

## Original Paper

# Social Contagion in COVID-19 Discussions Within the Belgian Reddit Community: Statistical and Modeling Study

Tim Van Wesemael<sup>1</sup>, MSc; Luis Enrique Correa Rocha<sup>2,3</sup>, PhD; Tijs W Alleman<sup>4,5</sup>, PhD; Jan M Baetens<sup>1</sup>, PhD

<sup>1</sup>BionamiX, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University, Ghent, Flanders, Belgium

<sup>2</sup>Department of Economics, Faculty of Economics and Business Administration, Ghent University, Ghent, Flanders, Belgium

<sup>3</sup>Department of Physics, Faculty of Science, Ghent University, Ghent, Flanders, Belgium

<sup>4</sup>Department of Public and Ecosystem Health, College of Veterinary Medicine, Cornell University, Ithaca, NY, United States

<sup>5</sup>Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

## Corresponding Author:

Tim Van Wesemael, MSc

BionamiX, Department of Data Analysis and Mathematical Modelling

Faculty of Bioscience Engineering, Ghent University

Coupure Links 653

Ghent, Flanders 9000

Belgium

Phone: 32 9 264 59 32

Email: [tim.vanwesemael@ugent.be](mailto:tim.vanwesemael@ugent.be)

## Abstract

**Background:** Understanding how sentiment toward COVID-19 mitigation measures evolves on social networks can help to inform infectious disease models and policymakers. Even though numerous studies have described social media interactions during the pandemic, few have modeled the underlying dynamics of sentiment contagion and polarization.

**Objective:** This study aimed to investigate topic emergence and sentiment evolution in COVID-19 mitigation discussions on r/Belgium, focusing on (1) whether discussion topics exhibited social contagion, (2) whether expressed sentiment displayed homophily, and (3) how this homophily formation can be captured by a mechanistic model.

**Methods:** We classified posts created on r/Belgium between January 1, 2020, and June 30, 2022, into lockdowns, masks, and vaccination, using a pretrained bidirectional encoder representations from transformers (BERT) topic model, and assigned English posts a sentiment, using a robustly optimized BERT pretraining approach (RoBERTa)-based sentiment classifier. We then examined temporal patterns of post volume and tested for social contagion in topic initiation. Sentiment homophily was quantified by comparing observed comment-parent sentiment pairs to null distributions. The novel smooth latent-expressed bounded confidence (SLEBC) model dynamically captured sentiment evolution, distinguishing between latent sentiment trajectories and noisy expressed sentiment. We tested the model against 2 alternatives, one with a linear update and one without the latent state, using the Watanabe-Akaike information criterion.

**Results:** Analysis of 655,642 posts made by 28,559 users revealed that post volume was associated with external events such as policy announcements and media reports. There was no evidence of within-Reddit social contagion in topic initiation. However, sentiment exhibited significant homophily, with comment sentiment correlating with parent comment sentiment. The SLEBC model reproduced observed sentiment patterns (Watanabe-Akaike information criterion: -28.5 to -18.4 across topics), outperforming both alternatives (-21.1 to -17.4 for the linear model and 6.8 to 692 for the one without the latent state). It slightly underestimated sentiment homophily but still outperformed the alternatives in this regard. In the SLEBC model, expressed sentiment adapts more strongly to the immediate parent comment than the user's latent state updates based on their interaction history (proportions of users showing this pattern: 0.74, 0.70, and 0.51 for lockdowns, masks, and vaccination).

**Conclusions:** Discussion topics on r/Belgium are not associated with social contagion within the platform, but sentiment dynamics are shaped by within-thread interactions. The SLEBC model suggests that users adapt their expressed sentiment to match the post they reply to, highlighting that expressed sentiment may poorly reflect underlying latent sentiment. Infodemic models for Reddit-like platforms could benefit from incorporating external information sources for topic seeding and from using bounded confidence rather than linear contagion mechanisms for sentiment spread.

**Keywords:** COVID-19 mitigation; Reddit; topic modeling; sentiment analysis; bounded-confidence model; social contagion

## Introduction

To mitigate the burden on health care systems during a pandemic, governments may be forced to introduce mandatory measures. The COVID-19 pandemic served as a prime example, as the myriad of mitigation measures had far-reaching societal consequences [1-3]. Initially in 2020, these were mainly nonpharmaceutical interventions, such as lockdowns, contact reductions, and mask mandates [4,5]. As the pandemic progressed, vaccination campaigns were initiated near the end of 2020, and the focus gradually shifted to vaccine distribution during 2021. The success of these interventions depended not only on their epidemiological effectiveness but also on public adherence, which was influenced by individual attitudes, information exposure, and social interactions. Therefore, understanding public sentiment toward these measures is crucial for health informatics, as it can inform epidemiological models, guide public health communication strategies, and support evidence-based policymaking [6,7].

An unprecedented surge of health-related information on social media accompanied the pandemic, leading to what has been termed an “infodemic,” the rapid spread of both accurate and inaccurate information that can hinder effective public health responses [8]. This phenomenon has given rise to infodemiology, the science of monitoring and analyzing digital health information to inform public health practice [9]. Social media platforms have become essential data sources for infodemiology, enabling real-time monitoring of public health sentiment and tracking information diffusion patterns [10]. Research has demonstrated that social media discussions about health topics are influenced by multiple factors. While traditional media coverage and official announcements drive much of the online discourse [11,12], the spread of information within social networks also plays a crucial role. Studies that have established correlations between sentiment expressed online and vaccine uptake highlight the potential of social media data as a proxy for public health attitudes [13,14].

However, infodemics are complex and multifaceted, as sentiment expressed on social media can reflect both support and opposition, and discussions are shaped by misinformation spread [15], echo chambers [16], and platform-specific dynamics. Twitter (now X) has been the most extensively studied platform for COVID-19 sentiment analysis [10,17-19], including research specific to Belgium on monitoring emotions [11] and support for specific mitigation measures [20].

Reddit, with its structured discussion threads and pseudo-anonymous communities, provides a different informative environment for studying online health discussions [21, 22]. Unlike Twitter’s brief posts, Reddit enables extended, threaded conversations. Previous research on COVID-19-related subreddits has documented both temporal persistence [23]

and shifts [24] in sentiment, regional differences in public concerns [25,26], and the role of external events in driving discussion volume and emotions [12,27,28]. Studies have also shown sentiment homophily, the tendency for users to interact with others with similar sentiment [16]. These patterns can manifest in different forms: pluralism (where individual sentiment is largely independent), consensus (where users converge toward similar sentiment), or polarization (where 2 or more sentiment clusters emerge) [29]. A large-scale experiment on Facebook showed that emotional states can spread through this online social network [30]. While echo chambers have been documented on other platforms, Reddit has shown less pronounced polarization in contexts such as the 2016 US election and vaccination [16,31].

Computational models of social contagion have been applied to understand the mechanisms behind these emergent patterns of sentiment on social platforms [32]. These models distinguish between simple contagion, where a single exposure can trigger behavioral change, and complex contagion, requiring multiple exposures [33,34]. Bounded confidence models, which assume that individuals only interact with others holding sufficiently similar opinions, have been used to reproduce observed steady-state patterns of consensus and polarization on social media [29,35]. In public health, these models have been used to model, for example, the spatial diffusion of vaccine hesitancy [36], or to couple opinion formation with epidemic spread on contact networks [37]. Nevertheless, most studies of online COVID-19 discourse have focused on describing sentiment patterns rather than developing mechanistic models that can explain the underlying dynamics. Here we applied a novel dynamic bounded confidence model to COVID-19 mitigation discussions on Reddit, improving mechanistic understanding of how sentiment evolves in threaded conversations.

This study examined discussions on COVID-19 mitigation measures within the Belgian Reddit community (r/Belgium) from January 1, 2020, to June 30, 2022, focusing on three key topics: lockdowns, masks, and vaccination. Topic modeling and sentiment analysis were used to characterize discussion volume and sentiment dynamics, in pursuit of 3 goals. First, we investigated whether discussion topics exhibited social contagion: a null model for contagion testing revealed no evidence of social contagion in topic initiation. Second, we examined whether expressed sentiment displayed homophily: combining interrupted time series analysis with homophily quantification, we found statistically significant sentiment homophily in the reply structure. Third, we modeled how this homophily formed under local interactions and long-term trends: we developed a novel stochastic bounded confidence model, the smooth latent-expressed bounded confidence (SLEBC) model, that distinguishes between a user’s latent state and the observable sentiment expressed in their comments. The SLEBC model reproduced sentiment distributions and homophily better than two ablations: one

that did not distinguish between the two sentiment states, and one that used linear instead of bounded confidence updates.

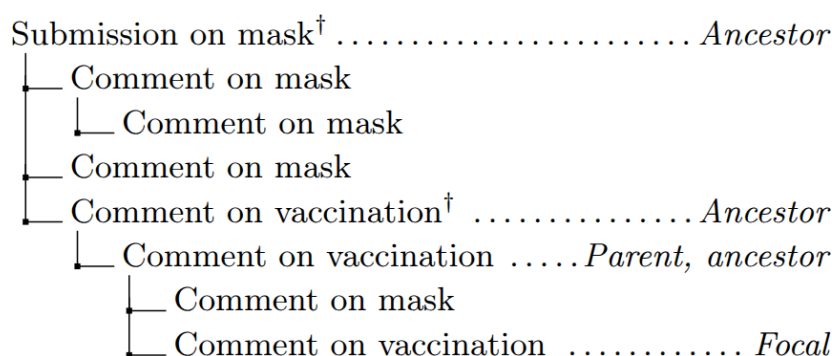
## Methods

### Data Source and Collection

Reddit is a social media platform organized into communities called subreddits, denoted by the prefix *r/*. Users operate under fixed pseudonymous usernames and typically share limited personal information. Discussions begin when a user

posts a submission consisting of a title, optionally accompanied by a hyperlink or text. Other users can comment on submissions and reply to specific comments, creating a tree-like conversation structure termed a thread ([Figure 1](#)). Throughout this work, the term post refers to both submissions and comments. Each comment's parent is the post it replies to, and its ancestors comprise the parent and the ancestors of the parent, always including a single submission at the top level. Users can upvote or downvote posts, generating a score reflecting community agreement [21]. A glossary of terms is available in [Multimedia Appendix 1](#).

**Figure 1.** Structure of a Reddit thread. The parent and ancestors of the focal comment are shown. The initiating posts for the mask and vaccination discussions are denoted by †; the others are participating.



Posts from the *r/Belgium* subreddit created between January 1, 2020, and June 30, 2022, were retrieved from the Pushshift dataset [38], a monthly-updated Reddit archive. This period encompassed the phases before, during, and after the COVID-19 outbreak in Belgium. Submissions containing

at least one of the keywords listed in [Textbox 1](#) in their title, text, or thread were initially selected. For each post, the unique identifier, creation timestamp, author, text, score, parent post, and submission title were extracted. Markdown tags, URLs, and quotes were removed from all texts.

#### Textbox 1. Keywords used to select submissions.

- corona
- virus
- covid
- mask
- masque
- lockdown
- confin
- quarant
- curfew
- avondklok
- couvre-feu
- couvre feu
- vaccin
- vax
- jab
- booster
- prik
- pique
- piquê
- pcr
- plf
- locator form
- cst
- safe ticket

Topic classification was performed using a multilingual bidirectional encoder representations from transformers (mBERT) model, trained specifically for the Belgian context and capable of processing Dutch, French, German, and English text [20]. Although the model identified eight specific measures plus other measures and not applicable categories, analysis focused on three primary measures: lockdowns, masks, and vaccination. These topics were selected for their sustained relevance throughout the pandemic and minimal risk of confusion with non-COVID-19 content. Submissions classified as not applicable with at least 90% of their comments similarly classified were excluded. Comments labeled not applicable or other inherited their parent's topic. For each topic, the distribution of posts per user was approximated by a power-law distribution  $P(k) \sim k^{-\gamma}$ , where  $\gamma$  was determined using maximum likelihood estimation and  $P(k)$  is the probability that a user was the author of  $k$  posts [39,40].

Language detection was performed using a robustly optimized BERT pretraining approach (RoBERTa), xlm-robertabase-language-detection [41]. For English-language posts, sentiment analysis was conducted with twitter-roberta-base-sentiment-latest (roberta-tbssl) [42], a classifier used for Reddit content [43,44]. The restriction to English avoided language-specific features in sentiment analysis, while incorporating the bulk of the posts (section *Dataset Characteristics*). The roberta-tbssl model outputs three scores (0 to 1) for positive, neutral, and negative tonality. These were combined into a single continuous sentiment value  $S_k$  per post  $k$  by subtracting the negative score from the positive score, yielding values from -1 (most negative) to 1 (most positive). This continuous representation was used in subsequent sentiment analyses (sections *Sentiment Homophily* and *Sentiment Evolution Model*).

## Temporal Analysis

The daily volume of posts related to each topic was calculated and visualized alongside key Belgian policy events, hospitalization counts, and administered vaccination doses [45,46]. A linear interrupted time series model was fitted to daily post counts, using the event dates in Table S1 in [Multimedia Appendix 2](#) as breakpoints [47]. The model assumed a linear segment between consecutive breakpoints. At each event  $j$ , both the intercept (level change  $\Delta\beta_{0,j}$ ) and the slope (trend change  $\Delta\beta_{1,j}$ ) may change, yielding a piecewise linear fit with ordinary least squares CIs. This allowed us to identify changes in posting volume associated with specific events ([Multimedia Appendix 3](#)).

To identify days with unusually negative sentiment, each post's sentiment was weighted by its score and aggregated per day and per topic. A day was marked as a significantly negative day for a topic if the median weighted sentiment on that day fell below the 0.275 quantile of the overall weighted sentiment distribution for that topic and at least 50 comments were submitted that day. These thresholds were chosen to identify 1 to 5 significantly negative days per topic ([Multimedia Appendix 3](#)).

## Topic Contagion

To assess whether users were more likely to initiate discussions on a topic after previously participating in one, each post in a thread sharing a common topic was classified as either initiating or participating. A post was classified as initiating if none of its ancestors shared the same topic; otherwise, it was participating ([Figure 1](#)). The author of an initiating post was designated the initiator, while all others were participants.

If contagion were present, exposure to a topic through participation would increase the probability of later initiating a new discussion on that topic. Hence, participations would accumulate at the start of a user's posting sequence. Conversely, if topic initiation is independent of prior participation, initiations and participations should be randomly interleaved across the sequence of a user's posts.

For each user, the number and sequence of initiations and participations per topic were recorded, along with the position in their posting sequence at which the first initiation occurred. The observed position was compared against a null model that preserves the number of participations and initiations, but treats each possible ordering as equally likely. Under this null model, the probability that the  $i$ 'th post is the first initiation has a closed-form combinatorial expression ([Multimedia Appendix 3](#)). For each user, the observed position in the sequence where the first initiation occurred was compared to the null model distribution. This yielded  $p(i)$ , the proportion of users with at least one initiation among their first  $i$  discussions for a topic, relative to the total number of users that initiated or participated in a discussion on that topic.

## Sentiment Homophily

The relationship between comment and parent sentiment was examined by comparing each comment's sentiment  $S_k$  to its parent's sentiment  $S_l$ . All such pairs  $(S_k, S_l)$  within a topic were collected into a 2D joint histogram with square bins of width  $w_H = 0.05$ , yielding an empirical joint probability distribution  $H$ .

A null model randomly paired comment and parent sentiment by sampling them independently from the observed distributions per topic. For comments, this distribution consisted of observed comment sentiment, while the parent distribution additionally included submission sentiment, making the null histogram  $\tilde{H}$  asymmetric. The difference between empirical and null histograms  $\Delta H = F_{.05}(H - \tilde{H})$ , revealed which sentiment pairs occurred more or less often than expected. The function  $F_{.05}$  sets entries to zero if their  $P$  value under the null model exceeds .05. That is,  $\Delta H_{l,m} > 0$  indicates that the sentiment pair  $(l, m)$  occurred significantly more frequently than expected under independent pairing, and  $\Delta H_{l,m} < 0$  that it occurred significantly less frequently. In the resulting histograms, pluralism appears as uniform patterns, consensus appears as positive mass near a single point on the

diagonal, and polarization appears as distinct on-diagonal clusters [29]. This 2D histogram was then summarized into a single homophily measure  $h(\Delta H)$  by weighting the value of each bin according to its distance from the diagonal  $l = m$ . Positive mass near the diagonal increases  $h$ , while off-diagonal mass decreases it. Higher  $h$  indicates stronger homophily, while negative values imply heterophily (Multimedia Appendices 3 and 4).

Two contexts served as extensions to comment-parent homophily. In each of them, we include more than one ( $n$ ) preceding comments. First, the ancestral context  $A_k^n$  comprises the  $n$  closest ancestors of comment  $k$ . Second, the user context  $U_k^n$  includes the parents of the comment  $k$  and of the  $n - 1$  other comments by the same author on the same topic. If users primarily adapted to recently viewed comments, ancestral homophily would dominate. If users had inherent sentiment and selectively interacted with similar users, user context homophily would prevail.

2D histograms were created for comment sentiment and mean context sentiment for context sizes one through five. Only comments with both ancestral and user contexts of at least size five were analyzed. Homophily values per context size were compared to null histograms  $\tilde{H}^n$  generated by pairing observed sentiment with the mean of  $n$  random observed sentiments.

## Sentiment Evolution Model

The SLEBC model was developed to capture the evolution of a user's sentiment. The model includes a latent  $u_i[t]$  and an expressed  $e_i[t]$  sentiment state for each user  $i$ . The former is an unobserved variable that accumulates sentiment influence across all interactions over time, whereas the latter quantifies

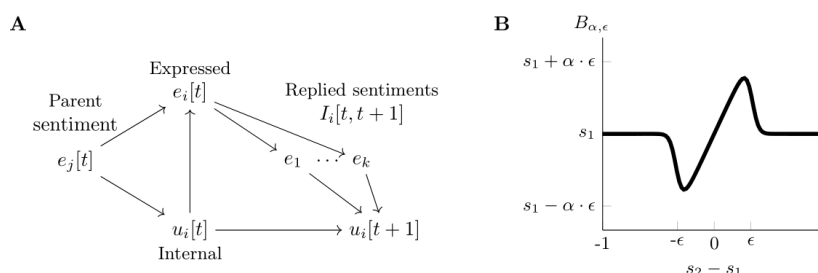
the sentiment of a specific comment and is observable. This structure is analogous to a state-space model, with  $u_i[t]$  as the latent state and  $e_i[t]$  as a parent-influenced observation. The latent state evolves based on the expressed sentiment of other users, that is, the posts that the user replies to, and the comments that the user receives. The expressed state updates according to the user's latent state and the expressed state of the immediate parent comment, allowing a local adaptation. Both states evolve stochastically according to a smooth bounded confidence kernel  $B_{\alpha, \epsilon}$ , each with its own update strength  $\alpha$  but a shared sentiment distance threshold  $\epsilon$ . Schematically, it is given by Figure 2A, and mathematically, it is given by the following:

$$e_i[t] \sim \widehat{\mathcal{N}}(B_{\alpha_e, \epsilon_i}(u_i[t-1], e_j[t]), \sigma_{e,i}) \quad (1a)$$

$$u_i[t] \sim \widehat{\mathcal{N}}\left(\bigotimes_{e_k \in I_i[t-1, t]} B_{\alpha_u, \epsilon_i}(u_i[t-1], e_k), \sigma_{u,i}\right) \quad (1b)$$

where  $t - 1$  is the time at which user  $i$  made its previous comment and  $\widehat{\mathcal{N}}(\mu, \sigma)$  is a truncated normal distribution with mean  $\mu$  and standard deviation  $\sigma$  and support  $[-1, 1]$ . In Equation 1a,  $e_j[t]$  refers to the expressed sentiment of the user  $j$  that user  $i$  is reacting to. In Equation 1b,  $I_i[t-1, t]$  is the set of sentiment of the parent comment and all replies user  $i$  received between time  $t - 1$  and  $t$ , and the operator  $\bigotimes$  indicates that the bounded confidence update is applied to each such interaction. The user-specific parameters  $\alpha_{e,i}$ ,  $\alpha_{u,i}$ ,  $\epsilon_i$ ,  $\sigma_{e,i}$ ,  $\sigma_{u,i}$  were sampled from a posterior distribution, given the likelihood of the observed expressed sentiments.

**Figure 2.** (A) Schematic representation of the smooth latent expressed bounded confidence (SLEBC, Equation 1). (B) The change in sentiment after applying  $B_{\alpha, \epsilon}$  (Equation A.8 in Multimedia Appendix 3) as a function of the sentiment difference  $s_2 - s_1$ .



In bounded confidence models, interacting agents linearly align their sentiment (with strength  $\alpha$ ), but only if their difference is smaller than a threshold  $\epsilon$ , which enables polarized states to emerge [48]. Here, a smoothed version  $B_{\alpha, \epsilon}$ , was used to ease calibration (Multimedia Appendix 3) [49]. Figure 2B illustrates the dependence of  $B_{\alpha, \epsilon}$  on the sentiment difference  $s_2 - s_1$ . Since  $\eta$  is finite, the maximum sentiment update remains smaller than  $\alpha \cdot \epsilon$ . To illustrate the advantage of the bounded confidence kernel  $B_{\alpha, \epsilon}$ , it was

compared to two alternatives. The first uses a linear kernel to update the sentiment in Equation 1 [50]. This corresponds to the bounded confidence update  $B_{\alpha, \epsilon}$  in Equation A.8 for  $\eta \rightarrow \infty$  and  $\epsilon = 2$ . Second, we ablated the internal sentiment state, so that expressed sentiment updated directly against the parent comment and received replies, without a separate latent trajectory (Equations A.9 and A.10 in Multimedia Appendix 3).

To fit the SLEBC model and its alternatives, users with at least 40 comments per topic were selected. For each user,

the sentiment of their comments and the posts they interacted with was extracted. The posterior distribution of the parameters was obtained using a Hamiltonian Monte Carlo method (Multimedia Appendix 3) [51,52].

To assess the quality of the models, the posterior distribution of predicted sentiment was scored using the Watanabe-Akaike information criterion (WAIC) [53]. A lower value is better and indicates how well the model can predict data it was not trained on (Multimedia Appendix 3). A boxplot synthesizes the homophily values given by  $h$  (Equation A.7 in Multimedia Appendix 3), showing their minimum, maximum, and 0.25, 0.5, and 0.75 quantiles. Mann-Whitney  $U$  tests were used to assess the relative strength of internal and expressed update strengths per user, checking whether the probability  $P(\alpha_{e,i} > \alpha_{u,i})$  is larger than  $P(\alpha_{e,i} < \alpha_{u,i})$  at the .05 level [54]. The proportion of users for which this is the case is denoted as. Inferred parameters enabled the reconstruction of the latent sentiment trajectory of the users over the period of interest. Monotonic trends in the population median of inferred latent sentiment were identified by a Hamed-Rao test, a modification of the Mann-Kendall test that accounts for autocorrelation [55]. With this test, we checked for each of the subsequent periods delimited by the events in Table S1 in Multimedia Appendix 2 if the population median showed such a monotonic trend with .05 significance.

## Ethical Considerations

This study was not subject to formal ethics board review, as Belgian law restricts such requirements to medical

experiments, animal studies, and dual-use research, consistent with Ghent University's institutional ethics framework [56]. Data were retrieved from the public subreddit, in line with Reddit's terms of service [57]. Although the raw extract contained identifiers, these were not included in the analytical dataset and were not published; all reported results are aggregated and anonymized. Therefore, this study follows the Association of Internet Researchers ethical guidelines for internet research [58].

## Results

### Dataset Characteristics

The final dataset comprised 28,559 unique users who created 645,280 comments and 10,362 submissions (Table 1). English was the predominant language, accounting for approximately 75% of the posts. Dutch posts represented roughly 10%, while French and German posts were substantially underrepresented, with less than 1% of the posts. Topic distribution was relatively balanced, with lockdowns being most prevalent (94,494 comments, 1009 submissions from 9987 users), followed by masks and vaccination.

**Table 1.** Number of users, comments, and submissions in the considered dataset of posts on r/Belgium (January 2020-June 2022) on COVID-19 mitigation topics.

|                 | Users, n | Comments, n | Submissions, n |
|-----------------|----------|-------------|----------------|
| Total           | 28,559   | 645,280     | 10,362         |
| Per language    |          |             |                |
| EN <sup>a</sup> | 24,850   | 488,468     | 6405           |
| NL <sup>b</sup> | 9344     | 71,940      | 3240           |
| FR <sup>c</sup> | 696      | 1036        | 145            |
| DE <sup>d</sup> | 272      | 335         | 4              |
| Per topic       |          |             |                |
| Lockdowns       | 9987     | 94,494      | 1009           |
| Masks           | 6824     | 48,500      | 437            |
| Vaccination     | 5552     | 41,700      | 590            |

<sup>a</sup>EN: English.

<sup>b</sup>NL: Dutch.

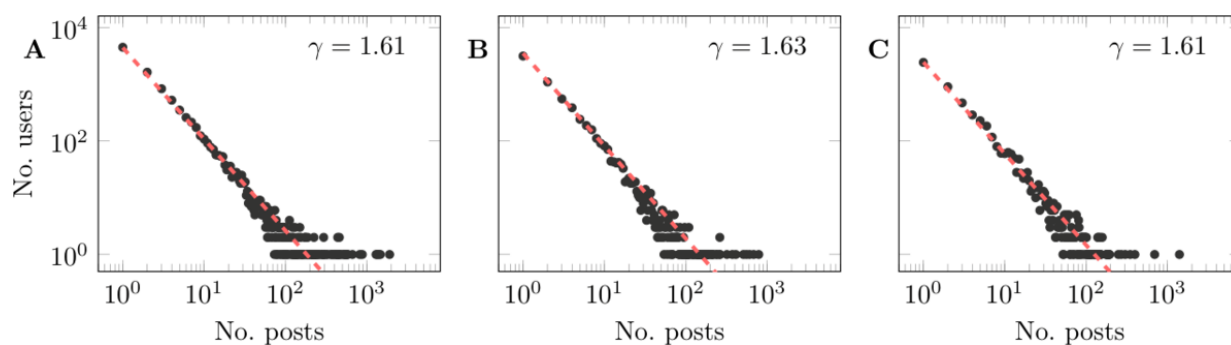
<sup>c</sup>FR: French.

<sup>d</sup>DE: German.

The distribution of posts per user exhibited a heavy-tailed pattern for all topics (Figure 3), consistent with previous observations on Reddit [59]. A small number of users created many posts, while the majority contributed only a single post,

limiting the ability to draw conclusions about individual user behavior patterns. Power-law approximations yielded similar exponents  $\gamma$  for all topics.

**Figure 3.** Distribution of posts per user in r/Belgium (January 2020–June 2022) on a double logarithmic scale for (A) lockdowns, (B) masks, and (C) vaccination, and a power-law approximation with exponent  $\gamma$  (dashed, red).

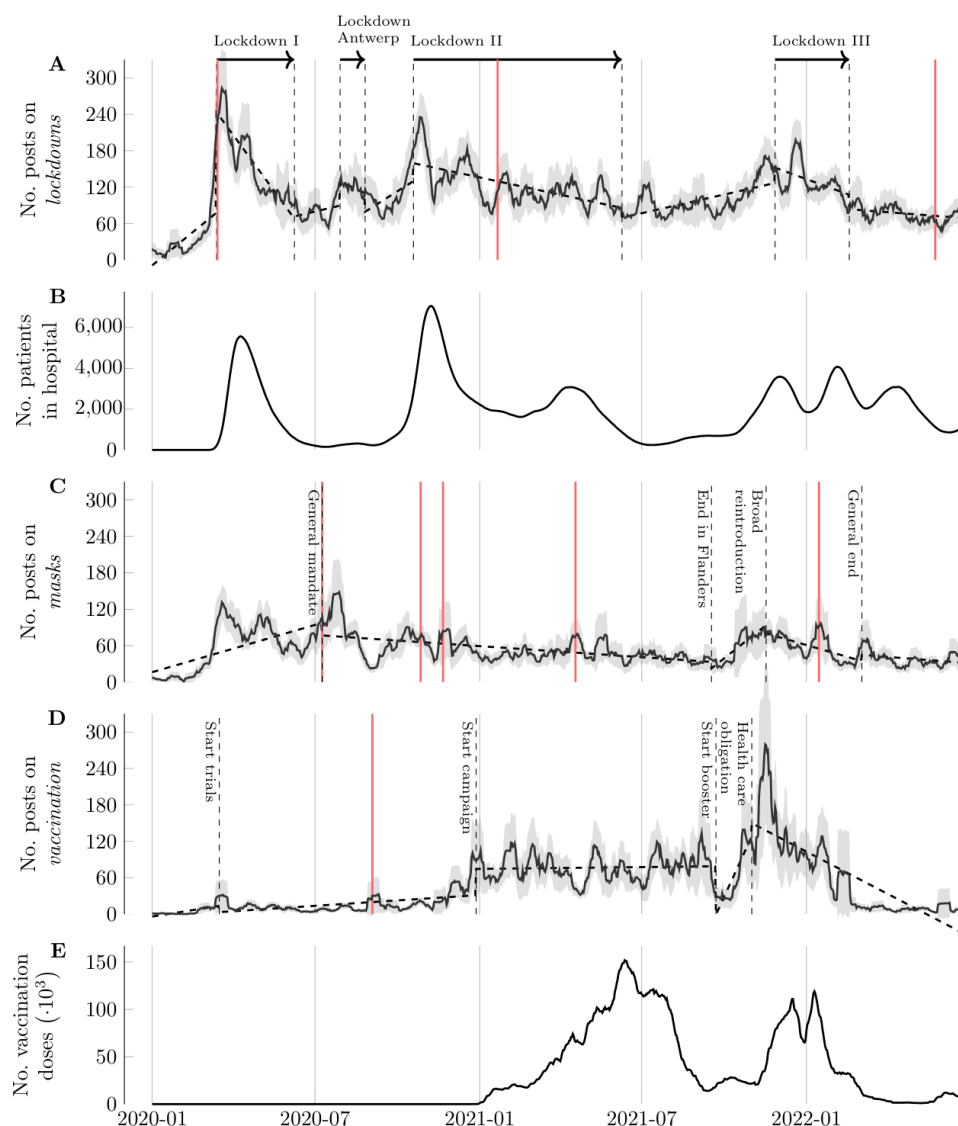


## Temporal Analysis

Figure 4 shows posting volume over time for the 3 topics, alongside key events and epidemiological indicators. It gives an overview of posting behavior over the period of the pandemic and associated events. The largest peaks in the number of posts coincided with the external events listed in Table S1 in [Multimedia Appendix 2](#). For example, the start of the first lockdown coincided with the single most active day across all topics, with a rolling mean of 282.79 (SD 126.68) posts on March 19, 2020. Smaller local maxima appeared during the Antwerp lockdown and at the start of the second national one. All events coincided with significant changes in

post volume. The piecewise linear trends overlaid on [Figure 4](#) (dashed lines) make these structural changes visible, with the slope and level shifts quantified in [Table 2](#). After these peaks, the volume of posts decreased and stabilized. Three days showed a significantly negative sentiment (days where the median score-weighted sentiment fell below the 0.275 quantile of the topic distribution, with at least 50 comments submitted): at the start of the first lockdown, in the middle of the second national one, and in February 2022, after the end of most mitigation measures. During the first one, negative comments were aimed at “lockdown parties,” social gatherings just before the restrictions came into effect.

**Figure 4.** The number of posts per day on r/Belgium on (A) lockdowns, (C) masks, and (D) vaccination; (B) the number of COVID-19 hospitalizations; and (E) vaccination doses in Belgium. All quantities are shown as their 2-week rolling mean. Piecewise linear trends are given by dashed lines. Red vertical lines mark significantly negative days. Important events are denoted by vertical dashed lines or gray regions (Multimedia Appendix 2).



**Table 2.** Results of the interrupted time series analysis. Change in level  $\Delta\beta_{0,j}$  and in slope  $\Delta\beta_{1,j}$  at the time of each event  $j$  per topic, with their 95% CIs.

| Events                 | Dates              | $\Delta\beta_{0,j}$ (95% CI)        | $\Delta\beta_{1,j}$ (95% CI) |
|------------------------|--------------------|-------------------------------------|------------------------------|
| <b>Lockdowns</b>       |                    |                                     |                              |
| Start lockdown 1       | March 13, 2020     | 165.17 <sup>a</sup> (151.39-178.96) | -3.22 (-3.53 to -2.91)       |
| End lockdown 1         | June 8, 2020       | 3.82 (-11.32 to 18.96)              | 2.36 (1.91-2.81)             |
| Start lockdown Antwerp | July 29, 2020      | 39.85 (19.77 to 59.93)              | -1.05 (-2.14 to 0.04)        |
| End lockdown Antwerp   | August 26, 2020    | -30.33 (-50.72 to -9.94)            | 1.64 (0.56-2.72)             |
| Start lockdown II      | October 19, 2020   | 29.43 (16.24-42.62)                 | -1.25 (-1.63 to -0.87)       |
| End lockdown II        | June 9, 2021       | -15.93 (-24.62 to -7.25)            | 0.65 (0.57-0.73)             |
| Start lockdown III     | November 27, 2021  | 25.16 (13.66-36.67)                 | -0.87 (-1.08 to -0.66)       |
| End lockdown III       | February 18, 2022  | -23.90 (-36.02 to -11.77)           | 0.42 (0.20-0.64)             |
| <b>Masks</b>           |                    |                                     |                              |
| General mandate        | July 9, 2020       | -20.73 (-27.74 to -13.72)           | -0.53 (-0.58 to -0.47)       |
| End in Flanders        | September 17, 2021 | -10.56 (-21.44 to 0.33)             | 1.27 (0.97 to 1.56)          |
| Broad reintroduction   | November 17, 2021  | -13.69 (-26.67 to -0.72)            | -1.59 (-1.91 to -1.27)       |
| General end            | March 4, 2022      | 10.10 (-0.63 to 20.82)              | 0.31 (0.15-0.48)             |

| Events                 | Dates              | $\Delta\beta_{0,j}$ (95% CI) | $\Delta\beta_{1,j}$ (95% CI) |
|------------------------|--------------------|------------------------------|------------------------------|
| Vaccination            |                    |                              |                              |
| First trials           | March 16, 2020     | -10.36 (-22.61 to 1.90)      | -0.14 (-0.39 to 0.11)        |
| Start campaign         | December 28, 2020  | 43.29 (35.32-51.26)          | -0.08 (-0.13 to -0.03)       |
| Start booster          | September 22, 2021 | -79.62 (-95.27 to -63.96)    | 3.21 (2.57 to 3.86)          |
| Health care obligation | November 19, 2021  | 21.46 (5.20-37.73)           | -4 (-4.64 to -3.35)          |

<sup>a</sup>Italics indicate significant changes.

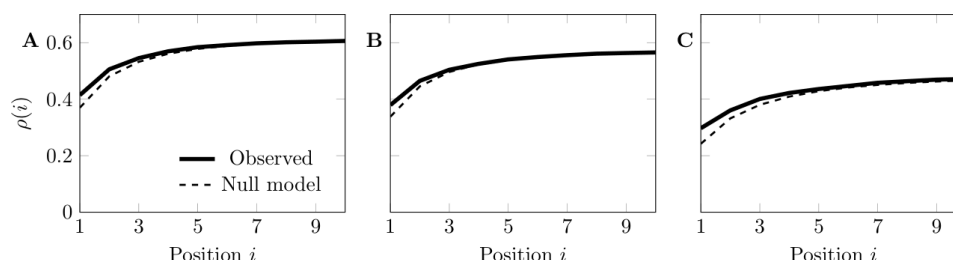
Discussions on masks began at the onset of the COVID-19 outbreak in Belgium, four months before the announcement of a general mandate (July 9, 2020), which itself did coincide with a small increase in posts, and a significantly negative day. During the latter, negative comments criticized the late decision of the government to introduce the mandate, and people not following it. Activity on mask-related discussions fell after the first lockdown but rose again during the second mandate period (November 2021-March 2022), which included another significantly negative day. Another 3 significantly negative days occurred during the first mandate. All events were associated with a significant change in trend, and after the 2 moments the mandate was expanded, the level changed significantly as well (Table 2). Post volume generally increased before mask mandates were tightened and started decreasing afterward.

Similar to the lockdown-related events, major vaccine-related events or news corresponded well with the volume of vaccination posts on Reddit. Broad discussions did not start until the run-up to the vaccination campaign, with the only nonsignificant change in level or volume occurring at the start of the trials (Table 2). The announcement that vaccination would be mandatory for health care personnel preceded the month with the highest posting activity, culminating in a rolling mean peak of 279.64 (SD 124.17) posts on November 16, 2021. The single significantly negative day, September 3, 2020, fell during the trials. Comments reacted to an article stating that 30% of the Belgian population would refuse a COVID-19 vaccine [60]. The negative reactions targeted both refusers and the “rushed” development of the new mRNA technology.

## Topic Contagion

We tested whether prior participation in a discussion on a topic increases the probability of subsequently initiating a new one, which serves as an indication of social contagion. Under the null hypothesis (no contagion), initiations and participations are ordered uniformly at random. Under contagion, on the other hand, participation would increase the probability of later initiating a new discussion, causing participations to accumulate earlier in users' sequences. Figure 5 shows the proportion of users with at least one initiation among their first discussions, relative to those with a participation or initiation,  $\rho(i)$ . The solid line represents the observed cumulative proportion and the dashed line is the expected proportion under the null model. If contagion were present,  $\rho(i)$  would lie below the null for small  $i$ , reflecting an accumulation of participations before the first initiation. For all topics, the observed curve lies above the null at low values of  $i$ , indicating initiations occurred earlier in posting sequences than expected. This contradicts the contagion hypothesis: initiations were more likely near the start of sequences. Consequently, users who participated in a discussion were not more likely to subsequently initiate one on the same topic. No evidence of social contagion was found for discussion initiation on the 3 mitigation topics within r/Belgium, neither simple nor complex. As  $\rho(i)$  converges, it approaches the proportion of users with at least one initiation: 0.612 for lockdowns, 0.572 for masks, and 0.481 for vaccination. Hence, not only did discussions on vaccination appear later than those on the other topics, they were initiated by a relatively smaller fraction of the user base.

**Figure 5.** Proportion of users  $\rho(i)$  that have at least one initiated post in the first  $i$  discussions for the three COVID-19 mitigation topics—(A) lockdowns (9987 users), (B) masks (6824 users), and (C) vaccination (5552 users)—in r/Belgium (January 2020-June 2022). The full line represents the observed proportion, and the dashed line represents the expected proportion under the null model.



## Sentiment Homophily

We examined the relationship between the sentiment of a comment and that of its parent, looking for correlations between these, namely homophily. Figure 6 presents heatmaps of comment-parent sentiment pairs  $H$  and  $\Delta H$  for

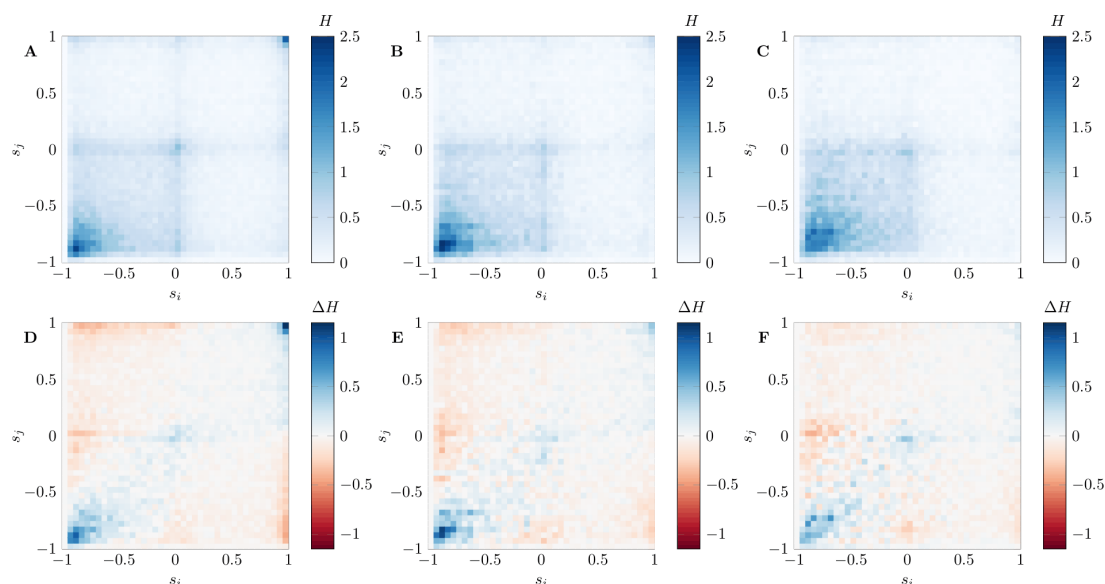
each topic. Across all topics, the center of mass of observed interactions  $H$  lay in the fourth quadrant, with a global maximum near  $(-1, -1)$ , indicating most posts expressed negative sentiment. For lockdowns, another local maximum appears in the first quadrant corner. Discussions around

masks and vaccination exhibited consensus (predominantly negative), while those on lockdowns showed polarization (negative replies to negative, positive to positive) [29].

Both consensus and polarization fall under the broader term homophily, the tendency to interact with others expressing similar sentiment. Homophily is expected when the underlying distribution is narrow, offering few interaction alternatives. However, the heatmaps of  $\Delta H$  in Figure 6 demonstrate homophily for all topics, irrespective of sentiment distribution shape. Positive values center around the diagonal, indicating more interactions between similar sentiment than expected. Negative values appear in the second and third quadrants, and in regions representing interactions between negative and

neutral sentiment. Discussions on lockdowns were the most polarized, with clear peaks in the equally-signed corners. The discrepancy between observed and expected values reveals greater polarization than  $H$  alone suggested for masks, with positive values of  $\Delta H$  near (1,1). The homophily measure  $h$  (Equation A.7 in Multimedia Appendix 3) is 0.246, 0.212, and 0.144 for lockdowns, masks, and vaccination, respectively. As Figure 6 suggests, homophily is most pronounced for lockdowns and least pronounced for vaccination. This mirrors the pattern reported in sections *Temporal Analysis* and *Topic Contagion*, where lockdowns and vaccination produce the most differing results, with masks showing more similarities to lockdowns than to vaccination.

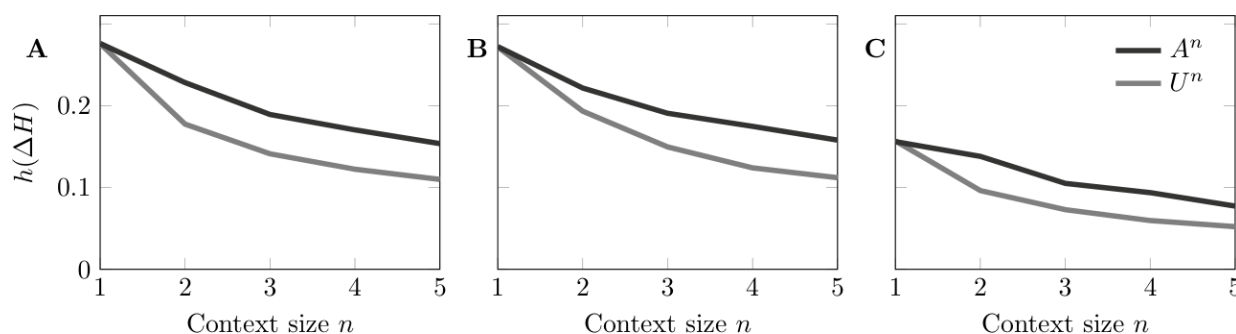
**Figure 6.** Normalized histograms of (A, B, C) observed sentiment interactions on r/Belgium (January 2020–June 2022) and (D, E, F) the difference with the null model for the COVID-19 mitigation topics (A, D) lockdowns (94,494 pairs), (B, E) masks (48,500 pairs), and (C, F) vaccination (41,700 pairs).



We examined which context type exhibited stronger homophily: ancestral or  $A_k^n$  user  $U_k^n$ . The former comprises posts the focal user most recently read, while the latter comprises posts with which the user most recently interacted. For this, a subsample of comments allowing a sufficient size for both contexts was considered (9067, 5372, and 4893 comments for lockdowns, masks, and vaccination, respectively). Figure 7 presents the homophily  $h$  measure for all 3 topics and both context types as a function of context size  $n$ , the number of preceding posts that are taken into

account. With  $h$  being 0.276 (lockdowns), 0.272 (masks), and 0.156 (vaccination) for  $n = 1$ , homophily in this subset is more pronounced, but the relative ordering of topics persists. The homophilic effect decreased with increasing context size  $n$ , regardless of type. Homophily within the ancestral context consistently exceeded that in the user context for all  $n$ . Comment sentiment was thus more strongly related to the immediate posting environment than to sentiment that previously prompted replies from the focal user. Comments on vaccination exhibited less homophily than the other topics.

**Figure 7.** Measure of homophily  $h$  between the sentiment of a comment and its context size  $n$  (ancestral context: black; user context: gray) for the COVID-19 mitigation topics (A) lockdowns (9067 comments), (B) masks (5372), and (C) vaccination (4893 comments) on r/Belgium (January 2020-June 2022).



## Sentiment Evolution Model

The SLEBC model (Equation 1) was set up to give mechanistic insights into the results of section *Sentiment Homophily*. To fit the parameters of this model and its alternatives, we considered 209, 101, and 98 users for lockdowns, masks, and vaccination, respectively, each with at least 40 comments. These users were the most active and thus yielded sufficient data, though the homophily within this population differed from the complete set (relative difference of  $-10.5\%$  for lockdowns,  $+12.5\%$  for masks, and  $+13.7\%$  for vaccination). Consequently, the results might not apply to the less active users.

Table 3 compares the performance of the 3 models, with lower WAIC and  $|\Delta h|$  indicating better model fit. All models systematically underestimated homophily, as

indicated by negative  $\Delta h$ , with only SLEBC having 95% credible intervals that encompass zero. The SLEBC model outperformed both alternatives across all topics, achieving the lowest WAIC values and homophily deficits closest to zero. The two alternatives each have their weaknesses. The linear model (Equation A.9) in Multimedia Appendix 3) achieved comparable predictive power to SLEBC, but substantially underestimated homophily. The stateless model (Equation A.10 in Multimedia Appendix 3), which ablated the latent state, achieved a worse fit, but more accurately predicted homophily. These results suggest the bounded confidence kernel is better suited to model the sentiment homophily, with the model lacking it, underestimating  $h$  more severely. Moreover, the latent state improves capturing the sentiment distribution over time by accounting for fluctuating expressed sentiment.

**Table 3.** Model comparison showing Watanabe-Akaike information criterion and homophily difference  $\Delta h$  (with 95% credible interval [CrI]) for the smooth latent-expressed bounded confidence model and its two alternatives, the linear model  $L_\alpha$  and the stateless model  $\bar{e}$ . Applied to highly active users ( $\geq 40$  comments) of r/Belgium (January 2020-June 2022) in COVID-19 mitigation topics: lockdowns (209 users, 16,762 comments), masks (101 users, 4277 comments), and vaccination (98 users, 3169 comments).

| Model                           | WAIC <sup>a</sup> | $\Delta h$ (95% CrI)         |
|---------------------------------|-------------------|------------------------------|
| <b>Lockdowns</b>                |                   |                              |
| $L_\alpha$ (Equation A.9)       | -21.11            | -0.0973 (-0.1084 to -0.0720) |
| $\bar{e}$ (Equation A.10)       | 692               | -0.0702 (-0.1078 to -0.0441) |
| SLEBC (Equation 1)              | -28.5             | -0.0273 (-0.0511 to 0.0015)  |
| <b>Masks</b>                    |                   |                              |
| $L_\alpha$ (Equation A.9)       | -17.4             | -0.1317 (-0.1651 to -0.0972) |
| $\bar{e}$ (Equation A.10)       | 169               | -0.0759 (-0.1534 to -0.0276) |
| SLEBC <sup>b</sup> (Equation 1) | -18.4             | -0.0671 (-0.0816 to 0.0103)  |
| <b>Vaccination</b>              |                   |                              |
| $L_\alpha$ (Equation A.9)       | -20.6             | -0.1032 (-0.1393 to -0.0732) |
| $\bar{e}$ (Equation A.10)       | 6.81              | -0.0672 (-0.1395 to -0.0410) |
| SLEBC (Equation 1)              | -21.2             | -0.0337 (-0.0829 to 0.0046)  |

<sup>a</sup>WAIC: Watanabe-Akaike information criterion.

<sup>b</sup>SLEBC: smooth latent-expressed bounded confidence.

Figure 8 compares per topic the posterior sentiment distributions and homophily values obtained from the SLEBC model to those observed in the data. The SLEBC model predicted sentiment distributions that placed most mass near the negative extreme, while spreading more evenly over

their support than observed distributions and lacking their tri-modal structure (Figure 8A-C). In particular, the local maximum around zero (neutral sentiment) present in the data was not captured. On the positive extreme, both observed and predicted distributions exhibited a local maximum

for lockdowns, though this peak was substantially more pronounced in the data. Applying the homophily analysis from section *Sentiment Homophily* to the predictions showed that the SLEBC model generated homophily (Figure 8-D), but underestimated it for all topics, with median values exceeding 0.75 quantiles of predicted values for all topics, and even the maximum for masks.

**Figure 8.** Outcomes of the fitted posterior of the smooth latent-expressed bounded confidence model on highly active users ( $\geq 40$  comments) of r/Belgium (January 2020-June 2022). Normalized distributions  $D$  of the predicted sentiment for the topics (A) lockdowns (209 users, 16,762 comments), (B) masks (101 users, 4277 comments), and (C) vaccination (98 users, 3169 comments) in the smooth latent-expressed bounded confidence model, compared to the observed sentiment. (D) The predicted values of  $h$  by the smooth latent-expressed bounded confidence model, with minimums, maximums, and 0.25, 0.5, and 0.75 quantiles. The dashed red line represents the observed value.

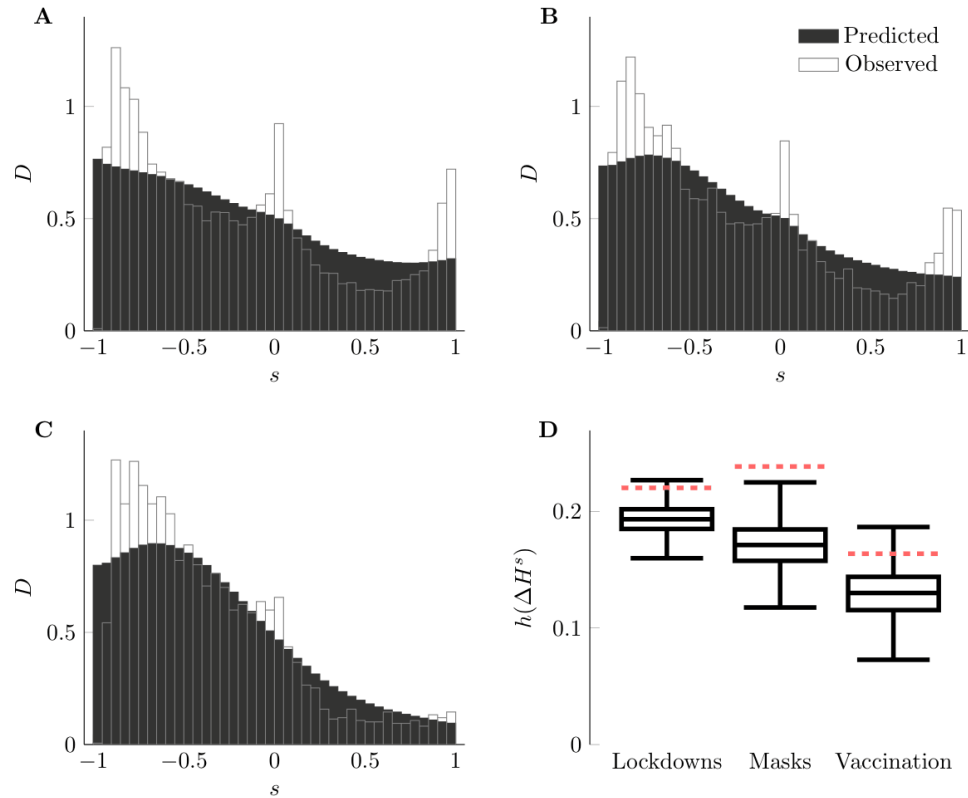


Table 4 presents a quantitative summary of the sentiment threshold  $\epsilon$  and update strength parameters  $\alpha_u$  and in the SLEBC model (Equation 1). For all topics and most users, expressed sentiment was more strongly influenced by parent comment sentiment than latent sentiment was by other interactions, as  $\kappa$  exceeded one half, although only marginally for vaccination. Therefore, adaptation to parent comment sentiment constituted an important mechanism behind the observed homophily, consistent with findings in section *Sentiment Homophily*. The latent update strength  $\alpha_u$  varied less across topics and users within a topic, as shown by its narrower quantile intervals, indicating that the degree of expressed sentiment adaptation to the parent depended more on user-specific behavior.

**Table 4.** The population mean values of  $\epsilon$ ,  $\alpha_u$ , and  $\alpha_e$  with their 95% quantile intervals (95% QIs) in the smooth latent-expressed bounded confidence model. Equation 1 sampled from the posterior distribution, together with  $\kappa$ , the proportion of users for which  $P(\alpha_{u,i} < \alpha_{e,i}) > P(\alpha_{u,i} > \alpha_{e,i})$ , with  $P < .05$  for highly active users, in COVID-19 mitigation discussions on r/Belgium (January 2020-June 2022) on lockdowns (209 users), masks (101 users), and vaccination (98 users).

| Topic       | $\epsilon$ | 95% QI        | $\alpha_u$ | 95% QI        | $\alpha_e$ | 95% QI        | $\kappa$ |
|-------------|------------|---------------|------------|---------------|------------|---------------|----------|
| Lockdowns   | 0.973      | (0.367-1.566) | 0.414      | (0.126-0.706) | 0.871      | (0.269-3.694) | 0.742    |
| Masks       | 0.874      | (0.372-1.510) | 0.417      | (0.238-0.609) | 0.656      | (0.265-1.933) | 0.703    |
| Vaccination | 0.971      | (0.445-1.484) | 0.425      | (0.139-0.788) | 0.508      | (0.234-1.389) | 0.510    |

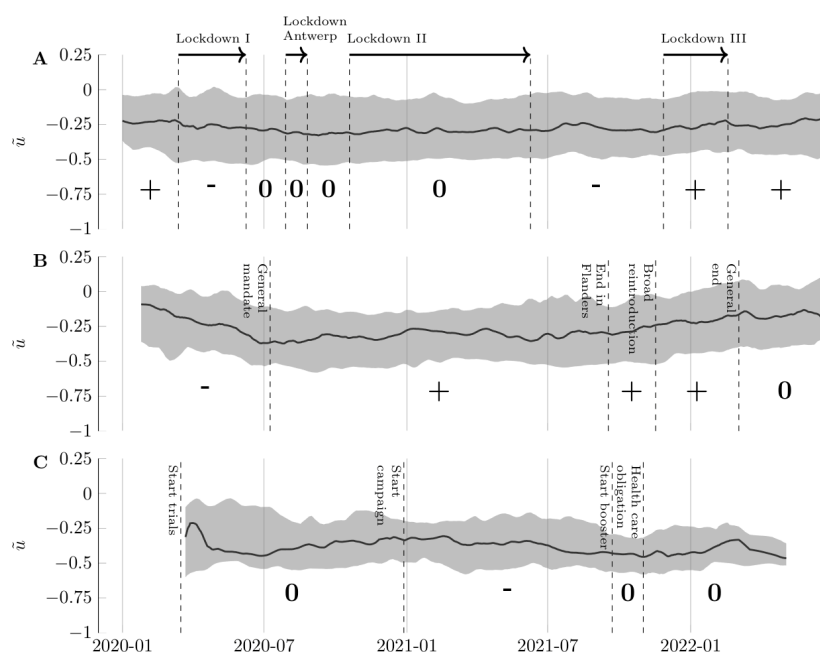
The mean sentiment threshold  $\epsilon$  is lower for masks than the other two topics, while the quantile interval for vaccination is the narrowest. For all topics, the mean value is just under half of the maximum meaningful value, two. The quantile intervals indicate that only a small number of users have very low or very high ones. The SLEBC model thus infers that users were moderately open to influence: they adapted their sentiment toward that of others unless opinions diverged by more than roughly half the scale, for example, between negative and neutral. In contrast, only a limited

number of users displayed no sentiment adaptation ( $\epsilon \approx 0$ ), or indiscriminate adaptation ( $\epsilon \approx 2$ ).

A key feature of the SLEBC model is its ability to separate the influence of a user's latent sentiment from that of the parent comment on the expressed sentiment. For assessing sentiment toward a mitigation measure, the former is particularly informative, as it captures the less noisy, long-term trends. Figure 9 presents the evolution of the inferred latent state of the SLEBC model,  $u$ , for the studied

population. For masks, median latent sentiment declined during the first half of 2020 before gradually increasing. For lockdowns, latent sentiment remained stably negative throughout the study period. The Hamed-Rao test identified both increasing and decreasing trends at the start and end of the considered period; however, the absolute change was small. This is the case for vaccination as well, although sentiment initially showed a short positive tendency around the start of the trials.

**Figure 9.** Median latent sentiment  $u$  of highly active users over time (full black curve) for the COVID-19 mitigation topics (A) lockdowns (209 users), (B) masks (101 users), and (C) vaccination (98 users) on r/Belgium (January 2020–June 2022). The area between the first and third quartile is shaded. Hamed-Rao monotonic test results are given per period delimited by events in Multimedia Appendix 2 (+: increasing, -: decreasing, 0: no significant trend).



## Discussion

### Principal Findings

This study analyzed discussions on COVID-19 mitigation measures on the Belgian Reddit community (r/Belgium) from January 2020 to June 2022, examining how discussion volume and sentiment evolved around lockdowns, masks, and vaccination. We found that discussion volume was primarily associated with external events such as policy announcements and media coverage rather than with social contagion within the platform. Peaks in posting activity aligned with major policy changes, including the first lockdown and the announcement of mandatory vaccination for health care personnel, suggesting that traditional media and official communications may play an important role in shaping health discourse in r/Belgium, as has been confirmed on Twitter [11, 61] and other Reddit communities [62]. This contrasts with observations for other online behaviors where complex social contagion mechanisms have been documented [34].

Although topic-level contagion was not significant, sentiment exhibited homophily, with users adapting their sentiment to match that of their parent comment. The

observed homophily measure  $h$  ranged from 0.144 to 0.246 across the 3 topics, confirming patterns previously documented on Reddit [16,43]. Our analysis suggested that this homophily operated primarily through immediate conversational adaptation rather than through selective interaction with like-minded users. To capture this dynamic, we developed the SLEBC model, which distinguishes between a user's unobserved latent sentiment and the observed sentiment expressed in comments. The model successfully reproduced observed patterns and revealed that most users (51%–74% depending on topic) adjusted their expressed sentiment more strongly to match that of the parent than they updated their latent sentiment state.

The SLEBC model, building on bounded confidence modeling [48], outperformed linear models, indicating that sentiment alignment occurred only when users encountered other users with sufficiently similar sentiment. Our results expand on the existing literature on using the bounded confidence model in steady-state opinion distributions on social media [29], and spatial patterns of vaccine hesitancy [36]. Nevertheless, a linear model explicitly using social network structure has been successfully calibrated on Italian vaccine hesitancy survey data, with the authors naming the

bounded confidence model as a possible extension [63]. By also incorporating a latent sentiment state, the SLEBC model corrected for the fluctuating nature of the expressed sentiment and better inferred the observed distribution of sentiment over time. This latent state may better represent a user's underlying sentiment trajectory than their fluctuating expressions alone, making it a stronger signal for relating online discourse to behavioral indicators. Discussions on different mitigation measures showed distinct patterns on r/Belgium, similarly to Twitter [20]. Discussions on lockdowns dominated in volume and emerged when restrictions were first imposed. This aligns with the increase in the volume of COVID-19-related posts observed in other country-related subreddits during a lockdown period [28]. Masks were discussed from the pandemic's onset, before any official mandate. Vaccination, by contrast, did not attract broad discussion until the run-up to the campaign. A possible explanation for the temporal differences is that mask wearing was a personal choice before the official mandate, whereas vaccination was impossible before the government launched the campaign.

The 3 topics also differed in sentiment. Discussions on lockdowns showed stronger polarization with distinct positive and negative sentiment clusters. In contrast, discussions on masks and vaccination both exhibited predominantly negative sentiment, though vaccination showed weaker homophily ( $h=0.144$ ) than masks ( $h=0.212$ ). This is mirrored in the inferred parameters of the SLEBC model. For lockdowns, the mean external update strength is strongest, followed by masks and vaccination. In contrast, population distributions of sentiment thresholds and latent update strengths showed more similarities over the topics. Aforementioned differences underscore the importance of analyzing mitigation topics separately rather than aggregating them into a single "COVID-19 sentiment" variable.

Existing coupled epidemiological-social models may implicitly treat social media activity as a direct proxy for the infection dynamics, whereas our results indicate that a communication layer should be considered as well [64]. In particular, modeling of discussion volume may benefit from an intermediate layer representing official communications and traditional media such as the OxCGRT stringency index [65] or GDELT dataset [66]. Users can then react to activity in this layer, allowing for an increase in discussion volume without within-platform contagion. Sentiment dynamics are better represented by complex contagion models such as bounded confidence kernels rather than by simple ones. The methodological framework of combining topic modeling, sentiment analysis, null model comparisons, and mechanistic modeling provides a template for digital health surveillance deployment during health emergencies, with established platforms like Reddit offering advantages as data infrastructure and analysis pipelines can be prepared in advance.

## Limitations and Future Work

This study has several limitations. The r/Belgium community represents a self-selected, nonrepresentative sample of the Belgian population. Reddit users tend to belong to a specific demographic, often young and male [31]. This effect

intersects with our focus on English-language posts (approximately 75% of the dataset). However, because English acts as a lingua franca on r/Belgium (and the majority of other Reddit communities) that bridges the country's distinct linguistic communities, restricting the analysis to English avoids introducing confounding variables from isolated subgroups that communicate exclusively in Dutch, French, or German. Moreover, SLEBC parameters were inferred from highly active users, whose homophily values differed from the full sample. Therefore, inferred parameters and latent trajectories may not represent the sentiment states of casual users. Nevertheless, qualitative findings such as the adaptation of expressed sentiment to the parent post are furthermore supported by the full dataset analysis in section *Sentiment Homophily*. Still, our findings represent platform- and user-specific dynamics that may not generalize, as highly active, English-proficient Reddit users may differ systematically in demographics, political orientation, or health attitudes from other communities. Applying our approach across different communities and platforms would help clarify how specific platform architectures influence sentiment homophily [15,25,61].

Sentiment analysis captures linguistic tone rather than explicit policy support, introducing ambiguity where negative sentiment could reflect criticism of either insufficient or excessive measures. The rapid increase in ability and availability of large language models can enhance our methodology by extracting more nuanced signals such as support for public health measures and willingness to adhere. The inferred latent states are unobservable mathematical constructs. While our model successfully reconstructs latent sentiment, it remains to be established whether these trajectories serve as reliable proxies for true offline attitudes. Future work could address this by relating the inferred latent states to behavioral surveys or adherence metrics.

## Conclusions

Our study showed that discussions on COVID-19 mitigation measures on Reddit's r/Belgium were not associated with social contagion within the platform but aligned with external events and traditional media. In contrast, sentiment dynamics were shaped by within-thread interactions. Users exhibited sentiment homophily by adapting their expressed sentiment to match that of preceding comments in the thread. The SLEBC model captures this by distinguishing between an unobserved latent sentiment state and sentiment expressed in comments, suggesting that expressed sentiment adapts strongly to align with that of the parent comment and therefore may poorly reflect users' underlying sentiment.

These findings have implications for Reddit-based digital health surveillance and epidemic modeling that uses social media signals. First, rather than assuming purely endogenous social contagion, models benefit from an intermediate communication layer of external official communications and traditional media as the driver of topic emergence. Second, when integrating social sentiment as a proxy for public adherence, models should account for interaction structure. In COVID-19 mitigation discussions on r/Belgium,

expressed sentiment adapted primarily to the immediate parent comment rather than to a user's broader interaction history, and this adaptation was better captured by a bounded confidence kernel than by a linear update, with the latter substantially underestimating observed homophily. Therefore, raw expressed sentiment is a noisy, perturbed signal. The

inferred latent trajectory is more stable and, if validated against behavioral outcomes, is a stronger candidate for model input. Finally, mitigation measures exhibited distinct patterns in our research and should not be collapsed into a single "COVID-19 sentiment" variable, which would obscure relevant differences.

---

## Acknowledgments

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation) taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: code optimization, process automation, proofreading and editing, reformatting, and publication support. The GenAI tools used were ChatGPT 5.1 (OpenAI) and Claude Sonnet 4.5 (Anthropic). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

---

## Funding

This work was supported by the Research Foundation — Flanders (FWO; grant numbers G0G0122N and W001625N [LECR and JB]), Special Research Fund (grant number 2024/01/709 [LECR]), Atlantic Coast Center for Infectious Disease Dynamics and Analytics - ACCIDDA (CDC CFA; grant NU38FT000012 [TWA]), and National Institutes of Health (grant number 5R24GM153920-02 [TWA]). Computational infrastructure was provided by the Flemish Supercomputer Center (VSC). This paper was also published with the support of the University Foundation of Belgium. The organizations did not in any way influence the design, execution or results of this work.

---

## Data Availability

Instructions on how to obtain the used dataset [38,68] and reproduce the findings (Figures 3-9, Tables 1-4) are available in the GitHub repository [67]. According to the Reddit Terms of Service, their data are freely usable for nonprofit purposes [57].

---

## Authors' Contributions

Conceptualization: TVW, LECR, JMB

Formal analysis: TVW

Funding acquisition: LECR, JMB

Investigation: TWA

Methodology: TVW, LECR, JMB

Project administration: JMB

Supervision: TWA

Validation: TWA

Visualization: TVW

Writing – original draft: TVW

Writing – review and editing: LECR, TWA, JMB

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Glossary.

[PDF File (Adobe File), 196 KB-Multimedia Appendix 1]

---

## Multimedia Appendix 2

Overview of selected events included in this study.

[PDF File (Adobe File), 230 KB-Multimedia Appendix 2]

---

## Multimedia Appendix 3

Supporting equations and methods.

[PDF File (Adobe File), 253 KB-Multimedia Appendix 3]

---

## Multimedia Appendix 4

Sensitivity analysis of homophily measure.

[PDF File (Adobe File), 243 KB-Multimedia Appendix 4]

## References

- Piette CH, Tielens J. How Belgian firms fared in the COVID-19 pandemic? *Natl Bank Belg Econ Rev*. 2022;1-33. URL: <https://hdl.handle.net/11159/631033> [Accessed 2026-06-25]
- Reiriz M, Donoso-González M, Rodríguez-Expósito B, Uceda S, Beltrán-Velasco AI. Impact of COVID-19 confinement on mental health in youth and vulnerable populations: an extensive narrative review. *Sustainability*. Feb 7, 2023;15(4):3087. [doi: [10.3390/su15043087](https://doi.org/10.3390/su15043087)]
- Alleman TW, Schoors K, Baetens JM. A dynamic disequilibrium input–output model of the Belgian COVID-19 pandemic. *Econ Syst Res*. Feb 26, 2026;1-25. [doi: [10.1080/09535314.2026.2622365](https://doi.org/10.1080/09535314.2026.2622365)]
- van den Broek-Altenburg E, Atherly A. Adherence to COVID-19 policy measures: behavioral insights from the Netherlands and Belgium. *PLoS One*. 2021;16(5):e0250302. [doi: [10.1371/journal.pone.0250302](https://doi.org/10.1371/journal.pone.0250302)] [Medline: [34048441](https://pubmed.ncbi.nlm.nih.gov/34048441/)]
- Luyten J, Schokkaert E. Belgium’s response to the COVID-19 pandemic. *Health Econ Policy Law*. Jan 2022;17(1):37-47. [doi: [10.1017/S1744133121000232](https://doi.org/10.1017/S1744133121000232)] [Medline: [34219632](https://pubmed.ncbi.nlm.nih.gov/34219632/)]
- West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nat Hum Behav*. May 2020;4(5):451-459. [doi: [10.1038/s41562-020-0887-9](https://doi.org/10.1038/s41562-020-0887-9)] [Medline: [32377018](https://pubmed.ncbi.nlm.nih.gov/32377018/)]
- Krawczyk K, Chelkowski T, Laydon DJ, et al. Quantifying online news media coverage of the COVID-19 pandemic: text mining study and resource. *J Med Internet Res*. Jun 2, 2021;23(6):e28253. [doi: [10.2196/28253](https://doi.org/10.2196/28253)] [Medline: [33900934](https://pubmed.ncbi.nlm.nih.gov/33900934/)]
- Cinelli M, Quattrocioni W, Galeazzi A, et al. The COVID-19 social media infodemic. *Sci Rep*. Oct 6, 2020;10(1):16598. [doi: [10.1038/s41598-020-73510-5](https://doi.org/10.1038/s41598-020-73510-5)] [Medline: [33024152](https://pubmed.ncbi.nlm.nih.gov/33024152/)]
- Cinelli M, Gesualdo F. Infodemic versus viral information spread: key differences and open challenges. *JMIR Infodemiology*. May 7, 2025;5(1):e57455. [doi: [10.2196/57455](https://doi.org/10.2196/57455)] [Medline: [40334284](https://pubmed.ncbi.nlm.nih.gov/40334284/)]
- Alamoodi AH, Zaidan BB, Zaidan AA, et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review. *Expert Syst Appl*. Apr 1, 2021;167:114155. [doi: [10.1016/j.eswa.2020.114155](https://doi.org/10.1016/j.eswa.2020.114155)] [Medline: [33139966](https://pubmed.ncbi.nlm.nih.gov/33139966/)]
- Kurten S, Beullens K. #Coronavirus: monitoring the Belgian Twitter discourse on the severe acute respiratory syndrome coronavirus 2 pandemic. *Cyberpsychol Behav Soc Netw*. Feb 2021;24(2):117-122. [doi: [10.1089/cyber.2020.0341](https://doi.org/10.1089/cyber.2020.0341)] [Medline: [32857607](https://pubmed.ncbi.nlm.nih.gov/32857607/)]
- Park B, Jang IS, Kwak D. Sentiment analysis of the COVID-19 vaccine perception. *Health Informatics J*. 2024;30(1):14604582241236131. [doi: [10.1177/14604582241236131](https://doi.org/10.1177/14604582241236131)] [Medline: [38403926](https://pubmed.ncbi.nlm.nih.gov/38403926/)]
- Lyu H, Wang J, Wu W, et al. Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. *Intell Med*. Feb 2022;2(1):1-12. [doi: [10.1016/j.imed.2021.08.001](https://doi.org/10.1016/j.imed.2021.08.001)] [Medline: [34457371](https://pubmed.ncbi.nlm.nih.gov/34457371/)]
- Cheng T, Han B, Liu Y. Exploring public sentiment and vaccination uptake of COVID-19 vaccines in England: a spatiotemporal and sociodemographic analysis of Twitter data. *Front Public Health*. Aug 17, 2023;11:1193750. [doi: [10.3389/fpubh.2023.1193750](https://doi.org/10.3389/fpubh.2023.1193750)] [Medline: [37663835](https://pubmed.ncbi.nlm.nih.gov/37663835/)]
- Pierri F, DeVerna MR, Yang KC, Axelrod D, Bryden J, Menczer F. One year of COVID-19 vaccine misinformation on Twitter: longitudinal study. *J Med Internet Res*. Feb 24, 2023;25:e42227. [doi: [10.2196/42227](https://doi.org/10.2196/42227)] [Medline: [36735835](https://pubmed.ncbi.nlm.nih.gov/36735835/)]
- Cinelli M, De Francisci Morales G, Galeazzi A, Quattrocioni W, Starnini M. The echo chamber effect on social media. *Proc Natl Acad Sci U S A*. Mar 2, 2021;118(9):e2023301118. [doi: [10.1073/pnas.2023301118](https://doi.org/10.1073/pnas.2023301118)] [Medline: [33622786](https://pubmed.ncbi.nlm.nih.gov/33622786/)]
- Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An “Infodemic”: leveraging high-volume Twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infect Dis*. Jul 2020;7(7):ofaa258. [doi: [10.1093/ofid/ofaa258](https://doi.org/10.1093/ofid/ofaa258)] [Medline: [33117854](https://pubmed.ncbi.nlm.nih.gov/33117854/)]
- Lanier HD, Diaz MI, Saleh SN, Lehmann CU, Medford RJ. Analyzing COVID-19 disinformation on Twitter using the hashtags #scamdemic and #plandemic: retrospective study. *PLoS One*. Jun 22, 2022;17(6):e0268409. [doi: [10.1371/journal.pone.0268409](https://doi.org/10.1371/journal.pone.0268409)] [Medline: [35731785](https://pubmed.ncbi.nlm.nih.gov/35731785/)]
- Xie T, Ge Y, Xu Q, Chen S. Public awareness and sentiment analysis of COVID-related discussions using BERT-based infoveillance. *AI*. Mar 16, 2023;4(1):333-347. [doi: [10.3390/ai4010016](https://doi.org/10.3390/ai4010016)]
- Scott K, Delobelle P, Berendt B. Measuring shifts in attitudes towards COVID-19 measures in Belgium. *arXiv*. Preprint posted online on Apr 20, 2021. [doi: [10.48550/arXiv.2104.09947](https://doi.org/10.48550/arXiv.2104.09947)]
- Medvedev AN, Lambiotte R, Delvenne JC. The anatomy of Reddit: an overview of academic research. In: Ghanbarnejad F, Roy RS, Karimi F, Delvenne JC, Mitra B, editors. *Dynamics On and Of Complex Networks III*. Springer Nature; 2019:183-204. [doi: [10.1007/978-3-030-14683-2\\_9](https://doi.org/10.1007/978-3-030-14683-2_9)]
- Corradini E. Deconstructing cultural appropriation in online communities: a multilayer network analysis approach. *Inf Process Manag*. May 2024;61(3):103662. [doi: [10.1016/j.ipm.2024.103662](https://doi.org/10.1016/j.ipm.2024.103662)]

23. Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: a call to action for strengthening vaccine confidence. *J Infect Public Health*. Oct 2021;14(10):1505-1512. [doi: [10.1016/j.jiph.2021.08.010](https://doi.org/10.1016/j.jiph.2021.08.010)] [Medline: [34426095](https://pubmed.ncbi.nlm.nih.gov/34426095/)]
24. Liu Y, Whitfield C, Zhang T, Hauser A, Reynolds T, Anwar M. Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. *Health Inf Sci Syst*. Jun 25, 2021;9(1):25. [doi: [10.1007/s13755-021-00158-4](https://doi.org/10.1007/s13755-021-00158-4)] [Medline: [34188896](https://pubmed.ncbi.nlm.nih.gov/34188896/)]
25. Yurtsever MME, Shiraz M, Ekinici E, Eken S. Comparing COVID-19 vaccine passports attitudes across countries by analysing Reddit comments. *J Inf Sci*. Aug 2025;51(4):831-842. [doi: [10.1177/01655515221148356](https://doi.org/10.1177/01655515221148356)]
26. Yan C, Law M, Nguyen S, Cheung J, Kong J. Comparing public sentiment toward COVID-19 vaccines across Canadian cities: analysis of comments on Reddit. *J Med Internet Res*. Sep 24, 2021;23(9):e32685. [doi: [10.2196/32685](https://doi.org/10.2196/32685)] [Medline: [34519654](https://pubmed.ncbi.nlm.nih.gov/34519654/)]
27. Basile V, Cauteruccio F, Terracina G. How dramatic events can affect emotionality in social posting: the impact of COVID-19 on Reddit. *Future Internet*. Jan 27, 2021;13(2):29. [doi: [10.3390/fi13020029](https://doi.org/10.3390/fi13020029)]
28. Hu M, Conway M. Perspectives of the COVID-19 pandemic on Reddit: comparative natural language processing study of the United States, the United Kingdom, Canada, and Australia. *JMIR Infodemiology*. 2022;2(2):e36941. [doi: [10.2196/36941](https://doi.org/10.2196/36941)] [Medline: [36196144](https://pubmed.ncbi.nlm.nih.gov/36196144/)]
29. Valensise CM, Cinelli M, Quattrocioni W. The drivers of online polarization: fitting models to data. *Inf Sci (Ny)*. Sep 2023;642(C):119152. [doi: [10.1016/j.ins.2023.119152](https://doi.org/10.1016/j.ins.2023.119152)]
30. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci U S A*. Jun 17, 2014;111(24):8788-8790. [doi: [10.1073/pnas.1320040111](https://doi.org/10.1073/pnas.1320040111)] [Medline: [24889601](https://pubmed.ncbi.nlm.nih.gov/24889601/)]
31. De Francisci Morales G, Monti C, Starnini M. No echo in the chambers of political interactions on Reddit. *Sci Rep*. Feb 2, 2021;11(1):2818. [doi: [10.1038/s41598-021-81531-x](https://doi.org/10.1038/s41598-021-81531-x)] [Medline: [33531520](https://pubmed.ncbi.nlm.nih.gov/33531520/)]
32. Kozitsin IV. A general framework to link theory and empirics in opinion formation models. *Sci Rep*. Apr 1, 2022;12(1):5543. [doi: [10.1038/s41598-022-09468-3](https://doi.org/10.1038/s41598-022-09468-3)] [Medline: [35365685](https://pubmed.ncbi.nlm.nih.gov/35365685/)]
33. Guilbeault D, Becker J, Centola D. Complex contagions: a decade in review. In: Lehmann S, Ahn YY, editors. *Complex Spreading Phenomena in Social Systems*. Springer, Cham; 2018:3-25. [doi: [10.1007/978-3-319-77332-2\\_1](https://doi.org/10.1007/978-3-319-77332-2_1)]
34. State B, Adamic L. The diffusion of support in an online social movement: evidence from the adoption of equal-sign profile pictures. Presented at: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15; Mar 14-18, 2015:1741-1750; Vancouver, British Columbia, Canada. [doi: [10.1145/2675133.2675290](https://doi.org/10.1145/2675133.2675290)]
35. Baumann F, Lorenz-Spreen P, Sokolov IM, Starnini M. Modeling echo chambers and polarization dynamics in social networks. *Phys Rev Lett*. Jan 31, 2020;124(4):048301. [doi: [10.1103/PhysRevLett.124.048301](https://doi.org/10.1103/PhysRevLett.124.048301)] [Medline: [32058741](https://pubmed.ncbi.nlm.nih.gov/32058741/)]
36. Haensch A, Dragovic N, Borgers C, Boghosian B. A geospatial bounded confidence model including mega-influencers with an application to Covid-19 vaccine hesitancy. *JASSS*. Jan 31, 2023;26(1):8. [doi: [10.18564/jasss.5027](https://doi.org/10.18564/jasss.5027)]
37. Zeng R, Chang X, Liu B. Evolutionary modeling and analysis of opinion exchange and epidemic spread among individuals. *Front Phys*. Nov 2024;12:1501807. [doi: [10.3389/fphy.2024.1501807](https://doi.org/10.3389/fphy.2024.1501807)]
38. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift Reddit dataset. *ICWSM*. May 26, 2020;14(1):830-839. [doi: [10.1609/icwsml.v14i1.7347](https://doi.org/10.1609/icwsml.v14i1.7347)]
39. Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. *SIAM Rev*. Nov 4, 2009;51(4):661-703. [doi: [10.1137/070710111](https://doi.org/10.1137/070710111)]
40. Thukral S, Meisheri H, Kataria T, et al. Analyzing behavioral trends in community driven discussion platforms like Reddit. Presented at: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); Aug 28-31, 2018:662-669; Barcelona, Spain. [doi: [10.1109/ASONAM.2018.8508687](https://doi.org/10.1109/ASONAM.2018.8508687)]
41. Papariello L. Xlm-roberta-base-language-detection. Hugging Face. URL: <https://huggingface.co/papluca/xlm-roberta-base-language-detection/tree/be68b4f03c79c99170af209e5648a5a2a2dba5c8> [Accessed 2026-06-26]
42. Loureiro D, Barbieri F, Neves L, Espinosa Anke L, Camacho-Collados J. TimeLMs: diachronic language models from twitter. Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations; May 22-27, 2022:251-260; Dublin, Ireland. [doi: [10.18653/v1/2022.acl-demo.25](https://doi.org/10.18653/v1/2022.acl-demo.25)]
43. Bonifazi G, Cauteruccio F, Corradini E, et al. A space-time framework for sentiment scope analysis in social media. *BDCC*. Nov 2, 2022;6(4):130. [doi: [10.3390/bdcc6040130](https://doi.org/10.3390/bdcc6040130)]
44. Huang X, Tang X. Analyzing replies and interactions among users with different stances: a case study of the Russia-Ukraine conflict. In: Tang X, Huynh VN, Xia H, Bai Q, editors. *Knowledge and Systems Sciences*. Springer, Singapore; 2025:109-123. [doi: [10.1007/978-981-96-0178-3\\_8](https://doi.org/10.1007/978-981-96-0178-3_8)]

45. Alleman TW, Vergeynst J, De Visscher L, et al. Assessing the effects of non-pharmaceutical interventions on SARS-CoV-2 transmission in Belgium by means of an extended SEIQRD model and public mobility data. *Epidemics*. Dec 2021;37:100505. [doi: [10.1016/j.epidem.2021.100505](https://doi.org/10.1016/j.epidem.2021.100505)] [Medline: [34649183](https://pubmed.ncbi.nlm.nih.gov/34649183/)]
46. COVID-19 pandemic overview. Epistat by Sciensano. 2024. URL: <https://epistat.sciensano.be/covid/> [Accessed 2024-08-11]
47. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. Feb 1, 2017;46(1):348-355. [doi: [10.1093/ije/dyw098](https://doi.org/10.1093/ije/dyw098)] [Medline: [27283160](https://pubmed.ncbi.nlm.nih.gov/27283160/)]
48. Hegselmann R, Krause U. Opinion dynamics and bounded confidence models, analysis, and simulation. *J Artif Soc Soc Simul*. 2002;5(3). URL: <https://jasss.soc.surrey.ac.uk/5/3/2/2.pdf> [Accessed 2026-07-14]
49. Brooks HZ, Chodrow PS, Porter MA. Emergence of polarization in a sigmoidal bounded-confidence model of opinion dynamics. *SIAM J Appl Dyn Syst*. Jun 30, 2024;23(2):1442-1470. [doi: [10.1137/22M1527258](https://doi.org/10.1137/22M1527258)]
50. Degroot MH. Reaching a consensus. *J Am Stat Assoc*. Mar 1974;69(345):118-121. [doi: [10.1080/01621459.1974.10480137](https://doi.org/10.1080/01621459.1974.10480137)]
51. Ge H, Xu K, Ghahramani Z. Turing: a language for flexible probabilistic inference. Presented at: Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics; Apr 9-11, 2018:1682-1690; Playa Blanca, Lanzarote, Canary Islands. URL: <https://proceedings.mlr.press/v84/ge18b.html> [Accessed 2026-06-26]
52. Hoffman MD, Gelman A. The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res*. Apr 14, 2014;15(47):1593-1623. URL: <https://jmlr.org/papers/volume15/hoffman14a/hoffman14a.pdf> [Accessed 2026-06-25]
53. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Opper M, editor. *J Mach Learn Res*. 2010;11(116):3571-3594. URL: <https://jmlr.org/papers/v11/watanabe10a.html> [Accessed 2026-06-29]
54. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statist*. Mar 1947;18(1):50-60. [doi: [10.1214/aoms/1177730491](https://doi.org/10.1214/aoms/1177730491)]
55. Hamed KH, Ramachandra Rao A. A modified Mann-Kendall trend test for autocorrelated data. *J Hydrol (Amst)*. Jan 1998;204(1-4):182-196. [doi: [10.1016/S0022-1694\(97\)00125-X](https://doi.org/10.1016/S0022-1694(97)00125-X)]
56. Federale Overheidsdienst Volksgezondheid, Veiligheid van de Voedselketen en leefmilieu, wet inzake experimenten op de menselijke persoon, ejustice [Article in Dutch]. Federale Overheidsdienst Justitie. May 2004. URL: [https://www.ejustice.just.fgov.be/cgi/article\\_body.pl?language=nl&caller=summary&pub\\_date=04-05-18&numac=2004022376](https://www.ejustice.just.fgov.be/cgi/article_body.pl?language=nl&caller=summary&pub_date=04-05-18&numac=2004022376) [Accessed 2026-03-21]
57. Public content policy. Reddit. URL: <https://support.reddithelp.com/hc/en-us/articles/26410290525844-Public-Content-Policy> [Accessed 2026-03-21]
58. Franzke AS, Bechmann M, Zimmer C, Ess CM. Internet research: ethical guidelines 3.0. Association of Internet Researchers; 2020. URL: <https://aoir.org/reports/ethics3.pdf> [Accessed 2026-06-26]
59. Mancini A, Desiderio A, Di Clemente R, Cimini G. Self-induced consensus of Reddit users to characterise the GameStop short squeeze. *Sci Rep*. Aug 12, 2022;12(1):13780. [doi: [10.1038/s41598-022-17925-2](https://doi.org/10.1038/s41598-022-17925-2)] [Medline: [35962174](https://pubmed.ncbi.nlm.nih.gov/35962174/)]
60. Galindo G. 30% of belgians not in favour of getting coronavirus vaccine. *The Brussels Times*. URL: <https://www.brusselstimes.com/129760/30-of-belgians-not-in-favour-of-getting-coronavirus-vaccine-survey> [Accessed 2024-12-16]
61. Kobayashi R, Takedomi Y, Nakayama Y, et al. Evolution of public opinion on COVID-19 vaccination in Japan: large-scale Twitter data analysis. *J Med Internet Res*. Dec 22, 2022;24(12):e41928. [doi: [10.2196/41928](https://doi.org/10.2196/41928)] [Medline: [36343186](https://pubmed.ncbi.nlm.nih.gov/36343186/)]
62. Gozzi N, Tizzani M, Starnini M, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: mixed-methods analysis. *J Med Internet Res*. Oct 12, 2020;22(10):e21597. [doi: [10.2196/21597](https://doi.org/10.2196/21597)] [Medline: [32960775](https://pubmed.ncbi.nlm.nih.gov/32960775/)]
63. Ancona C, Iudice FL, Garofalo F, De Lellis P. A model-based opinion dynamics approach to tackle vaccine hesitancy. *Sci Rep*. Jul 12, 2022;12(1):11835. [doi: [10.1038/s41598-022-15082-0](https://doi.org/10.1038/s41598-022-15082-0)] [Medline: [35821508](https://pubmed.ncbi.nlm.nih.gov/35821508/)]
64. Bedson J, Skrip LA, Pedi D, et al. A review and agenda for integrated disease models including social and behavioural factors. *Nat Hum Behav*. Jul 2021;5(7):834-846. [doi: [10.1038/s41562-021-01136-2](https://doi.org/10.1038/s41562-021-01136-2)] [Medline: [34183799](https://pubmed.ncbi.nlm.nih.gov/34183799/)]
65. Hale T, Angrist N, Goldszmidt R, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav*. Apr 2021;5(4):529-538. [doi: [10.1038/s41562-021-01079-8](https://doi.org/10.1038/s41562-021-01079-8)] [Medline: [33686204](https://pubmed.ncbi.nlm.nih.gov/33686204/)]
66. Alipour S, Galeazzi A, Sangiorgio E, et al. Cross-platform social dynamics: an analysis of ChatGPT and COVID-19 vaccine conversations. *Sci Rep*. Feb 2, 2024;14(1):2789. [doi: [10.1038/s41598-024-53124-x](https://doi.org/10.1038/s41598-024-53124-x)] [Medline: [38307909](https://pubmed.ncbi.nlm.nih.gov/38307909/)]
67. Covid-reddit-belgium. GitHub. URL: <https://github.com/TimVWese/covid-reddit-belgium> [Accessed 2026-06-26]
68. Subreddit comments/submissions 2005-06 to 2023-12. Academic Torrents. URL: <https://academictorrents.com/details/56aa49f9653ba545f48df2e33679f014d2829c10> [Accessed 2026-06-26]

---

**Abbreviations**

**BERT:** bidirectional encoder representations from transformers

**RoBERTa:** robustly optimized BERT pretraining approach

**SLEBC:** smooth latent-expressed bounded confidence

**WAIC:** Watanabe-Akaike information criterion

---

*Edited by Amaryllis Mavragani; peer-reviewed by Jing Kersey, Simon Zollo, Zhao Liu; submitted 14.Nov.2025; final revised version received 11.May.2026; accepted 03.Jun.2026; published 29.Jul.2026*

*Please cite as:*

Van Wesemael T, Rocha LEC, Alleman TW, Baetens JM

Social Contagion in COVID-19 Discussions Within the Belgian Reddit Community: Statistical and Modeling Study

J Med Internet Res 2026;28:e87723

URL: <https://www.jmir.org/2026/1/e87723>

doi: [10.2196/87723](https://doi.org/10.2196/87723)

© Tim Van Wesemael, Luis Enrique Correa Rocha, Tijs W Alleman, Jan M Baetens. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 29.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.