system and other institutions, prevent fear from being spread, and even expressed often, on Twitter. However, there are a number of other factors likely to play a role in these findings. First, the results are produced by a model, with the biases and limitations that derive from the data it was trained on. For one, the data used for the finetuning of the model on the task of emotion detection came from English tweets translated using automatic tools into the Nordic languages of this study, with the loss of information inherent to this process. Also, the training dataset was unbalanced in the amount of emotions inferred from those tweets (e.g., many tweets expressed joy and anger, but not so many expressed sadness), leading to further difficulties in detecting those emotions that were underrepresented - a well known limitation ubiquitous to machine learning algorithms. Even with these factors in consideration, the results seem to indicate that the amount of fear is

much lower in the Nordic Twitter than we would have expected based on results from other countries, but that is not to say that people in the Nordic countries experienced no fear during the pandemic. As we consider the Twitter culture and the profile of Twitter users in the Nordic countries, it also seems unlikely for our data to be an accurate representation of the Nordic population on its own. Complementing our dataset with Facebook data might show different results in terms of the amount of fear being expressed online. Unfortunately, Facebook data is not as widely accessible for academic research as Twitter data.

Anger versus joy

The second question we ask in our study is, if fear is not a predominating emotion in the Nordic Twittersphere during Covid-19, what is? And is that different between the Nordic countries taking part in this study? From our results, we infer that the two main emotions during that period of time were joy and anger in any of the Nordic countries. While it might be tempting to draw conclusions regarding the presence of joy in what are believed to be among the happiest countries on Earth, a word of caution comes from previous research showing that high-arousal emotions, such as anger and joy, tend to become viral and therefore be overrepresented on social media in contrast to data sampled through other methods (e.g., surveys). And as mentioned above, our model was more able to detect some emotions than others, with joy and anger being the 'easiest' emotions for the model to recognise. We further assessed differences between the countries of study, finding that joy is always the predominant emotion detected across countries, although the difference between the amount of joy and fear is much larger in Finland than in the other countries. This especially goes for Sweden, where the amounts of joy and anger are closest to a balance. However, this analysis was conducted under the assumption that the model performs equally well across languages, but further research would be needed in order to verify this assumption.

Emotions' evolution over time

Lastly, we focused on the fluctuations of joy across time, to assess potential relationships with variables indicative of the pandemic evolution, such as number of hospitalisations and time spent at home. We found no relationship between fluctuations in joy and neither of these two variables for any of the countries, in contrast to similar studies addressing these questions. Instead, we observed that spikes in joy could be found around national celebrations, such as Christmas and New years eve. Given the comprehensive data collection implemented for this study, which did not involve restricting data to Covid19-related tweets, a hidden relationship between these variables might become visible if we refine our dataset around Covid-19, removing unrelated tweets.

Future work

All in all, this study provides insights about the public emotions in the Nordic countries during the second wave of the pandemic. It does so through an exploratory analysis of - to our knowledge - the largest Nordic Twitter dataset collected during the Covid-19 crisis. Our findings highlight a lack of fear in the Nordic Twittersphere in contrast to what we would have expected based on the results from studies on data from other countries during the pandemic. Instead, we found that most tweets in our data expressed joy or anger, and that joy remained stable independently of the rising number of hospitalisations and increasing number of hours spent at home. This was the case across all Nordic countries studied in this project. Future research could address the generalizability of our findings across social media platforms, and add complementary information through data from other countries during this same time period, in order to see whether additional research supports our preliminary findings.

Preprint paper under submission as part of the report

Joy during the Covid-19 pandemic in the Nordic countries - what is this all about?

Marina Charquero-Ballester, Jessica G. Walter, Astrid Sletten Rybner, Ida A. Nissen, Kenneth Christian Enevoldsen & Anja Bechmann

1.0 Introduction

Signatures of public emotions are present throughout social media platforms, such as Twitter, providing an opportunity to study public emotions at scale. The Covid-19 pandemic provides an interesting **probe** for the study of public emotions, since sharing stressful experiences amplifies the perceived stress and the accompanying emotions (Nahleen et al., 2019). Here we focus on four of the Nordic countries (Denmark, Finland, Norway and Sweden), referred to as the Nordic countries in the remainder of this text for simplicity purposes. On one hand, the Nordic countries share characteristics such as healthy democracies (Transparency International, 2019), comprehensive healthcare systems (OECD, 2021b), high levels of education (OECD, 2021a) and high levels of trust in organizations and the media (Delhey & Newton, 2005; OECD, 2020). Furthermore, during the Covid19 pandemic, the legacy media managed to hold its place as the main source of information (Ohlsson et al., 2021). These common factors unify the Nordic countries in terms of expected resilience in the face of a global crisis like the covid-19 pandemic. On the other hand, the Nordic countries managed the Covid-19 crisis in different ways. In particular, the degree to which measures for managing the pandemic were enforced by law has varied widely; from the rapid work and school lockdown in Denmark and Norway to the more lax approaches followed in Finland, and specially, in Sweden (Saunes et al., 2022). This had varying degrees of impact on each of these countries' economies (Ohlsson et al., 2021). These factors of similarity and divergence allow us to use the Nordic context as a series of case studies of high trust countries with slight variations in Covid-19 management.

Previous research on data from other countries has shown that the stringency of the restrictions has an impact on public emotions and sometimes even mental health (Aknin et al., 2022; Brooks et al., 2020; Marroquín et al., 2020), that can be measured through social media data (Lwin et al., 2020; Zhang et al., 2022). Higher policy stringency led on average to higher psychological distress, while perception of physical distancing measures and the government's handling mediate this relation (Aknin et al., 2022). However, most of the data used for these studies comes from countries with lower indices of trust in the government and the media (e.g., the US; (Zhang et al., 2022), and as such they are not generalizable to the context of high trust and resilience expected from the Nordic countries. Furthermore, the relationship between mental health and the stringency of measures has been shown to be mediated by trust in the government (O'Hara et al., 2020). Understanding the emotional implication of public health crises is important for tailoring public health messages and services that cover the needs of the population, and investigating resilient states is as important as studying those less able to cope.

In the context of the Nordic countries, available evidence remains scarce. A study focusing on the emotional landscape during the first wave of the pandemic, and evaluating it in Sweden and Norway, found that the positive sentiment dropped below the average for Norway much earlier than for Sweden. This was interpreted by the authors as a consequence of an earlier lockdown being imposed in Norway (Imran et al., 2020) and is in accordance with studies from other countries showing an impact of stringency of the restrictions on public emotion. Another study including data from the Nordic countries separated countries according to whether they were following an elimination strategy (countries that aimed to eliminate community transmission of SARS-CoV-2 within their borders) or a mitigation strategy (countries that aimed to control SARS-CoV-2 transmission). They found that life evaluations remained stable in the countries with a near-elimination strategy (i.e. Norway, Denmark and Finland) but showed a significant decrease in Sweden, who followed a mitigation strategy (Aknin et al., 2022). When zooming into the association between policy stringency and psychological distress, they found that it did not differ significantly between countries that followed a mitigation strategy or near-elimination strategy. While the study by Imran et al., (2020) focused on Twitter data collected during an earlier stage of the pandemic (12th of March to 29th April 2020) the study by Aknin et al., (2022) used a battery of surveys and clinical measures over a more extended period of time, from the 27th April 2020 to 28th June 2021.

Here we chose to investigate public emotions using a large Twitter dataset collected between August 2020 and March 2021. Signatures of public emotions have been shown to be present throughout social media platforms, such as Twitter data originating from different countries (Xue et al., 2020; Zhang et al., 2022). This provides an opportunity to study public emotions at scale and across time, in contrast to survey data, which is too costly to collect on such a continuous basis. In particular, over the course of the covid-19 crisis, many studies have highlighted fear and anxiety or distress as one of the main emotional responses in the population (Aknin et al., 2022; Johansson et al., 2021; Sher, 2020; Widmann, 2022) some of them observing signs of this fear and anxiety on Twitter (Lwin et al., 2020; Zhang et

NORDIS

al., 2022). The study by Lwin et al., 2020 found furthermore that public emotions shifted strongly from fear to anger, with sadness and joy also coming to the surface over the course of two months at the beginning of the pandemic. However, these studies were conducted on data from mixed geolocations (Xue et al., 2020; Lwin et al., 2020) or from countries where less resilience would be expected in the face of crisis due to e.g., their lower indices on trust in the government and the media among other factors (e.g., Zhang et al., 2020, Twitter data from the US). It is also worth noting that most of these studies address earlier stages of the covid-19 crisis, but little is known about the emotional responses at later stages of the pandemic, when the initial coping mechanism has been exhausted, with the population's resilience having been probed for months.

Many of the previous studies studying public emotions on Twitter from a language psychology perspective have used Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015). While this tool has the advantage of being simple to implement and scalable, it often detects false signals as words can appear in different contexts and take on different meanings. Furthermore, the latest version (LIWC 2015) has not been translated into all languages present in this study and, as previous research has shown, within-language standardization is needed when analyzing texts from different languages (Dudău & Sava, 2021). Applying the LIWC in this dataset would have required translation of the entire dataset, with the consequent loss of information inherent to automatic translation of such short and context-dependent excerpts of text. Therefore, instead of using a lexicon-based approach, we decided to use the XLM-RoBERTa-base model (Conneau et al., 2020). This model is based on a deep Bidirectional Encoder Representation from Transformers (BERT) model (Devlin et al., 2019). The model shares the advantages of traditional transformer models, which use attention mechanisms to extract relational context and even long-range dependencies in a sentence. Additional advantages for bidirectional transformer models are its bidirectionality, which instead of only accessing the left context of a token allows to fuse the left and the right context of a sentence, allowing for a better context extractor. The availability of pretrained multilingual models, as is the case of XLM-RoBERTa, allowed us to use the same model to analyse tweets in 4 different languages. Additionally, we opted for a sampling method aimed at collecting as many tweets as possible without any hashtag or covid-related keyword specifications (see Methods), which avoids introducing bias related to those hashtags in our dataset. We opted for this approach as it is challenging with such an ubiquitous crisis to determine when something is unrelated to it; furthermore, we expect anxiety and other Covid-19 related emotions to permeate all aspects of life.

The aim of this study is twofold. First, we aim to investigate the emotional landscape of the Nordic countries during a subacute phase of the pandemic. Based on previous research highlighting fear and anxiety in earlier stages of the pandemic (Aknin et al., 2022; Harper et al., 2020; Johansson et al., 2021; Sher, 2020; Widmann, 2022), we were interested in evaluating whether traces of such anxiety can be observed on Twitter data from the Nordic countries. In contrast to previous studies, which select covid-19 related tweets, we trace a comprehensive Twitter dataset (see Method, Twitter data) since we expect any traces of anxiety to permeate nearly all aspects of life, especially given the implemented measures of social distancing. Secondly, this study aims to investigate whether the public emotion landscape in the Nordic countries can be used as a measure of well-being and resilience. To do so, we investigate the relationship between public emotion and two variables related to the covid-19 pandemic, namely the number of deaths and mobility data.

2.0 Results

As detailed in the Methods section, evaluation of the performance of the XLM-RoBERTa model on the validation set revealed high variability in the f1-scores used to assess the ability of the model to detect each individual emotion. We decided to focus only on those emotions that the model could detect best by setting a threshold at f1 > 0.65, above which we consider that the model can reliably detect a particular emotion in a tweet. Unfortunately, anticipation obtained an extremely low score (f1= 0.10) and had to be excluded from further analysis. This precluded the possibility to find any tweets expressing anxiety, at least according to the theoretical model of emotions (Plutchik, 2001) adopted in this study, which argues that anxiety consists of a combination of fear and anticipation. In the following, we report the results taking fear as a proxy for anxiety.

Our results indicate that fear was not a predominant emotion at any point between August 2020 and March 2021, and in fact, our model did not detect any tweets expressing fear for that time period for any of the Nordic countries **(Figure 1)**. Emotions present in our data were joy, optimism, anger and disgust. Linear regressions on the emotional time series show significant positive correlations when looking at emotions with the same polarity (Anger vs. Disgust and Joy vs. Optimism), and significantly anticorrelated when looking at emotions with opposite polarities (i.e., Anger vs. Joy) (see **Supplementary table 1**).

Figure 1. We did not find any traces of fear on Nordic Twitter for the period starting August 2020 to March 2021. *The time series shows the percentage of tweets per day classified as containing one of the 5 emotions that our model was able to detect reliably (anger, disgust, fear, joy, optimism). We observe that fear was not present in any of the Nordic countries' data. Furthermore, the percentage of tweets expressing positive emotions (joy and optimism) was higher than the percentage of tweets expressing negative emotions (anger and disgust) across the entire time frame for each of the Nordic countries.*

In the following, we focus the analysis on the main positive and negative emotion timeseries, namely joy and anger. From **Supplementary figure 2**, we can see that the amount of joy is consistently higher across countries, although the difference might differ between countries. First we tested whether this difference between joy and anger is significant for all 4 countries. Our results show that this is indeed the case, with all countries showing a significant difference between joy and anger as tested using non-parametric Wilcoxon Rank-sum test (*Denmark: p < 4.12e-80, effect size = 0.8651247, Norway: p < 4.34e-80, effect size = 0.8650045, Finland: p < 4.12e-80 effect size = 0.8651247, Sweden: p < 8.46e-67, effect size = 0.7881046*). It is important to note though, that given the difference in model performance regarding each individual emotion (i.e., f1=0.81 for joy and f1=0.74 for anger) we cannot conclude that there is more joy than anger on the Nordic Twitter sphere. However, we can assume that the model should be performing comparably across languages, and hence we can compare that difference between countries to evaluate whether the relationship between joy and anger is significantly different among them.

As shown in **Figure 2** and **Table 1**, we compare the difference between the expression of joy and the expression of anger across the four nordic countries and find a significant difference for all pairs. However, the effect sizes differed across comparisons. *In particular, we see the largest effect sizes when comparing Sweden to all other countries, supporting that Sweden has the smallest difference between joy and anger across the Nordic countries. The comparison between FInland and Norway also reached a large effect size, whereas the comparison Denmark vs. Norway and Denmark vs. Finland only reached moderate effect sizes (see **table 1**).*

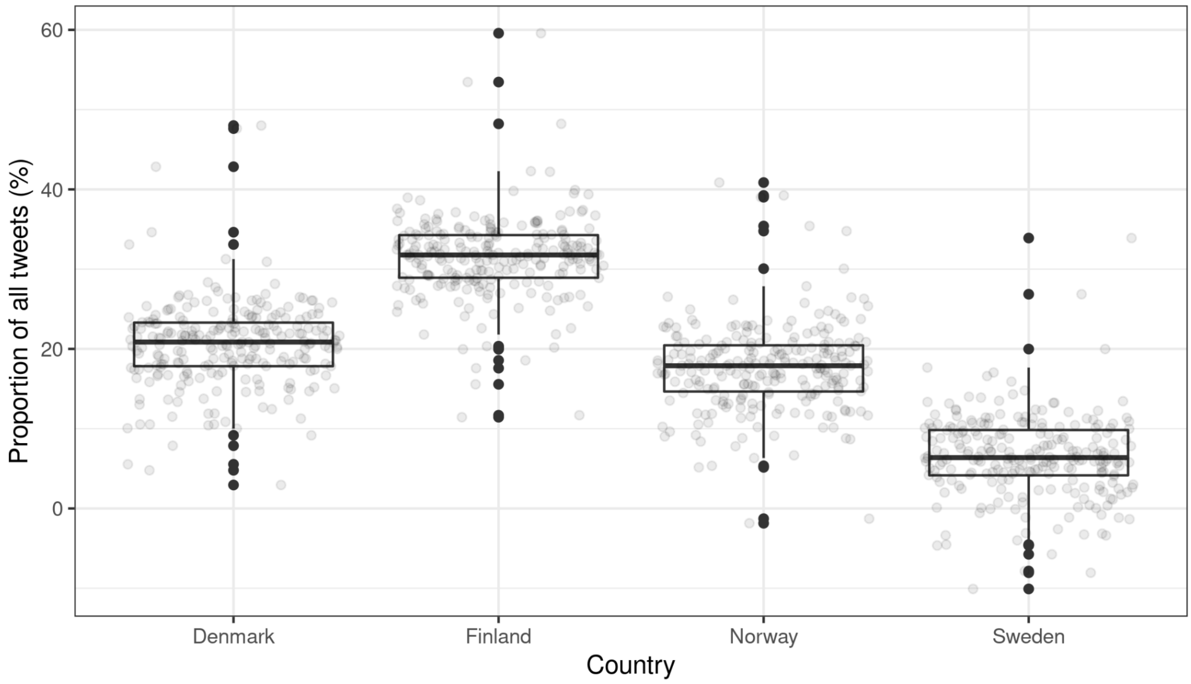


Figure 2. The biggest difference between the expression of joy and the expression of anger on Twitter is found in Finland, followed by Denmark, Norway and Sweden, in

that order. Difference between the expression of joy and the expression of anger in different countries across the time-period of 240 days ranging between the 8th of August 2020 and the 31st March 2021. Each data point corresponds to the percentage of tweets from a single day labeled with joy minus the percentage of tweets labeled with anger. Using a pairwise Wilcoxon rank sum test, revealed that these differences were significantly different for all pairs of Nordic countries although the effect sizes differed across comparisons (see **Table 1**).

Table 1: Differences in the level of joy vs anger between countries

DK-FI	$W = 3322$, $p < 4.23e-63$* effectsize = 0.4585161 (large)
DK-NO	$W = 38907$, $p < 2.90e-11$* effectsize = 0.4548513 (moderate)
DK-SV	$W = 55797$, $p < 1.27e-70$* effectsize = 0.8161610 (large)
FI-NO	$W = 55218$, $p < 1.05e-67$* effectsize = 0.6893061 (large)
FI-SV	$W = 57339$, $p < 1.06e-78$* effectsize = 0.8364074 (large)
NO-SV	$W = 54359$, $p < 1.72e-63$ * effectsize = 0.7175729 (large)

Following these results we evaluate the relationship between emotion expressed on Twitter and two indicators of the evolution of the Covid-19 pandemic in the Nordic countries, to understand whether they have an impact on the public emotional landscape. First, we look at the relationship between joy - the emotion which our model was able to detect best - and the number of daily deaths, for which we would expect a negative relationship **(Figure 4).** However, we found that the correlations between these two variables are very low, ranging from r=-0.0143 in Norway to r=0.2481 in Sweden. Second, we look at the relationship between joy and time spent at home during this time period **(Figure 5)**. Again, we find very low correlations, ranging from r=0.0740 in Norway to r=0.02285 in Sweden.

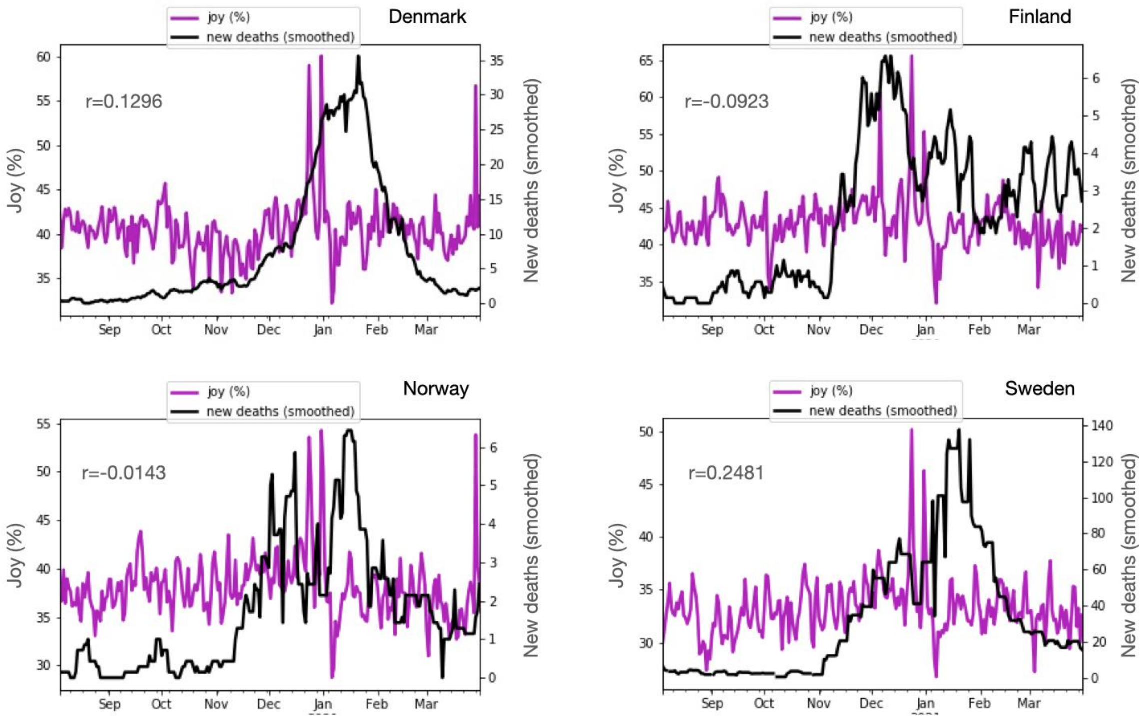


Figure 4. The expression of joy on Twitter is not correlated with fluctuations in the number of daily deaths due to Covid-19 in any of the Nordic countries. Pearson correlation coefficients for each country were as follows: Denmark r=0.1296, Finland r=-0.0923, Norway r=-0.0143 and Sweden r=0.2481.

NORDIS

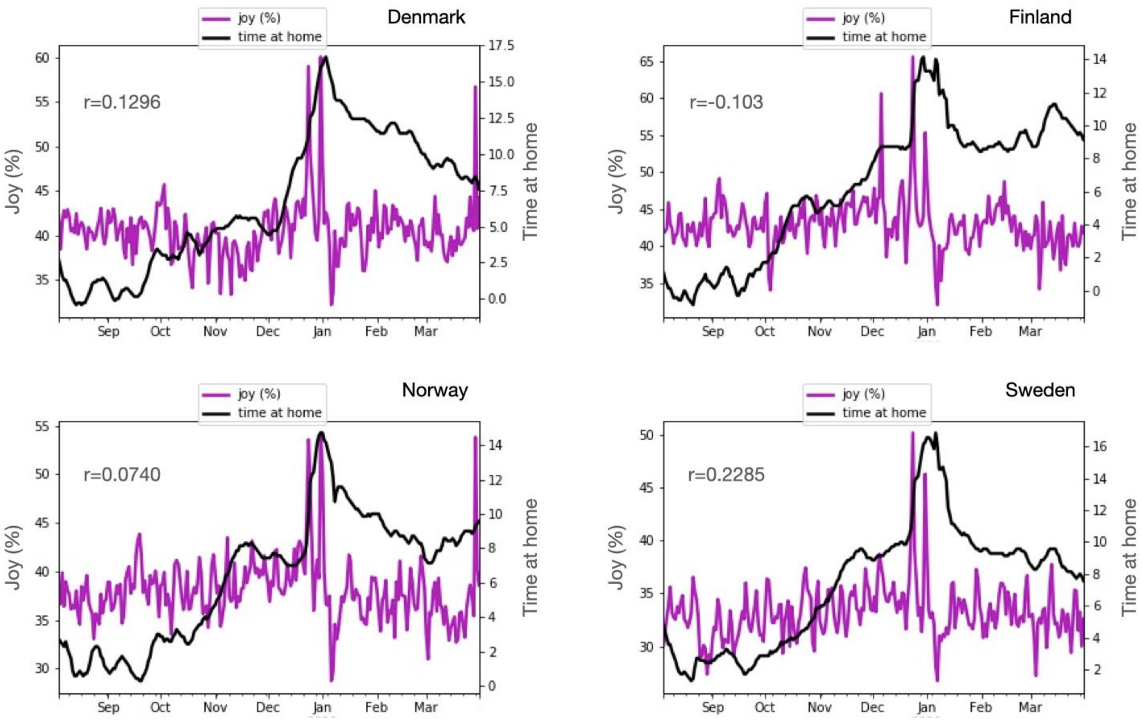


Figure 5. The expression of joy on Twitter is not correlated with the amount of time spent at home due, at least partly, to the social distancing measures applied by each government. Pearson correlation coefficients for each country were as follows: Denmark r=0.1296, Finland r=-0.103, Norway r=-0.0740 and Sweden r=0.2285.

3.0 Discussion

The main goal of this study was to investigate public emotions on Twitter of the Nordic countries during the second wave of the pandemic, with a special focus on any potential traces of fear and anxiety in a context of high trust and resilience. While previous studies found that fear and anxiety were very prevalent on Twitter during earlier stages of the Covid-19 crisis (e.g., Zhang *et al.*, 2022), our results show that this was not the case for tweets from the Nordic countries during the second wave of the pandemic. In fact, our model detected no tweets containing fear across all the Twitter data from all 4 Nordic countries **(Figure 1)**. This could be due to the high levels of trust on the government and the media that characterise the Nordic countries (Delhey & Newton, 2005; OECD, 2020), which has been shown to mediate the relationship between mental health and stringency of measures during the Covid-19 crisis (O'Hara et al., 2020). In other words, the high resilience that the population from the Nordic countries expect from their governments, health care system and other institutions, prevent anxiety from being spread, and even expressed often, on social media. While this is a possible explanation, we cannot exclude other factors. In particular, our results could also be due to the Twitter culture of the Nordic countries being more oriented towards sharing information rather than personal views and emotions than in other countries.

Beyond the lack of fear and anxiety, we observed a strong predominance of joy and anger, which was the case across all Nordic countries**.** This observation, while it could be reflecting the nature of the Nordic Twittersphere to some degree, it is likely to be strongly influenced by the performance of the model. Joy and anger are the two emotions that the model can detect best, and as such, it is not surprising that these two emotions appear to be predominant in our data. An additional consideration is the physiological arousal of these two emotions; previous research has shown that content that elicits high-arousal, whether positive (joy) or negative (anger, anxiety or fear), becomes more viral (Berger & Milkman, 2012) e.g., is retweeted more. Since we aimed to take the emotional pulse of Nordic Twitter, we did not remove retweets from our sample, and hence higher-arousal emotions were likely to have more weight than low-arousal emotions (e.g., sadness).

Finland appeared to have the largest difference between levels of joy and anger, although the difference between the levels of these two emotions were significant for all Nordic countries **(Figure 2).** We also compared the difference between joy and anger levels between countries, and our analyses show that the comparison was significant for all country pairs **(Figure 3),** although the effect size varied from large (e.g., Sweden vs. all other countries) to moderate (e.g., Denmark vs. Norway and Finland)**. H**ere, due to possible differences in model performance across languages, we can not be certain that the results are uniquely driven by differential amounts of emotions between countries. We further found no relationship between levels of joy and the number of daily deaths, and levels of joy and the amount of time spent at home. This suggests that, at least when considering a comprehensive sample of the Nordic Twittersphere instead of focusing solely on Covid-19 related tweets, we cannot detect emotion fluctuations related to the evolution of the pandemic. This is the case independently of the stringency of the measures adopted to counteract the pandemic adopted by each individual country.

This study provides insights about the public emotions in the Nordic countries during the second wave of the pandemic. It does so through an exploratory analysis of - to our knowledge - the largest Nordic Twitter dataset collected during the Covid-19 crisis. However, this study also has some limitations, some of which could be addressed in future studies. In particular, the availability of training data in the Nordic languages is very scarce, and as such we had to automatically translate an English dataset into the languages of the study, with the consequent loss of information that this involves. Furthermore, our training dataset was unbalanced, containing different amounts of tweets expressing each of the emotions. As such, the model was less able to detect emotions such as anticipation, trust and even fear, which were of interest for this study. In addition, some phenomena like the use of irony and sarcasm, very present on Twitter during earlier stages of the pandemic at least in other countries (Vicari & Murru, 2020), will have gone unnoticed and might have resulted in some tweets being erroneously labeled as joy, when they should have been labeled as anger. Also, in order to get more insights into the Twittersphere, crisis and emotions in the Nordic (digital) media countries, we need comparative studies including Nordic and other European countries. We also need to consider the representability of the Twitter data for the general population. While our study suggests that the public emotion was not dominated by fear, without any traces of anxiety having been found, we cannot exclude the possibility that a different scenario might be revealed by looking at data from different social media platforms, such as Facebook that unfortunately is not widely available for academics. Future studies could make use of a better training dataset and investigate public emotions using data from various social media platforms when available, as well as include data from a more diverse sample of countries. All in all, this study provides preliminary evidence supporting the absence of public anxiety and fear and the stability of public joy across time during the second wave of the pandemic as studied through Twitter in the Nordic countries.

4.0 Materials and methods

We used the XLM-RoBERTa base model proposed by Conneau et al., (2019). It is a large multi-lingual language model, trained on 2.5TB of filtered CommonCrawl data and based on the initial RoBERTa model released in 2019 (arXiv:1907.11692v1).

4.1 Training data

We used the SemEval 2018, task 1 (Mohammad & Bravo-Marquez, 2017) as the training dataset to finetune the model for emotion detection. This dataset was manually annotated through a crowdsourcing project, with each tweet being annotated by, on average, 7 annotators (for further details, see Mohammad and Bravo-Marquez, 2017). The possible emotions were the following: 1) anger (also includes annoyance, rage) 2) anticipation (also includes interest, vigilance) 3) disgust (also includes disinterest, dislike, loathing) 4) fear (also includes apprehension, anxiety, terror) 5) joy (also includes serenity, ecstasy) 6) love (also includes affection) 7) optimism (also includes hopefulness, confidence) 8) pessimism (also includes cynicism, no confidence) 9) sadness (also includes pensiveness, grief) 10) surprise (also includes distraction, amazement) 11) trust (also includes acceptance, liking, admiration) 12) neutral or no emotion. Each of the 11 emotions were considered whether present in each of the tweets, resulting in a multilabel training dataset in which each tweet can contribute to multiple emotions. Since we needed the training dataset to be a available in the four nordic languages of this study, we translated the English version into all 4 nordic languages using the eTranslation tool provided by the European Commision (https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-etranslation_en). The english version of the dataset consists of 10,983 tweets in total, 6,838 for the training, 886 for the validation and 3,259 for the test set. The training set, used for the finetuning of the model, contains a total of 160 anxiety tweets (i.e., tweets labeled as both expressing anticipation and fear). Accordingly, we expect the fine-tuned xlm-Roberta model to be able to detect anxiety if present in our Twitter dataset.

4.2 Twitter data

In this study we analysed a large Twitter dataset containing tweets in 4 of the nordic languages: Danish, Norwegian, Swedish and Finnish. The tweets in Danish, Norwegian, and Swedish were collected through the HOPE project (https://hope-project.dk/#/) using a set of stopwords in each of the languages to scrape through the Twitter API before the language tag was available. The stopwords were sourced from the Open Subtitles website (http://www.opensubtitles.org), and generated by selecting the top 100 most frequent words in the 4 Nordic languages of this study, full list available at the following github repository: https://github.com/centre-for-humanities-computing/stopwords-danish-distinct. For each language, words were cross-checked with the lists from any of the other Nordic languages lists relevant for this study and removed if the word was present across languages. The reason for doing so was to create a list of frequently used words in each of the languages that was still as differentiated as possible from the other Nordic languages. In the case of Norwegian, additional Nynorsk words were added to the Bokmal stopwords list after checking for their absence in the other languages. The tweets in Finnish were collected using the equivalent set of stopwords in that language.

Our dataset includes tweets that were posted between August 2020 and March 2021. This time period was chosen due to the inclusion of a low risk phase (summer 2020) as well as a second wave accompanied by social distancing measures. Furthermore, the population had not been vaccinated yet, and hence we expected to find traces of fear or anxiety in the population of the Nordic countries. The Twitter dataset contained a total of 57.828.980 tweets (29.088.137 in Swedish, 6.168.893 in Norwegian, 7.369.613 in Danish, and 15.202.337 in Finish).

4.3 Data analysis

First, we fine tuned the XLM-RoBERTa base for emotion detection using a learning rate of 2e5 as in the original publication (Devlin et al., 2019). We used a batch size of 16 and fine tuned over 4 epochs, but set an early stopping callback based on the evaluation loss, a strategy commonly used to prevent overfitting of the model. For evaluation of the fine tuning on the validation set, a threshold of 0.5 was set above which a tweet was considered to have been assigned to a specific emotion category according to the model. We then evaluated the performance of the model on the validation set focusing primarily on the f1-score, choosing this metric due to our imbalanced dataset. The f1-score combines the precision and the recall metrics - the two most common metrics that take into account class imbalance - into a single metric. In our case, this was important given the unequal amount of different emotions that can be found on our Twitter training dataset used for the emotion detection fine-tuning. Performance metrics after the finetuning can be found in the supplementary material (**Supplementary figure 1**). As shown in the Supplementary material, only 5 out of the 11 emotions obtained an f1-score above 0.65, which we set as the threshold above which we consider the model to be able to reliably detect the emotion. These were anger, disgust, fear, joy and optimism.

Before applying the fine-tuned model, we preprocessed the Twitter dataset by removing mentions and URLs. We then decoded emojis present in the text using Demoji 1.1.0 (https://github.com/bsolomon1124/demoji). After the text preprocessing was done, we estimated the probabilities for each emotion to be present in each tweet. Tweets with all emotional probabilities below 0.5 were considered neutral tweets and excluded from further analysis. We then selected tweets with a maximal score in one of the 5 emotions that the model is able to reliably estimate and focus our analysis in that subsample of tweets. This means, in the following, we present results only concerning the following emotions: anger, disgust, fear, joy and optimism.

Statistical analyses were carried out using Python and R software. We used non-parametric two-sample t-tests, in particular Wilcoxon rank sum for comparisons within countries and Kruskal-Wallis for comparisons between countries. Correction for multiple comparisons we carried out using Bonferroni correction; given that we carry out a total of 11 t-tests, the threshold for significance is $p < 0.05/11$ ($p = 0.0045$).

4.0 References

Aknin, L. B., Andretti, B., Goldszmidt, R., Helliwell, J. F., Petherick, A., De Neve, J.-E., Dunn, E. W., Fancourt, D., Goldberg, E., Jones, S. P., Karadag, O., Karam, E., Layard, R., Saxena, S., Thornton, E., Whillans, A., & Zaki, J. (2022). Policy stringency and mental health during the COVID-19 pandemic: A longitudinal analysis of data from 15 countries. *The Lancet Public Health*, *7*(5), e417–e426. https://doi.org/10.1016/S2468-2667(22)00060-3

Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet*, *395*(10227), 912–920. https://doi.org/10.1016/S0140-6736(20)30460-8

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). *Unsupervised Cross-lingual Representation Learning at Scale* (arXiv:1911.02116). arXiv. http://arxiv.org/abs/1911.02116

Delhey, J., & Newton, K. (2005). Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism? *European Sociological Review*, *21*(4), 311–327. https://doi.org/10.1093/esr/jci022

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. http://arxiv.org/abs/1810.04805

Dudău, D. P., & Sava, F. A. (2021). Performing Multilingual Analysis With Linguistic Inquiry and Word Count 2015 (LIWC2015). An Equivalence Study of Four Languages. *Frontiers in Psychology*, *12*, 570568. https://doi.org/10.3389/fpsyg.2021.570568

Fan, R., Zhao, J., Chen, Y., & Xu, K. (2014). Anger Is More Influential than Joy: Sentiment Correlation in Weibo. *PLoS ONE*, *9*(10), e110184. https://doi.org/10.1371/journal.pone.0110184

Gregory, K. (2018). *Anger in Academic Twitter: Sharing, Caring, and Getting Mad Online*. 18.

Harper, C. A., Satchell, L. P., Fido, D., & Latzman, R. D. (2020). Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-020-00281-5

Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets. *IEEE Access*, *8*, 181074–181090. https://doi.org/10.1109/ACCESS.2020.3027350

Johansson, F., Côté, P., Hogg-Johnson, S., Rudman, A., Holm, L. w., Grotle, M., Jensen, I., Sundberg, T., Edlund, klara, & Skillgate, E. (2021). Depression, anxiety and stress among Swedish university students before and during six months of the COVID-19 pandemic: A cohort study. *Scandinavian Journal of Public Health*, *49*(7), 741–749. https://doi.org/10.1177/14034948211015814

Larsen, M. E., Boonstra, T. W., Batterham, P. J., ODea, B., Paris, C., & Christensen, H. (2015). We Feel: Mapping Emotion on Twitter. *IEEE Journal of Biomedical and Health Informatics*, *19*(4), 1246–1252. https://doi.org/10.1109/JBHI.2015.2403839

Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020).

Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. *JMIR Public Health and Surveillance*, *6*(2), e19447. https://doi.org/10.2196/19447

Marroquín, B., Vine, V., & Morgan, R. (2020). Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources. *Psychiatry Research*, *293*, 113419. https://doi.org/10.1016/j.psychres.2020.113419

Mohammad, S. M., & Bravo-Marquez, F. (2017). Emotion Intensities in Tweets. *ArXiv:1708.03696 [Cs]*. http://arxiv.org/abs/1708.03696

OECD. (2020). *Trust in government* [Doi:https://doi.org/10.1787/1de9675e-en]. https://www.oecd-ilibrary.org/content/data/1de9675e-en

OECD. (2021a). *Education at a Glance 2021: OECD Indicators*. OECD. https://doi.org/10.1787/b35a14e5-en

OECD. (2021b). *Health at a Glance 2021: OECD Indicators*. OECD. https://doi.org/10.1787/ae3016b9-en

OECD. (2022). *Building Trust to Reinforce Democracy*. https://doi.org/10.1787/b407f99c-en

O'Hara, L., Abdul Rahim, H. F., & Shi, Z. (2020). *Gender and trust in government modify the association between mental health and stringency of social distancing related public health measures to reduce COVID-19: A global online survey* [Preprint]. Public and Global Health. https://doi.org/10.1101/2020.07.16.20155200

Ohlsson, J., Blach-Ørsten, M., & Willig, I. (2021). *Covid-19 och de nordiska nyhetsmedierna* (p. 138). Nordicom, University of Gothenburg. https://doi.org/10.48335/9789188855473

Pennebaker, J., Booth, R., Boyd, R., & Francis, M. (2015). *Linguistic Inquiry and Word*

Count: LIWC2015.

Saunes, I. S., Vrangbæk, K., Byrkjeflot, H., Jervelund, S. S., Birk, H. O., Tynkkynen, L.-K., Keskimäki, I., Sigurgeirsdóttir, S., Janlöv, N., Ramsberg, J., Hernández-Quevedo, C., Merkur, S., Sagan, A., & Karanikolos, M. (2022). Nordic responses to Covid-19: Governance and policy measures in the early phases of the pandemic. *Health Policy*, *126*(5), 418–426. https://doi.org/10.1016/j.healthpol.2021.08.011

Sher, L. (2020). COVID-19, anxiety, sleep disturbances and suicide. *Sleep Medicine*, *70*, 124. https://doi.org/10.1016/j.sleep.2020.04.019

Syvertsen, T., Enli, G., Mjøs, O. J., & Moe, H. (2014). *The Media Welfare State: Nordic Media in the Digital Era*. University of Michigan Press. https://doi.org/10.2307/j.ctv65swsg

Vicari, S., & Murru, M. F. (2020). One Platform, a Thousand Worlds: On Twitter Irony in the Early Response to the COVID-19 Pandemic in Italy. *Social Media + Society*, *6*(3), 205630512094825. https://doi.org/10.1177/2056305120948254

Widmann, T. (2022). Fear, Hope, and COVID‐19: Emotional Elite Rhetoric and Its Impact on the Public During the First Wave of the COVID‐19 Pandemic. *Political Psychology*, pops.12831. https://doi.org/10.1111/pops.12831

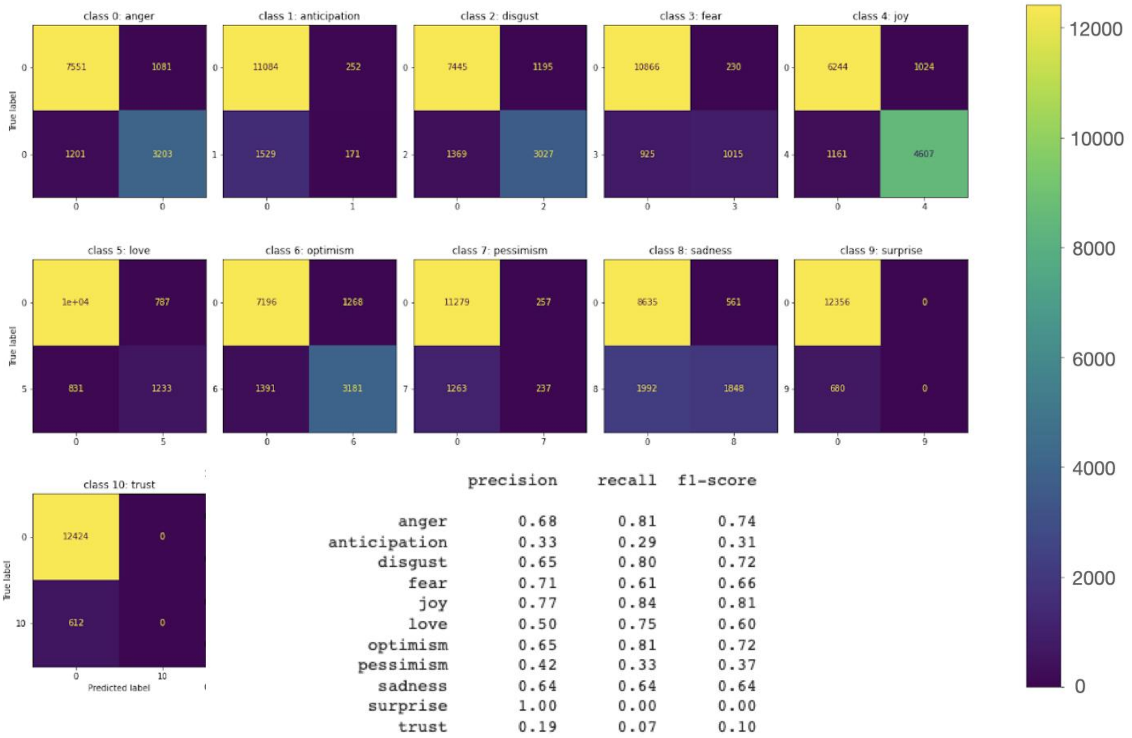
Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLOS ONE*, *15*(9), e0239441. https://doi.org/10.1371/journal.pone.0239441

Zhang, S., Sun, L., Zhang, D., Li, P., Liu, Y., Anand, A., Xie, Z., & Li, D. (2022). The COVID-19 Pandemic and Mental Health Concerns on Twitter in the United States.

NORDIS

Health Data Science, *2022*, 1–9. https://doi.org/10.34133/2022/9758408

1.4 Annexes

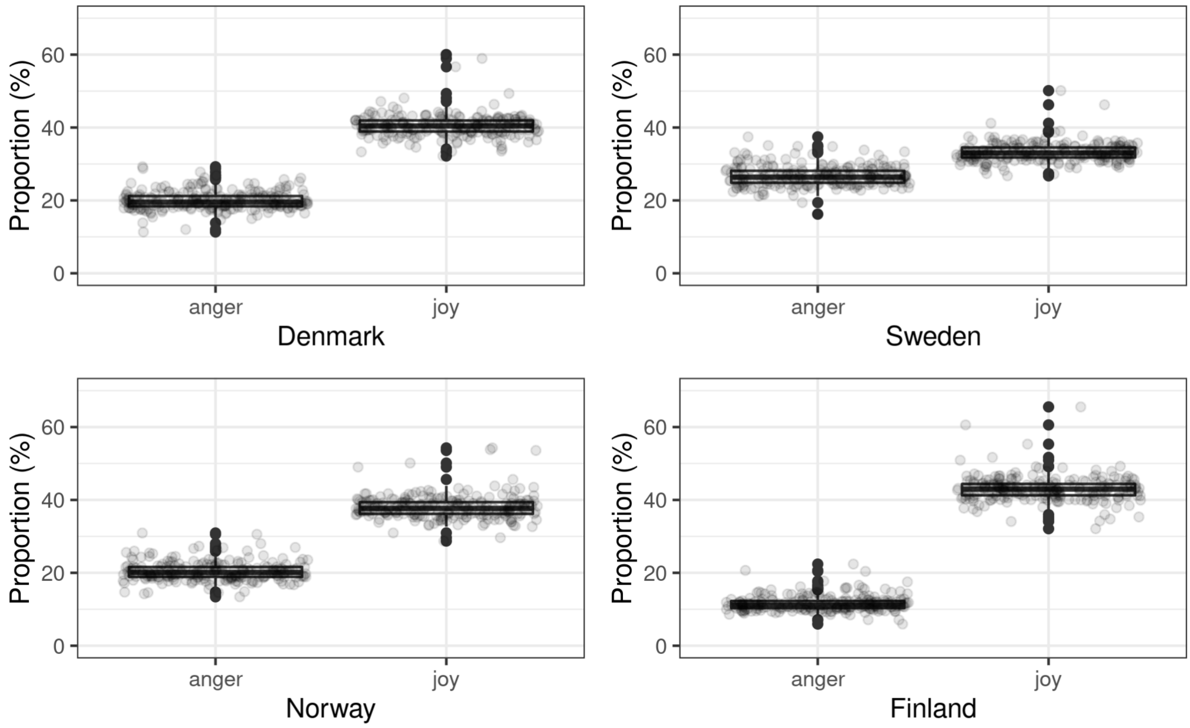


	precision	recall	f1-score
anger	0.68	0.81	0.74
anticipation	0.33	0.29	0.31
disgust	0.65	0.80	0.72
fear	0.71	0.61	0.66
joy	0.77	0.84	0.81
love	0.50	0.75	0.60
optimism	0.65	0.81	0.72
pessimism	0.42	0.33	0.37
sadness	0.64	0.64	0.64
surprise	1.00	0.00	0.00
trust	0.19	0.07	0.10

Supplementary figure 1. Confusion matrices and performance scores for each of the emotions

	Denmark	Finland	Norway	Sweden
Anger vs. Disgust	r=0.995, pvalue=1.344e-241	r=0.991, pvalue=2.212e-208	r=0.994, pvalue=7.412e-233	r=0.994, pvalue=1.348e-230
Joy vs. Optimism	rvalue=0.977, pvalue=9.154e-163	r=0.977, pvalue=1.264e-162	r=0.966, pvalue=1.611e-142	r=0.984, pvalue=4.898e-180
Anger vs. Joy	r=-0.836, pvalue=4.599e-64,	r=-0.733, pvalue=8.640e-42	r=-0.799, pvalue=1.599e-54,	r=-0.849, pvalue=6.161e-68

Supplementary table 1. Correlation coefficients and p-values resulting from the linear regressions on the emotion timeseries. Emotional time series show significant positive correlations when looking at emotions with the same polarity (Anger vs. Disgust and Joy vs. Optimism), and significantly anticorrelated when looking at emotions with opposite polarities (i.e., Anger vs. Joy)



Supplementary figure 2. All Nordic countries had significantly more tweets expressing joy than anger during the second wave of the pandemic. *Each data point corresponds to the proportion of all tweets on a day that were labeled either anger or joy. We used non-parametric t-tests (Wilcoxon Rank-sum test) to compare the amount of tweets expressing joy with the amount of tweets expressing anger and found that, while the amount of joy was significantly higher for all countries, the effect sizes varied across countries. Denmark: p < 4.12e-80, effect size = 0.8651247, Norway: p < 4.34e-80, effect size = 0.8650045, Finland: p < 4.12e-80 effect size = 0.8651247, Sweden: p < 8.46e-67, effect size = 0.7881046)*