Topic and Emotion Development among Dutch COVID-19 Twitter Communities in the early Pandemic

Boris Marinov⋆, Jennifer Spenader⋆, Tommaso Caselli⊖,
⋆Department of Artificial Intelligence, ⊖CLCG
⊖⋆University of Groningen, Groningen, The Netherlands
`boris.marinov96@gmail.com,`
`{j.k.spenader,t.caselli}@rug.nl,`

Abstract

The paper focuses on a large collection of Dutch tweets to gain insight into the perception and reactions of users during the early months of the COVID-19 pandemic. We focused on five major communities of users: government and health organizations, news media, politicians, the general public and conspiracy theory supporters, investigating differences among them in topic dominance and the expressions of emotions. Through topic modeling we monitor the evolution of the conversation about COVID-19 among these communities. Our results indicate that the focus on COVID-19 shifted from the virus itself to its impact on the economy between February and April. Surprisingly, the overall emotional public response appears to be substantially positive and expressing trust, although differences can be observed in specific groups of users.

1 Introduction

During the early COVID-19 pandemic, Twitter has played a key role in facilitating communication from government agencies and officials, but also among members of the general public. Twitter content has thus had a major influence on the public perception and sentiment surrounding the developing COVID-19 situation around the world. In this work, we conduct an exploratory study to investigate potential quantitative differences in topic dominance and emotional content between different user communities. To better focus our study, we restrict the the analysis to Dutch Twitter during the first three months of the COVID-19 pandemic.

Many individuals rely on Twitter for information. In the United States, 68% of people reported social media as their "news-outlets" (Matsa and Shearer, 2018), with a third of people also claiming that social media is an important source of health and science information (Hitlin and Olmstead, 2018). These figures are not surprising, since many official agencies, e.g. the government and health organizations, increasingly use Twitter to communicate key information to the public. This is in part because research has shown that Twitter can be particularly effective during rapidly developing events, such as disasters, political unrest, and outbreaks (Househ, 2016; LaLone et al., 2017; Daughton and Paul, 2019; Rogers et al., 2019). For example, during the 2015 Zika virus outbreak, credible organizations such as the World Health Organization (WHO) and the Center for Disease Control (CDC) used Twitter to circulate important health information (Stefanidis et al., 2017).

As there is no gate-keeping on the platform, content quality varies widely, from informal comments by private individuals to official communications from the government. Some content is made up of misinformation, rumors, and false statements, that are then massively circulated via the platform (Kumar and Geethakumari, 2014; Hamidian and Diab, 2016; Li et al., 2019). For this reason, studying the topics and emotional content of different user communities around a major public event (the COVID-19 pandemic) can give us important insights into both the public perception of the situation, the response of official agencies, and the general sentiment as the pandemic developed.

The first case in The Netherlands was reported on the 27th of February.[1] While a growing number of papers that analyze Twitter activity during the pandemic have already come out (Ordun et al., 2020; Chen et al., 2020; Schild et al., 2020), they largely focus on the situation in the United States, or other English speaking countries. Instead, we focus primarily on tweets from The Netherlands or Belgium, only considering Dutch language data. This allows us to study the direct relation between disease spread, government response and the public sentiment on a more local scale. Dutch tweets are almost exclusively tweeted by individuals or organizations within the Netherlands, the Flemish area of Belgium, or, to a lesser extent by Dutch individuals abroad.This contrasts with English tweets which are more heterogeneous since English has a larger group of speakers and users. In particular we address the following questions:

RQ1 How varied is the discourse around COVID-19 by Dutch speaking communities? What topics dominate?

RQ2 Do different users (e.g., politicians, journalists, influencers, active and casual users, among others) focus on different topics related to the COVID-19 outbreak?

RQ3 Do different users express different emotions about the COVID-19 outbreak?

The main contributions of this paper can be summarized as follows: (1) we conduct an extensive analysis of a newly created and publicly available Dutch Twitter dataset (Caselli and Basile, 2020) covering the first three months of the COVID-19 pandemic; (2) we found commonalities across user communities, but also different identifiable focuses, and (3) we identify community differences also in the emotional spectrum. Our analysis characterizes the evolution of the discourse around COVID-19 in Dutch speaking communities between February and April 2020, both in the topics that dominated and in the emotions expressed that reflect differences across Twitter communities during this virtual national conversation.

2 Related Works

Twitter has been largely used as a proxy of natural language data for the study of different socio-demographics issues of a target population. Previous work ranges from the identification of personality traits (Plank and Hovy, 2015), expressions of emotions (Hagen et al., 2015; Dini and Bittar, 2016), language use (Blodgett et al., 2016), age (Sloan et al., 2015), gender (Rangel et al., 2017), and authorship attributions (Schwartz et al., 2013; Rangel et al., 2017), among others. Previous attempts at community detection of Twitter users have used numerous methods and classification types, categorizing users based on their tweet context (Java et al., 2007), interaction and topics in the network (Darmon et al., 2014; Bakillah et al., 2015), sentiment oriented approaches (Abel et al., 2011; Cao et al., 2015) and following-to-followers ratio (Krishnamurthy et al., 2008).

Topic modeling has been widely and successfully used as a distant reading method to explore massive amounts of data and identify clusters of information (Yang et al., 2011; Ritter et al., 2010; Sarioglu et al., 2013; Schöch, 2017). The application of these approaches to Twitter data is a challenging task due to the short nature of the texts (Hong and Davison, 2010). Previous work has mainly applied Latent Dirichlet Allocation (LDA) based methods (Blei et al., 2003) with varying degree of success. The main issue is that LDA works under the assumption that a document can belong to multiple topics, where the membership is determined by some probability over the words. Since tweets are short snippets of text communicating often one single thought, they are instead more likely to contain only a single topic (2018). For this reason, we make use of the recently proposed semantics-assisted non-negative matrix factorization method (SeaNMF), (Shi et al., 2018).

There is already a growing number of published research studies where a range of different analyses, including sentiment and public response, have been applied to COVID-19 social media (Sharma et al., 2020; Ordun et al., 2020). However, as far as we know, no research has focused in detail on a single country and its social media response.

---

[1] https://www.rivm.nl/en/novel-coronavirus-covid-19/current-information

3 Data Collection and Cleaning

We conducted our analysis on a newly created dataset, `40twene_nl` (Caselli and Basile, 2020).[2] The messages have been extracted from an on-going and continuous collection of Twitter messages in Dutch (Sang, 2011; Bouma, 2015) by means of a selection of relevant keywords. Keyword identification was done manually by monitoring a trend website [3] from February to April 2020. For each day, top trending and COVID-19 related hashtags were extracted. To ensure variability and limit biases, the selection of the keywords was conducted by considering the trends during different moments of each day (i.e., morning, afternoon, evening, night). This resulted in 39 unique keywords which have been applied to obtain the messages in the selected periods of time. The keywords are not preprocessed (e.g., lemmatization or stemming) and instead only exact matches were looked for in hashtags or messages. Retweets have been excluded since they would have added noise. The dataset contains 2,390,596 tweets and only the Twitter IDs are distributed. We have used Hydrator[4] to retrieve the texts of the tweets and all associated metadata, including the geolocalization, the user screen name, the user name, the user description, the timestamp, the number of likes, the number of retweets, the geographical coordinates, and the place. The use of keywords may have introduced bias in the selection of the data by excluding non-Dutch speaking users living in The Netherlands or Belgium. This may result in a less representative analysis of the whole populations, with some minority voices "not being heard".

3.1 Preprocessing

Basic preprocessing is conducted to prepare the data. In particular, we have cleaned the data by (1) removal of duplicate tweets and tweets that have been deleted; (2) removal of hyperlinks (3) lower casing (4) removal of all keywords used for collecting the data; (5) removal of stop words;[5] (6) lemmatization and removal of closed-class words;[6] (7) removal of numerals and user mentions.

After preprocessing, 2,390,487 tweets remained (55,600 for February, 1,396,408 for March and 938,479 for April). Figure 1 illustrates the daily frequency of the tweets in the dataset, as well as some key dates of the evolution of the pandemic in the Netherlands.[7] Peaks around key dates indicate a link between the frequency and the occurring events, with a pattern similar to that of previous studies during past outbreaks (Shin et al., 2016).
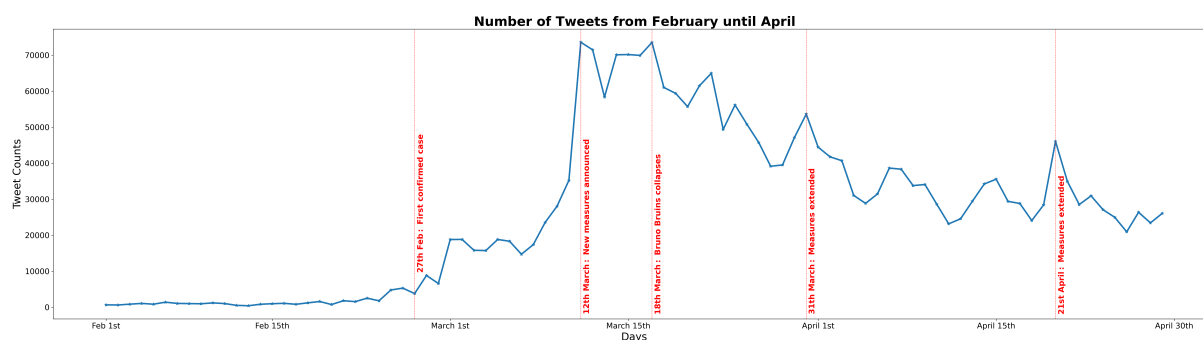

Figure 1: Daily Tweet frequencies from February to April

4 What's the Buzz About #COVID19_NL?

The `40twene_nl` collection has been specifically designed to identify messages about the COVID-19 in Dutch. The first step in our analysis is to determine whether this conversation about the pandemic has

---

[2]https://osf.io/pfnur/?view_only=5cc01c6cadc8441eb47659459fd5db10

[3]https://getdaytrends.com

[4]https://github.com/DocNow/hydrator

[5]We extended the original list for Dutch available in NLTK with additional data from this web-page: https://eikhart.com/blog/dutch-stopwords-list

[6]We used Frog (Bosch et al., 2007); closed-class words have been deleted if Frog assigned a probability higher or equal than 0.7. We did not further evaluate the quality of Frog on tweets.

[7]https://en.wikipedia.org/wiki/COVID-19_pandemic_in_the_Netherlands#1120_March

been characterized by the presence of different issues.

Hashtags and mentions are an important Twitter feature and can give a good initial overview into conversations on the platform. Mentions are used by users to notify others of the post, while hashtags allow for easier searching and grouping of tweets. Table 1 summarizes the top 5 hashtags and mentions for each month (hashtags containing *corona* or *covid* not included in count). For February the hashtags are more concerned with events happening outside of the Netherlands, however for March and April more relevant and local events can be seen (food hoarding, wearing of masks, working from home). The mentions for all months are largely centered around RIVM and particular politicians as one would expect. The number of mentions towards politicians does however increase in March and April compared to February, suggesting a reliance on the countries' leaders.

Table 1: Top 5 hashtags and mentions across February, March and April, with English translations\ explanations.

February		March		April	
Hashtag	English	**Hashtag**	English	**Hashtag**	English
1. #rivm	*Dutch CDC*	1. #rivm	*Dutch CDC*	1. #blifthuis	*stay home*
2. #china	*China*	2. #blifthuis	*stay home*	2. #persconferentie	*press conference*
3. #virus	*virus*	3. #hamsteren	*hoarding*	3. #rivm	*Dutch CDC*
4. #wuhan	*Wuhan*	4. #rutte	*Dutch PM*	4. #rutte	*Dutch PM*
5. #italie	*Italy*	5. #thuiswerken	*work from home*	5. #thuiswerken	*work from home*
Mention	English	**Mention**	English	**Mention**	English
1. @rivm	*Dutch CDC*	1. @rivm	*Dutch CDC*	1. @rivm	*Dutch CDC*
2. @nos	*Nat. TV News*	2. @minpres	*Prime minister*	2. @minpres	*Prime minister*
3. @telegraaf	*Nat. Newspaper*	3. @telegraf	*Nat.newspaer*	3. @telegraaf	*Nat.newspaper*
4. @nunl	*Nu.nl, News site*	4. @nos	*Nat.TV news*	4. @nos	*Dutch PM*
5. @coronanederland	*Corona Netherlands*	5. @theirrybaudet	*Right wing politician*	5. @hugodejonge	*Health minister*

4.1 Topic Modeling

We next conducted a topic modeling analysis as a method to aggregate messages and identify informative clusters of similar words, i.e., topics. Given that the keywords in `40twene_nl` are stable across time, this collection appears to be particularly suitable both for monitoring topic changes over time (i.e., from February to April 2020) and for comparing topic dominance differences between groups of users.

We apply a recently proposed method based on semantics-assisted non-negative matrix factorization (SeaNMF) (Shi et al., 2018). Existing NMF methods learn topics by decomposing the term-document matrix into lower ranked matrices, demonstrating strong performances in dimension reductions and clustering for high-dimensional data (Choo et al., 2015), with the approach being successfully applied to topic modeling (Kim et al., 2015). SeaNMF builds on top of this by leveraging word-context semantic correlations during training, overcoming the sparse problems of short texts and outperforming typically used LDA models.

Two important parameters of SeaNMF are the number of topics and the α value (weight for factorizing the word semantic correlation matrix). Setting α to 1 leads to best results,[8] while the number of topics is varied to investigate if there are differences across the three months concerning the narratives around COVID-19. Four evaluations are used to judge the fit of the model: **Average Pointwise Mutual Information (APMI)** and **Normalized PMI (NPI)** indicate how well the words in each topic relate to each other (relying on co-occurance); while **topic diversity (TD)** and **rank-biased overlap (RBO)** measure the diversity of the topics (Bianchi et al., 2020). We report in Table 2 the results of the evaluation of the SeaNMF models for each month separately. The number of topics ranges between 30 and 150, with increasing steps of 20, thus exploring different granularities of aggregation.

4.2 Topic Modeling Findings

In general, we obtain relatively high scores compared to other works on clustering twitter data (Cheng et al., 2014; Bianchi et al., 2020).We consider this as additional evidence of the homogeneity of the

---

[8]The threshold of the α value is based on evidence from literature, namely Shi et al. (2018)

Table 2: Topic modeling evaluation per month (scores for the optimal number of topics are in bold). NPI was used to finally select the optimal number of topics

Number Topics	February				March				April			
	APMI	NPMI	TD	RBO	APMI	NPMI	TD	RBO	APMI	NPMI	TD	RBO
30	3.0529	0.2589	0.9933	0.9998	2.0278	0.1637	0.8833	0.9930	2.2268	0.1752	0.9500	0.9975
50	3.3987	0.2669	0.9860	0.9997	2.1632	0.1854	0.8640	0.9938	2.4016	0.2055	0.9140	0.9954
70	3.7174	0.2719	0.9786	0.9996	2.4076	0.2111	0.8514	0.9937	2.6059	0.2242	0.9243	0.9980
90	3.9307	0.2896	0.9756	0.9997	**2.4083**	**0.2138**	**0.8389**	**0.9952**	2.6777	0.2297	0.9111	0.9984
110	**4.1662**	**0.3093**	**0.9773**	**0.9998**	2.3602	0.2059	0.8291	0.9958	**2.7358**	**0.2306**	**0.8955**	**0.9982**
130	4.1329	0.2911	0.9669	0.9997	2.3992	0.2093	0.8223	0.9964	2.7059	0.2269	0.8885	0.9983
150	4.2502	0.3016	0.9707	0.9997	2.5140	0.2177	0.8160	0.9969	2.777	0.2289	0.882	0.9985

`40twene_nl`, and as a cue that the topics we have induced are potentially indicating specific sub-topics on COVID-19 in Dutch.

The topic evaluation measures each emphasize different aspects. The scores for TD and its sister measure RBO are both high, indicating that the induced topic clusters are well differentiated from each other. However, TD has a different behavior when compared to RBO in assessing topic diversity. In February, TD and RBO substantially remained unchanged; while we observe a lowering of TD when larger numbers of topics are selected, but scores for both measures are in the range of 0.90s. A similar behavior can be observed in April, although the scores of both measures start to diverge at 110 topics, suggesting a lower quality of the generated topics. March, however, stands out. While RBO remains substantially unchanged across the number of topics, TD keeps degrading, suggesting less diversity in the topics. Similar observations hold for the scores of the APMI and NPMI. When combining all measures together, it appears that NPMI could be used to discriminate what would be an optimal number of topics for each month. In particular, as soon as the NPMI score stops increasing, we selected the corresponding number of topics as the best. For February and April we identify this threshold at 110, while for March this lies at 90.

The variation across the months is small, despite the fact that March has substantially more tweets than the other two months (1,396,408 *vs.* 55,600 for February and 938,479 for April). However, this is an indication of a variation in the way Twitter users were talking about the outbreak. Table 3 outlines the top 5 topics per month. Labels have been manually assigned by one of the authors according to the keywords. Looking at the table we see that in February, the outbreak was still perceived as something far away from the Netherlands or Belgium, with focus on China and global events. This changes in March where the situation moves from global to local, with a larger focus on the infections and dealing with the new way of life. In April, the infections are improving, with focus on the COVID-19 consequences on the economy and a renewed interest in the world-wide situation.

Table 3: Top 5 topics for each month

February	March	April
COVID-19 China: 31.26%	COVID-19 Netherlands: 15.02%	Economy: 11.58%
Early COVID-19: 20.36%	Infections: 9.79%	Government: 5.38%
Global Issues: 9.04%	Economy: 6.86%	Global Issues: 5.35%
COVID-19 Europe: 6.47%	Government: 6.64%	Measures: 5.27%
Measures: 3.97%	Global Issues: 4.73%	COVID-19 Europe: 4.94%

5 Emotional and Discourse Analysis of Users' Language

To gain better insightswe computed two emotional prior scores for the messages by means of a lexicon look up approach. The lexicon used for this was the Dutch version of the NRC Emotional Lexicon (Mohammad and Turney, 2013) that contains emotion association scores for 14,182 words for 10 different emotions, each represented by a score between 0 or 1. Two additional columns for the word lemma and stem have also been added, to increase the likelihood of finding a match.

The first emotional score, called **Polarity Score**, assigns a positive or negative score to every message on the basis of the scores associated to the entries in the NRC Lexicon. To make the score operational,

emotions that in the lexicon that are deemed to be negative (e.g., "Anger", "Disgust", "Fear" and "Sadness"), have their scores inverted and ambiguous ones have been excluded. To compute this score, we go over each word in every tweet and attempt to find a match in the lexicon. If an entry is found, all associated scores for the emotional categories are added up. In case multiple entries are found, the scores are again added up and an average score is returned. This score indicates the actual emotion directionality carried by a tweet, influenced by positive and negative words within it, rather than the dominant emotion label.

The second emotional score, that we called **Emotional Load**, aims at assessing the emotional weight of the messages, i.e., how many emotions a message has. We compute the users' Emotional Load by assigning a point any time that a word in a message matches an entry in the NRC lexicon. The procedure of computing the score is similar to that of the **Polarity Score**, however the lexicon is left unchanged and negative emotions are not inverted. For each found word in a tweet, the scores in the entry are added up. Similarly as before, in case multiple entries are found, the total added scores are averaged out.

Emojis are also included in the score calculation of both measures. A different lexicon was used for this (Novak et al., 2015), with the lexicon containing sentiment scores for each emoji. For the Emotional Load, each detected emoji simply adds 1 to the score, while for the Polarity Score, the actual emotion score of the corresponding emoji is used.

We applied these measures to specific groups of users corresponding to socio-demographic categories of interest. In particular, we wanted to investigate if there are differences among five major societal actors: governmental and public health organizations, news media (i.e., TVs, radios, newspapers), politicians, the general public, and promoters/supporters of conspiracy theories. These groups have been selected as they appear to play different roles with respect to the impact of COVID-19. Governmental and public health organizations are in charge of everyday management and decisions which impact on the population; news media are responsible for setting the tone and the narratives about COVID-19; politicians represent the connecting elements between the government and the different groups of interest they represent; the general population is somehow a passive actor who undergoes the decisions of the government; conspiracy theory activists promote "alternative facts" on COVID-19, are responsible for spreading misinformation, and build narratives that directly oppose those of governments and news media.

The identification and clustering of the users in socio-demographic categories has been conducted as follows: first, we have distinguished the authors of the tweets into verified and non-verified users. User names are cleaned to only include alphabetic characters. We then conducted a semi-automatic automatic entity linking by associating the screen names to Wikipedia entry. Ideally the screen name would return a single entry with the needed information in the form of a summary page. Each retrieved Wikipedia summary is then further matched with a set of pre-defined keywords characterizing every target group extracted from the user description. In case a screen name returned multiple Wikipedia entries, a summary is taken from each and the matching process repeated. After this first pass, we manually checked that the automatically assigned labels were correct. A limitation of this strategy is that the lack of Wikipedia entry places the target users in the general population. As a strategy to compensate this potential generalization of users in the general public demographic, we further distinguished in the general population all verified accounts from the rest. Verified accounts in Twitter are used to signal the authenticity of the accounts of users of public interest. The conspiracy theory activists are found by searching for users whose messages contained the "5G" keyword.

5.1 Emotional Content Findings

Table 4 reports the percentages of tweets per user category for which, respectively, no matches, a single match, or multiple matches were found in the NRC Dutch version of the lexicon. [9] Quite surprisingly, we observe that for all user categories, between 88% to 96% of the tweets have one or more corresponding entries in the lexicon, besides some translation errors.

Table 5 illustrates the results of the Polarity Score and the Emotional Load for each category of users.

---

[9]For the General Population (Non-Verified) a random sample of 40,000 messages was selected for the scores

Table 4: Lexicon coverage per user category.

User Category	0 matches (%)	1 match (%)	> 1 matches (%)
Politicians	0.66	2.37	96.97
News Media	2.85	8.49	88.66
Government/Health Organizations	0.61	2.94	96.45
General Population (Verified)	2.5	7.3	90.20
General Population (Non-Verified)	2.14	6.24	91.62
Conspiracy Theory	1.71	5.16	93.13

Table 5: Polarity Score and Emotional Load scores for each of the selected category of users.

User Category	Number of Users	Number of Tweets	Polarity Score	Emotional Load
Politicians	150	4826	0.7387	9.0370
News Media	142	26072	-0.3051	5.7403
Government/Health Organizations	62	3101	0.9848	8.3099
General Population (Verified)	1082	43037	0.2113	6.4115
General Population (Non-Verified)	226146	2366469	0.1150	7.3628
Conspiracy Theory	8195	14291	-0.2446	7.3212

For completeness, we also report the number of identified users per category and their associated tweets.

The Polarity Score clearly indicates that there are differences in the directionality of the expressed emotions in the messages across the various categories. We can observe that Politicians and Governmental/Health Organizations express very high positive emotions with scores very near to the maximum level, i.e., 1. This clearly marks a distinctive element for both types of users: they are the main actors in the management and containment of the virus among the population. Their communication is oriented to express positive messages showing leadership, unity, and solutions to the problem. On the opposite side, we find the News Media and the Conspiracy Theory groups. In both cases, negative polarity scores are expected, although for different reasons. News Media in the time period we have considered were focused on reporting the number of new infected cases, number of death, people in intensive care units, job losses, and, more in general, the impact of the lockdown on the Dutch society. Although there were also positive messages, the majority of the news events that were covered is associated with negative emotions, boosting an already existing trend of News Media on bad news (Thompson et al., 2017). One of the main activities of conspiracy theories is counteract the official narratives. In this specific case, their messages have an opposite tone with respect to the Politicians and Governmental/Health Organizations. Indeed, a recurrent theme is accusing both these categories of lying and having creating this crisis to increase their control on the population. The remaining two categories, expression of the population at large, are associated with a positive Polarity Score. For both groups the scores are the lowest when compared to the other user groups, and with the verified account being higher than the non-verified ones (0.2113 *vs.* 0.1150, respectively). It thus appears that verified accounts have contributed to promote a positive attitude in the time of the crisis, in line with the communication of other public figures such as Politicians and Government/Health Organizations. Such a positive attitude is also mirrored, although with lesser intensity, by the population at large (i.e., the non-verified accounts).

Table 6 reports the distribution of the polarity per class (i.e., scores are not summed) and the average polarity values.

Table 6: Polarity Scores: percentages of tweets per polarity class and average polarity per class for each of the user categories.

User Category	% Positive	Average score positive	% Negative	Average score negative
Politicians	50.27	3.54	30.33	-3.44
News Media	33.38	2.81	40.49	-3.07
Government/Health Organizations	52.11	3.70	28.64	-3.30
General Population (Verified)	41.01	3.14	34.17	-3.15
General Population (Non-Verified)	40.72	3.30	34.89	-3.52
Conspiracy Theory	37.26	3.16	39.54	-3.60

We observe that the averages for both the positive and negative Polarity Scores are roughly similar across all user categories, with the News Media being a slight exception with the lowest positive average

of 2.81. The lower Polarity Scores observed in Table 5 is due to the distribution of the messages in the positive and negative classes, indicating that contrasting emotions are at play and tends to compensate in our computation of the Polarity Score.

The only analysis to which we make a tentative comparison with our results is EmoItaly[10], a study that quantified emotions in Italy from January to May 2020. EmoItaly uses a similar approach to automatically identify the emotions, i.e., dictionary lookup using the Italian version of the NRC lexicon. However, it differs from our method because it computes the global "amount" for each emotion label (i.e., a weight for each emotion, rather than a dominant emotion). Besides this difference, on the basis of the reported results, it appears that the general population in Italy had a much more negative attitude than in the Netherlands during the same period.

When focusing on the Emotional Load, quite surprisingly Politicians and Governmental/Health Organizations are the groups of users that tend to have high levels of emotions in their messages. On the other hand, News Media accounts are least emotionally loaded. However, this low level of emotions, which may reflect specific lexical choices in the framing of news, is not mirrored in low levels of Polarity Scores. In contrast, the remainder of the user groups have comparable scores. However, the connection between Emotional Load and Polarity Score can be used to get additional insights on the behavior of these category of users. Given the way the Polarity Score is computed, these low values (all ranging between 0.24 and 0.10) should also be read as cues of conflicting emotions.

To better assess the differences across the categories we ran two statistical tests. First, we used a Kruskal-Wallis test across all groups. The results indicates that both for the Polarity Score (Chi square = 3047, DF = 5, p-value$<$0.05) and the Emotional Load (Chi square = 3047, DF = 5, p-value$<$0.05), the differences are statistically significant. Further we ran a pairwise comparison using a Wilcoxon rank sum test. The results show that both for the Polarity Score and the Emotional Load, the differences across the groups are significant (p-value$<$0.05), with the exception of the pair General Population (Non-Verified) and Conspiracy Theory (p-value$>$0.05).

5.2 Top Emotions and Topics Across User Groups

We further investigated two additional aspects connected with the selected groups, namely the dominant emotion and the associated topics, and their evolution in the selected time period. The emotion labels correspond to prior values, obtained by means of a dictionary look-up approach using the NRC Dutch Lexicon. Results are illustrated in Table 7.

As far as the topics are concerned, the patterns we have already observed in Table 3 is actually mirrored by the users. In February, COVID-19 is still perceived as something that is not affecting the Netherlands. Politicians mention COVID-19 as something that is related to China but are still concerned with other, more generic, issues (e.g., Politics, News Outlets, and Global Issues). On the other hand, COVID-19 is already quite central in the discussions of all the other groups. News Media are reporting potential early cases of COVID-19 and are initiating a discussion on the economic impact of the disease. The General Population (Non-Verified) discusses the spread of the disease across Europe and also about the measures being adopted to limit the spreading. This latter aspect is also present in the Conspiracy Theory group, although it represents a larger number of messages than in all other groups (i.e., 7.37%), showing a bigger sensitivity to the topic. Governmental and Health Organizations both stress that COVID-19 is a Chinese issue. However, we can observe that a portion of the messages are also dedicated to the discussion of measures against COVID-19 and its impact in the country. After the onset of the outbreak, March sees big changes in all groups. References to China have basically disappeared from the public conversation. The focus now is on the Netherlands and in Europe. With the exception of Politicians, all other groups are mainly concerned about the number of daily infections. Governmental/Health Organizations also appear to be mainly concerned about making the population aware of the measures to control the spreading (i.e., wash your hand frequently and remain at home as much as possible). Politicians, on the other hand, are mainly concerned with the economic impact of COVID-19. April shows a further change in

---

[10]http://corpora.ficlit.unibo.it/EmoItaly/

Table 7: Top 5 dominant topics and emotions for each user category.

User Category	February Topics		February Emotions		March Topics		March Emotions		April Topics		April Emotions	
Politicians	COVID-19 China	49.3%	Trust	25.35%	Economy	14.93%	Trust	30.22%	Economy	22.07%	Trust	29.44%
	Early COVID-19	18.31%	Fear	18.31%	COVID-19 NL	14.05%	Sadness	19.86%	Government	7.32%	Anticipation	20.17%
	News Outlets	5.63%	Anticipation	12.68%	Infections	9.43%	Anticipation	17.05%	Global Issues	6.21%	Sadness	19.01%
	Politics	4.23%	Disgust	11.27%	COVID-19 EU	5.27%	Fear	11.16%	COVID-19 EU	5.75%	Anger	11.87%
	Global Issues	4.23%	Sadness	9.86%	Government	5%	Anger	10.93%	Measures	5.7%	Fear	10.15%
News Media	COVID-19 China	58.6%	Sadness	19.14%	Infections	26.75%	Sadness	21.56%	Economy	20.62%	Sadness	24.10%
	Early COVID-19	11.72%	Anticipation	16.99%	Economy	13.68%	Anticipation	17.60%	Infections	7.63%	Anticipation	18.67%
	Global Issues	7.42%	Fear	15.16%	COVID-19 NL	6.92%	Trust	15.32%	Global Issues	6.69%	Trust	14.72%
	COVID-19 EU	5.05%	Disgust	15.05%	Hospitals	5.44%	Fear	12.26%	Measures	6.33%	Anger	13.4%
	Economy	3.76%	Anger	12.04%	Press Conference	4.59%	Anger	11.99%	Government	5.03%	Fear	12.33%
Governmental/Health Organizations	COVID-19 China	60.98%	Trust	21.95%	Infections	22.83%	Trust	28.64%	Economy	16.87%	Anticipation	18.67%
	Early COVID-19	16.1%	Anticipation	18.54%	COVID-19 NL	10.83%	Anticipation	23.16%	Measures	9.68%	Trust	24.36%
	Global Issues	11.22%	Anger	16.10%	Measures	9.46%	Sadness	16.50%	Life Online	5.94%	Sadness	17.83%
	Measures	2.93%	Fear	14.63%	Economy	9%	Fear	11.48%	Global Issues	5.8%	Anger	11.74%
	COVID-19 NL	2.44%	Sadness	12.68%	Government	7.18%	Anger	8.15%	Government	5.28%	Fear	9.02%
General Population (Verified)	COVID-19 China	36.47%	Sadness	17.35%	Infections	16.86%	Sadness	19.27%	Economy	16.74%	Sadness	21.49%
	Early COVID-19	18.36%	Anticipation	17.22%	COVID-19 NL	10.12%	Trust	19.25%	Measures	7.46%	Anticipation	19.98%
	Global Issues	11.13%	Fear	15.74%	Economy	9.44%	Anticipation	18.59%	Global Issues	6.76%	Trust	18.33%
	COVID-19 EU	7.72%	Disgust	12.87%	Government	7.01%	Anger	11.96%	Government	5.22%	Anger	12.61%
	Sports	4.39%	Trust	12.15%	Measures	5.2%	Fear	11.87%	COVID-19 EU	4.69%	Fear	11.1%
General Population (Non-Verified)	COVID-19 China	30.58%	Anticipation	17.25%	COVID-19 NL	15.06%	Trust	20.00%	Economy	11.48%	Trust	20.19%
	Early COVID-19	20.55%	Trust	16.59%	Infections	9.71%	Anticipation	18.31%	Government	5.38%	Anticipation	19.35%
	Global Issues	9.02%	Fear	15.32%	Economy	6.84%	Sadness	16.79%	Global Issues	5.32%	Sadness	18.53%
	COVID-19 EU	6.48%	Sadness	14.23%	Government	6.64%	Fear	12.1%	Measures	5.24%	Anger	11.92%
	Measures	4.07%	Anger	11.47%	WFH	4.53%	Anger	11.65%	COVID-19 EU	4.95%	Fear	11.44%
Conspiracy Theory	COVID-19 China	37.89%	Sadness	16.84%	Infections	14.77%	Trust	18.28%	Conspiracies	28.9%	Trust	21.01%
	Early COVID-19	17.37%	Anger	14.74%	COVID-19 NL	13.19%	Sadness	18.17%	Economy	6.07%	Sadness	19.00%
	COVID-19 EU	7.37%	Fear	14.74%	Economy	7.5%	Anticipation	17.74%	Environment	4.43%	Anticipation	16.91%
	Global Issues	7.37%	Anticipation	14.74%	Government	5.29%	Anger	13.3%	Global Issues	3.83%	Anger	14.75%
	Measures	7.37%	Trust	14.21%	Conspiracies	4.71%	Fear	12.27%	Measures	3.65%	Fear	12.29%

the narratives: the economic situation becomes the main issue of every group, with the exception of the Conspiracy Theory group, where it makes up only 6.07% of the messages. At the same time, it seems that information concerning the daily count of infections is not of interest anymore. The only group that basically keeps discussing these counts is the News Media.

When focusing on the emotions, a few patterns can be observed. The dominant emotion for the Politicians and Government/Health Organizations is Trust. This is actually a constant with minimal variations during the three month period. It is also interesting to observe, especially for the Politicians, how Fear has been continuously losing prominence, leaving room for Sadness, expressing sympathy for the deaths and infected, and Anticipation, an emotion expressing ane expectation of predictable future events. Sadness is, on the contrary, the prevalent emotion of the News Media users. When it comes to the General Population, the differences between the verified and non-verified accounts become minimal in March and are maintained in April. There are slightly variations in the order of the top three dominant emotions, but they remain unchanged: Trust, Sadness, and Anticipation. Finally, we conclude our analysis looking at the Conspiracy Theory group. Interestingly, the most dominant emotion for this group in March and April is Trust, a positive emotion. However, Table 5 indicates that the associated Polarity Scores for this user category are actually negative. As a sanity check and as a way to interpret this result, we manually checked all messages whose dominant emotion is Trust across all the categories of users. Our hypothesis is that for the Conspiracy Theory group, Trust has a negative reading corresponding to Mistrust. To verify our hypothesis, we calculated the number of messages containing a negation.[11]. It turns out that while for all other user categories, negations of messages labeled with Trust are 28.6% on average, for the Conspiracy Theory group the percentage jumps to 40.1%. This supports to our hypothesis that the Trust results is actually 'Mistrust' expressed as negated Trust in the messages of this group.

6 Conclusion

We examined Dutch language Twitter responses to the COVID-19 pandemic from February until April of 2020. The overall public response appears to be substantially positive, mainly expressing trust, although differences across groups of users are present. The results demonstrate the effectiveness and necessity of monitoring social media platforms in order to gauge a nation's response to such large scale events. Such findings were already documented in previous virus outbreaks, such as the MERS outbreak in 2015 (Shin

---

[11]We have used the following Dutch negation words: "niet" [not], "nooit"[never], "nimmer" [never], "nergens" [nowhere], "niemand"[none], "niks" [nothing], "geen" [no]

et al., 2016).

The trend of the most discussed topics changes from talking about COVID-19 in Europe, to focusing more on the economy and adjustments of living styles, such as working from home (**RQ1**). While different user groups mostly appear to discuss the same major issues (e.g., the development of the pandemic in the country, the measures adopted by the government, and the impact on the economy), there are differences in their priorities and the specific topics they emphasize. These differences seem to reflect the interests, the role in society, and the stance of the users during the COVID-19 pandemic (**RQ2**). Emotional differences were also observed between the user groups, both in the overall emotional content as well as their dominant emotions. The dominant emotions, like the topics, reflect the roles of the groups, with Politicians and Government and Health Organization inspiring trust, TV/Radio News outlining the sadness of the situation, the general public being worried but also feeling assured by the "people in charge", and some indication of mistrust by the users concerned with the conspiracy theories (**RQ3**).

Future work will focus on three directions: first, we want to investigate the topics and emotional reactions of more fine-grained categories of users by targeting different professions (e.g., teachers, doctors, nurses, musicians, entertainers, among others); second, we want to investigate if the emotional response is different across geographic areas of the country that have been affected differently by the COVID-19 outbreak. For instance, the Noord Brabant province was one of the major outbreak areas, while the Groningen province was minimally affected. We thus expect that messages from Noord Brabant should be characterized by more negative emotions than those from the Groningen area. Finally, in future research it would be useful to run similar studies on datasets in other European languages, in order to allow a comparison of how different national situations and governmental responses impacted public perception of the early stages of the pandemic.

Acknowledgments

References

Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. 2011. Analyzing user modeling on twitter for personalized news recommendations. In *international conference on user modeling, adaptation, and personalization*, pages 1–12. Springer.

Mohamed Bakillah, Ren-Yu Li, and Steve HL Liang. 2015. Geo-located community detection in twitter with enhanced fast-greedy optimization of modularity: the case study of typhoon haiyan. *International Journal of Geographical Information Science*, 29(2):258–279.

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2020. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *arXiv preprint arXiv:2004.03974*.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Su Lin Blodgett, Lisa Green, and Brendan O'Connor. 2016. Demographic dialectal variation in social media: A case study of African-American English. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1119–1130, Austin, Texas, November. Association for Computational Linguistics.

Antal van den Bosch, Bertjan Busser, Sander Canisius, and Walter Daelemans. 2007. An efficient memory-based morphosyntactic tagger and parser for dutch. *LOT Occasional Series*, 7:191–206.

Gosse Bouma. 2015. N-gram frequencies for dutch twitter data. *Computational Linguistics in the Netherlands Journal*, 5:25–36.

Nan Cao, Lu Lu, Yu-Ru Lin, Fei Wang, and Zhen Wen. 2015. Socialhelix: visual analysis of sentiment divergence in social media. *Journal of visualization*, 18(2):221–235.

Tommaso Caselli and Valerio Basile. 2020. 40twene_nl version 1.0. DOI: 10.17605/osf.io/pfnur.

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Covid-19: The first public coronavirus twitter dataset. *arXiv preprint arXiv:2003.07372*.

Xueqi Cheng, Xiaohui Yan, Yanyan Lan, and Jiafeng Guo. 2014. Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, 26(1-1).

Jaegul Choo, Changhyun Lee, Chandan K Reddy, and Haesun Park. 2015. Weakly supervised nonnegative matrix factorization for user-driven clustering. *Data Mining and Knowledge Discovery*, 29(6):1598–1621.

David Darmon, Elisa Omodei, and Joshua Garland. 2014. Followers are not enough: A question-oriented approach to community detection in online social networks. *arXiv preprint arXiv:1404.0300*.

Ashlynn R Daughton and Michael J Paul. 2019. Identifying protective health behaviors on twitter: observational study of travel advisories and zika virus. *Journal of medical Internet research*, 21(5):e13090.

Luca Dini and André Bittar. 2016. Emotion analysis on Twitter: The hidden challenge. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3953–3958, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Matthias Hagen, Martin Potthast, Michel Büchner, and Benno Stein. 2015. Webis: An ensemble for twitter sentiment detection. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 582–589.

Sardar Hamidian and Mona Diab. 2016. Rumor identification and belief investigation on Twitter. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 3–8, San Diego, California, June. Association for Computational Linguistics.

Paul Hitlin and Kenneth Olmstead. 2018. The Science People see on Social Media. Pew Research Center, March 21, 2018.

Liangjie Hong and Brian D Davison. 2010. Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics*, pages 80–88.

Mowafa Househ. 2016. Communicating ebola through social media and electronic news media outlets: A cross-sectional study. *Health informatics journal*, 22(3):470–478.

Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65.

Hannah Kim, Jaegul Choo, Jingu Kim, Chandan K Reddy, and Haesun Park. 2015. Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 567–576.

Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. 2008. A few chirps about twitter. In *Proceedings of the first workshop on Online social networks*, pages 19–24.

KP Krishna Kumar and G Geethakumari. 2014. Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, 4(1):1–22.

Nicolas LaLone, Andrea Tapia, Christopher Zobel, Cornelia Caraega, Venkata Kishore Neppalli, and Shane Halse. 2017. Embracing human noise as resilience indicator: twitter as power grid correlate. *Sustainable and Resilient Infrastructure*, 2(4):169–178.

Quanzhi Li, Qiong Zhang, Luo Si, and Yingchi Liu. 2019. Rumor detection on social media: Datasets, methods and opportunities. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 66–75, Hong Kong, China, November. Association for Computational Linguistics.

Katerina Eva Matsa and Elisa Shearer. 2018. News use across social media platforms 2018— pew research center. *Journalism and Media*.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *National Research Council, Canada*, 2.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of emojis. *PloS one*, 10(12).

Catherine Ordun, Sanjay Purushotham, and Edward Raff. 2020. Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs. *arXiv preprint arXiv:2005.03082*.

Barbara Plank and Dirk Hovy. 2015. Personality traits on Twitter—or—How to get 1,500 personality tests in a week. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 92–98, Lisboa, Portugal, September. Association for Computational Linguistics.

Francisco Rangel, Paolo Rosso, Martin Potthast, and Benno Stein. 2017. Overview of the 5th author profiling task at pan 2017: Gender and language variety identification in twitter. *Working notes papers of the CLEF*, pages 1613–0073.

Alan Ritter, Colin Cherry, and Bill Dolan. 2010. Unsupervised modeling of Twitter conversations. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 172–180, Los Angeles, California, June. Association for Computational Linguistics.

Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2019. Calls to action on social media: Detection, social impact, and censorship potential. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 36–44, Hong Kong, China, November. Association for Computational Linguistics.

Erik Tjong Kim Sang. 2011. Het gebruik van twitter voor taalkundig onderzoek. *TABU: Bulletin voor Taalwetenschap*, 39(1/2):62–72.

Efsun Sarioglu, Kabir Yadav, and Hyeong-Ah Choi. 2013. Topic modeling based classification of clinical reports. In *51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop*, pages 67–73, Sofia, Bulgaria, August. Association for Computational Linguistics.

Leonard Schild, Chen Ling, Jeremy Blackburn, Gianluca Stringhini, Yang Zhang, and Savvas Zannettou. 2020. "go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of covid-19. *arXiv preprint arXiv:2004.04046*.

Christof Schöch. 2017. Topic modeling genre: An exploration of french classical and enlightenment drama. *DHQ: Digital Humanities Quarterly*, 11(2).

Roy Schwartz, Oren Tsur, Ari Rappoport, and Moshe Koppel. 2013. Authorship attribution of micro-messages. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1880–1891.

Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, Aastha Dua, and Yan Liu. 2020. Coronavirus on social media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*.

Tian Shi, Kyeongpil Kang, Jaegul Choo, and Chandan K Reddy. 2018. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In *Proceedings of the 2018 World Wide Web Conference*, pages 1105–1114.

Soo-Yong Shin, Dong-Woo Seo, Jisun An, Haewoon Kwak, Sung-Han Kim, Jin Gwack, and Min-Woo Jo. 2016. High correlation of middle east respiratory syndrome spread with google search and twitter trends in korea. *Scientific reports*, 6:32920.

Luke Sloan, Jeffrey Morgan, Pete Burnap, and Matthew Williams. 2015. Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. *PloS one*, 10(3):e0115545.

Anthony Stefanidis, Emily Vraga, Georgios Lamprianidis, Jacek Radzikowski, Paul L Delamater, Kathryn H Jacobsen, Dieter Pfoser, Arie Croitoru, and Andrew Crooks. 2017. Zika in twitter: temporal variations of locations, actors, and concepts. *JMIR public health and surveillance*, 3(2):e22.

Paul Thompson, Raheel Nawaz, John McNaught, and Sophia Ananiadou. 2017. Enriching news events with meta-knowledge information. *Language Resources and Evaluation*, 51(2):409–438.

Tze-I Yang, Andrew Torget, and Rada Mihalcea. 2011. Topic modeling on historical newspapers. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 96–104.