# Appendix

**Robustness check using sentiment analysis**

As previously stated, there are two main approaches for measuring engagement based on OSNs and text-as-data: one that is based on ternary ratios, and another that is based on the sentiment of the tweets.

Sentiment analysis uses specialized dictionaries that contain terms that are either positive, negative, or neutral in nature that is referred to as "sentiment analysis." In order to determine the frequency of these terms over a wide variety of linguistic parameters in the given text or hashtags, researchers employ these dictionaries.

To conduct sentiment analysis on the collected data, we employ the tidytext package in the R programming language, which is well-suited to the emotive nature of social media platforms, particularly X. Analyzing a text containing both positive and negative sentiment necessitates a reader's ability to discern the presence and strength of each emotion.

Various methodologies and dictionaries exist for assessing the sentiment or emotions conveyed in a text. The tidytext package provides access to multiple sentiment lexicons, including AFINN, Bing, and NRC. Each of these lexicons is based on unigrams, single words that convey emotions. The lexicons contain a vast number of English terms, assigning them values based on the emotions they express, such as positive or negative sentiment, joy, anger, sorrow, and others.

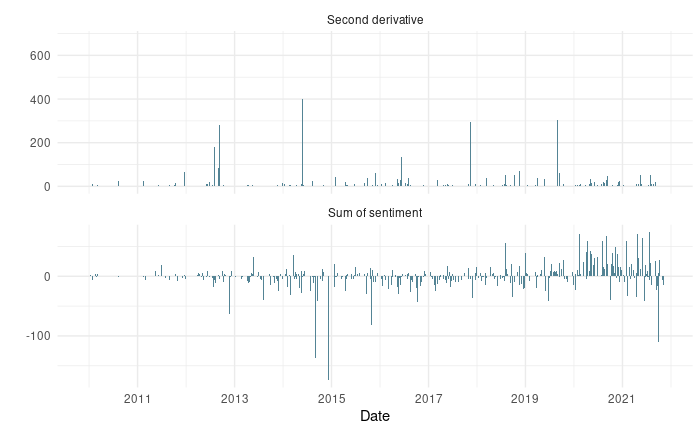
The NRC lexicon categorizes words using a binary classification system ("yes/no") into categories such as positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The Bing lexicon (Bing, 2012, 2015) also employs a binary classification system to divide words into positive and negative categories, with the former being more prevalent. We chose to utilize the AFINN lexicon for the sentiment analysis in our study, as it provides an overall score based on the volume of tweets received daily. The AFINN lexicon assigns a score to each word, ranging from -5 (strongly negative) to 5 (strongly positive), allowing for a more nuanced understanding of sentiment compared to binary classification systems.

To calculate the daily sentiment score, we employed two methods: the total sum and the mean of the scores. In the first method, we aggregated the AFINN scores of all words within the tweets for each day. This approach provided us with a daily total sentiment score, which represented the overall emotional tone of the tweets on that particular day. A positive score indicated a predominantly positive sentiment, whereas a negative score suggested a predominantly negative sentiment.

In the second method, we calculated the mean of the AFINN scores for each day, which involved summing up the scores and then dividing by the total number of words in the tweets for that day. This approach yielded an average sentiment score, enabling us to gauge the general sentiment per word. By comparing the mean scores across different days, we could assess the variations in sentiment over time and identify potential trends or patterns in the public's response to WHO’s communications on X.

Figure 11 depicts the positive second derivative of retweets in order to be able to visualize accelerations of conversations per day. We have also created a daily corpus aggregating all the texts from all the tweets posted in a day. We have run our sentiment analysis on each daily corpus. It allows us to easily visualize correlation levels between an acceleration of conversations and whether it is related to negative or positive sentiments.

**Figure 11.** Positive second derivative of retweets and the sum of sentiment associated to the tweet per day

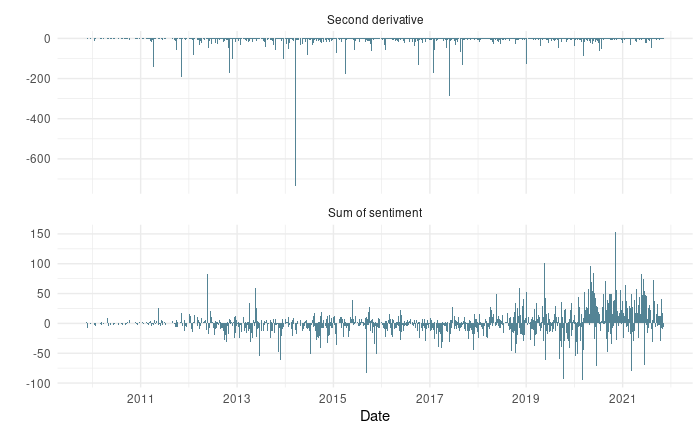


In Figure 11, an intriguing observation is that despite a clearly concerted effort by WHO to have more positively toned tweets during the COVID-19 period, the acceleration of the conversation around these tweets is minimal. This suggests that increased positivity by the IO has little impact (either positive or negative) on its popular legitimacy. Understanding these dynamics can help inform the organization’s communication strategies and contribute to the assessment of its legitimacy based on audience engagement on social media platforms.

Figure 12 presents the negative second derivative of retweets alongside the sum of sentiment associated with the daily corpus of text, which is created based on all the tweets of the day. This representation enables us to examine the relationship between the deceleration of retweets and the overall sentiment of the corresponding conversations.

By comparing these two metrics, we can gain insights into how the dynamics of audience engagement and the sentiment expressed in the conversations change over time. This analysis can help us understand whether negative sentiment is a driving factor in the deceleration of retweets or if other factors are at play. Such insights contribute to our assessment of the organization’s legitimacy based on public engagement and perception on social media platforms.

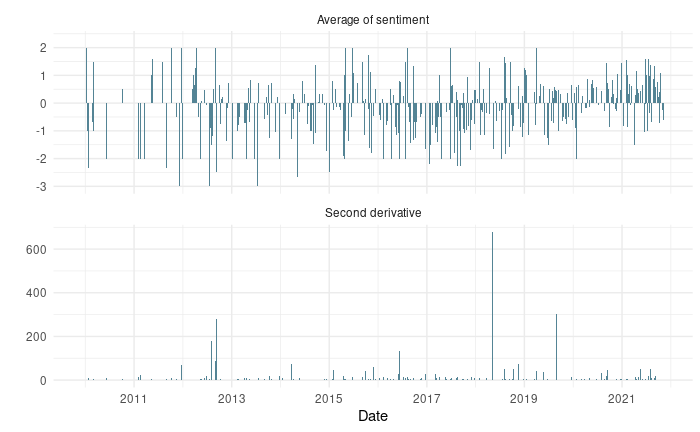
**Figure 12.** Negative second derivative of retweets and the sum of sentiment associated to the tweet per day



Like the positive sentiment analysis, there seems to be little correlation between the sum of sentiments and the negative second derivatives (i.e., the deceleration of a conversation). One possible explanation is that the overall quantity of tweeting has an affect on the level of public engagement. Not only is WHO tweeting more positively in the final two years in the sample, but also at a quantity of over 10 times that of the first three years (i.e., from 2008-2010). If tweeting is more frequent, then each tweet may carry less weight in the eyes of X users, thus inciting less variance in the reactions (meaning not extreme accelerations nor decelerations).

In the subsequent figures, Figures 13 and 14, we shift our focus to analyzing the mean sentiment per day instead of the sum of sentiments. By examining the mean sentiment, we can account for variations in daily conversation volume and better understand the overall sentiment trend on a per-tweet basis. This approach allows us to capture more accurately the general mood of the audience and further investigate the relationship between sentiment and the pace of conversations. Ultimately, such an analysis can contribute to a more comprehensive understanding of the factors affecting the organization’s legitimacy and public perception on social media platforms.

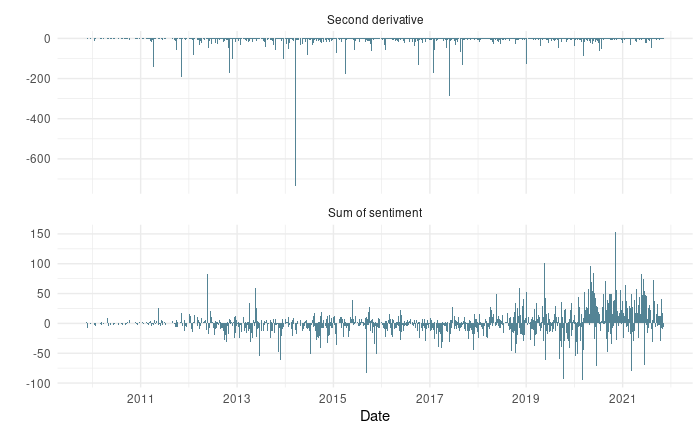
**Figure 13.** Positive second derivative of retweets and the mean of sentiment associated to the tweet per day



The pattern is similar to that outlined in Figures 11 and 12. There is a notably higher mean sentiment in the final two years of the sample (especially after the first half of 2020), but the impact this has on the positive second derivatives appears to be small in magnitude, and doesn’t differ greatly from previous years (despite those years having more negative sentiments on average).

Understanding this dynamic can provide valuable insights for organizations looking to enhance their legitimacy and public perception. By focusing on communicating messages that evoke positive sentiments, organizations can potentially encourage more active engagement, promote constructive conversations, and contribute to a favorable public image.

**Figure 14.** Negative second derivative of retweets and the mean of sentiment associated to the tweet per day



In conclusion, our analysis reveals that changes in the sum of sentiments of conversations seems to have little impact on the acceleration or deceleration of conversations. While it is beyond the scope of this paper, one possibility to consider is that the overall quantity of tweeting has a greater influence on the acceleration or deceleration of engagements than the sentiment of the original tweets. The strong increase in tweeting would also coincide with the changes we note in the retweet-to-reply ratio starting in 2020 and persisting through 2021.

These findings offer valuable insights for the research avenue proposed in this study, which focuses on using quantitative indicators to measure popular legitimacy based on text-as-data from OSNs. By further examining these dynamics, researchers may gain a more comprehensive understanding of how public sentiment and engagement patterns in OSNs can be utilized to assess the legitimacy of organizations like the World Health Organization, ultimately contributing to the development of more robust, data-driven methodologies in the realm of political science.