Emodynamics: Detecting and Characterizing Pandemic Sentiment Change Points on Danish Twitter

Rebekah Baglini[1], Sara Møller Østergaard[2], Stine Nyhus Larsen[2] and Kristoffer Nielbo[2]

[1] *School of Communication and Culture - Linguistics, Cognitive Science, and Semiotics, Aarhus University, Jens Chr. Skous Vej 2, Building 1485, DK-8000 Aarhus C*

[2] *Center for Humanities Computing Aarhus, Aarhus University, Jens Chr. Skous Vej 4, Building 1483,DK-8000 Aarhus C*

Abstract

In this paper, we present the results of an initial experiment using emotion classifications as the basis for studying information dynamics in social media ('emodynamics'). To do this, we used Bert Emotion [18] to assign probability scores for eight different emotions to each text in a time series of 43 million Danish tweets from 2019-2022. We find that variance in the information signals novelty and resonance reliably identify seasonal shifts in posting behavior, particularly around the Christmas holiday season, whereas variance in the distribution of emotion scores corresponds to more local events such as major inflection points in the Covid-19 pandemic in Denmark. This work in progress suggests that emotion scores are a useful tool for diagnosing shifts in the baseline information state of social media platforms such as Twitter, and for understanding how social media systems respond to both predictable and unexpected external events.

Keywords

Change point detection, Information theory, Social media, Covid-19

1. Introduction

The Covid-19 pandemic saw unprecedented activity on social media, as people's social networks became limited to the virtual sphere. During this time, Twitter saw record user activity on its platform, including in Denmark where Tweet activity spiked during the first lockdown period (March-April 2020) and has remained high since (Figure 1). In contrast to other social media platforms, engagement on Twitter is primarily driven by informational needs and desire to engage with and react to news in real time [8, 6, 17]. From the perspective of cultural dynamics, the COVID-19 pandemic provides a natural experiment that allows us to study the effect of a global catastrophe on the informational and emotional dynamics of social media, at some level reflecting the affective experience of a wide socio-cultural and political user spectrum. As

CHR 2022: Computational Humanities Research Conference, December 12 – 14, 2022, Antwerp, Belgium
rbkh@cc.au.dk (R. Baglini); smoe@cas.au.dk (S. M. Østergaard); 201808306@post.au.dk (S. N. Larsen); kln@cas.au.dk (K. Nielbo)
0000-0002-2836-5867 (R. Baglini); 0000-0002-0572-6391 (S. M. Østergaard); 0000-0000-0000-0000 (S. N. Larsen); 0000-0002-5116-5070 (K. Nielbo)

such, social media content during the pandemic functions as a proxy for how cultural information systems respond to unexpected external events. We explore the effect of Covid-19 on the dynamics of Danish Twitter by using methods derived from prior work on information dynamics which apply windowed relative entropy to unstructured texts in a time series. Specifically, we extract information signals of novelty and resonance from 2019 to the present based on emotion classifications of the content of Danish language tweets using BERT Emotion across eight categories [18], a method chosen to reflect the more affective and emotion-driven nature of short-format social media texts [23, 2].

1.1. Information dynamics

In line with developments in information theory, recent studies have used information-theoretic models to track the states and dynamics of socio-cultural systems as reflected in lexical data [9, 1, 5, 16, 10]. Both Shannon entropy and relative entropy have been used to detect changes in prevalent mental states due to the socio-cultural context (e.g., state censorship, degree of recognition, religious observation) [16, 12]. One specific information-theoretic approach applies windowed relative entropy to dense low-dimensional text representations to generate signals that capture information *novelty* as a reliable content difference from the past and *resonance* as the degree to which future information conforms to said novelty [1, 10]. Taking a more dynamic perspective on this approach, one recent study has shown that discussion boards on social media where the novelty signal displays both short-range correlations only and a particularly strong association with resonance are more likely to contain trending content [11]. Using the same approach, but combined with event detection, has also been shown to reliably predict major change points in historical data [25].

Previously, information dynamics in newspapers during the first phase of COVID-19 have been used to examine national response strategies to the pandemic in Denmark and Sweden [14]. A peculiar behavior could be observed in news media when the first wave of COVID-19 virus spread across the world. In response to the pandemic event, the ordinary rate of change in news content was disrupted because nearly every story became associated with COVID-19. On the one hand, content novelty went down, because nearly every story became more similar to previous stories (i.e., news suddenly became 'Corona news'), but on the other hand, the COVID-19 association became more prevalent, resulting in, at least initially, an increase in content persistence. A recent study, [13] argues that this behavior is an example of the *news information decoupling* (NID) principle, according to which information dynamics of news media are (initially) decoupled by temporally extended catastrophes such that the content novelty decreases as media focus monotonically on the catastrophic event, but the resonant property of said content increases as its continued relevance propagate throughout the news information system. The same study further indicated that NID can be used to detect significant change in news media that originate in catastrophic events. We wish to explore whether the emotional novelty and resonance shows a similar dynamics on social media during the same period.

1.2. From infodynamics to emodynamics

Prior studies of information dynamics in media have used lexical co-occurrence as the basis for extracting information signals. This paper investigates whether the emotional character of social media texts can similarly capture event-related inflection points using windowed relative entropy. There are two motivating factors behind this choice. First, sentiment analysis and emotional classification are commonly used methods to characterize different states on social media [23], due to the more pronounced emotional valence found in social media texts compared to e.g. newspaper articles [26, 2]. Second, LDA topic models—commonly used as the latent variables in measuring the information dynamics of texts—are difficult to apply to ultra-short texts such as tweets, without aggregating individual tweets into larger chunks or threads [19]. Finally, representing texts as probability distributions over emotional categories allows us to measure the emotional and affective impact of events in the social media sphere—that is, how users are reacting to and feeling about news and events in real time.

2. Methods

The dataset consists of 43,555,069 million Danish tweets (excluding retweets) from January 1 2019-August 16 2022 collected using the Twitter API V2 (academic track) (see Fig. 1) . These tweets were queried using the most common Danish, Swedish and Norwegian words from the Opensubtitles word frequency lists containing 50.000 words per language [7] (see Appendix A.1). The word lists were adjusted to remove words non-specific to Scandinavian languages, and the 100 highest frequency unique words from each list were then combined and used as queries. The Danish subset of the collection was then extracted using Twitter's native language classifier (which was found to be slightly more accurate than any of the language detection libraries for Python we compared against). Note that this sampling method, being based on Danish language queries, will not include data from multilingual Danes or non-Danish-speaking expats and immigrants, and therefore gives an extensive but not comprehensive representation of the daily discourse on the Danish Twittersphere.

2.1. Emotion classification

To obtain emotion classifications of the tweets, we used the Danish BERT Emotion model [18]. This model includes eight emotion categories (see Table 1) and outputs a predictive distribution over the emotion categories. Each tweet was then assigned a probability distribution across the eight emotion categories (Table 1) using DaNLP's pretrained BERT emotion models, finetuned on Danmarks Radio (DR) Facebook data using the Transformers library from HuggingFace, and based on pretrained Danish BERT representations by BotXO. The model classifying amongst eight emotions achieves an accuracy on 0.65 and a macro-f1 on 0.64 on the social media test set from DR's Facebook dataset containing 999 examples. By running the Danish Bert Emotion model on our twitter dataset, we generate a predictive emotion distribution for each tweet which serves as the document representation for the extraction of information signals.

We summarize these probability distributions by averaging the probability of each emotion over one day, thus giving us a mean daily probability for each emotion. Figure 2 shows mean

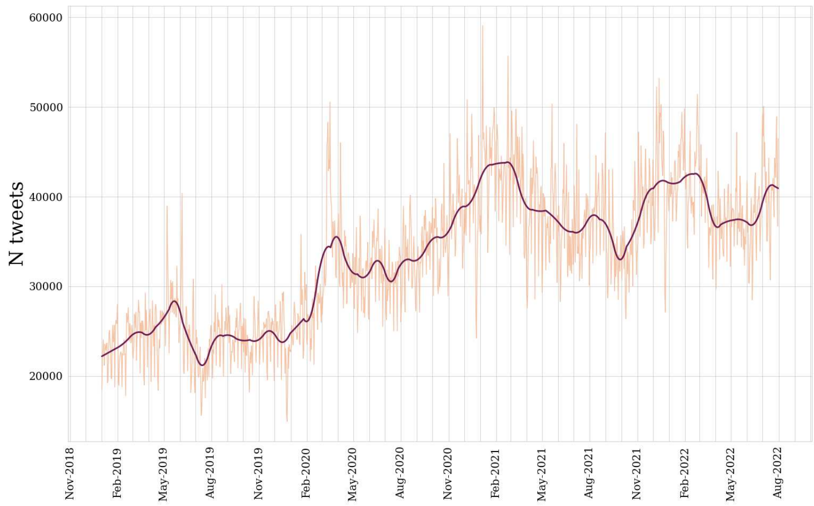

Figure 1: Daily count of Danish tweets from Jan 1 2019-Aug 16 2022.

Table 1
The eight emotion categories in the Danish BERT Emotion model together with our English translations.

Danish Labels	Translations
Glæde/Sindsro	Happiness/Calmness
Tillid/Accept	Trust/Acceptance
Forventning/Interrese	Expectation/Interest
Overasket/Målløs	Surprised/Speechless
Vrede/Irritation	Anger/Irritation
Foragt/Modvilje	Contempt/Reluctance
Sorg/Trist	Grief/Sadness
Frygt/Bekymret	Fear/Worry

daily emotion distributions as time series signals.

2.2. Windowed relative entropy

The summarized daily probability distributions of emotion scores are then used to generate signals that capture information *novelty* as a reliable content difference from the past and *resonance* as the degree to which future information conforms to said novelty, [1, 10], with the latent variables being the emotion distribution.

We used Jensen-Shannon divergence (JSD) to quantify the amount of surprise between two probability distributions. The advantage of JSD over the closely related Kullback-Leibler divergence (KLD) is that it is symmetrical and smoothed, making it a distance metric [20]. JSD is calculated as

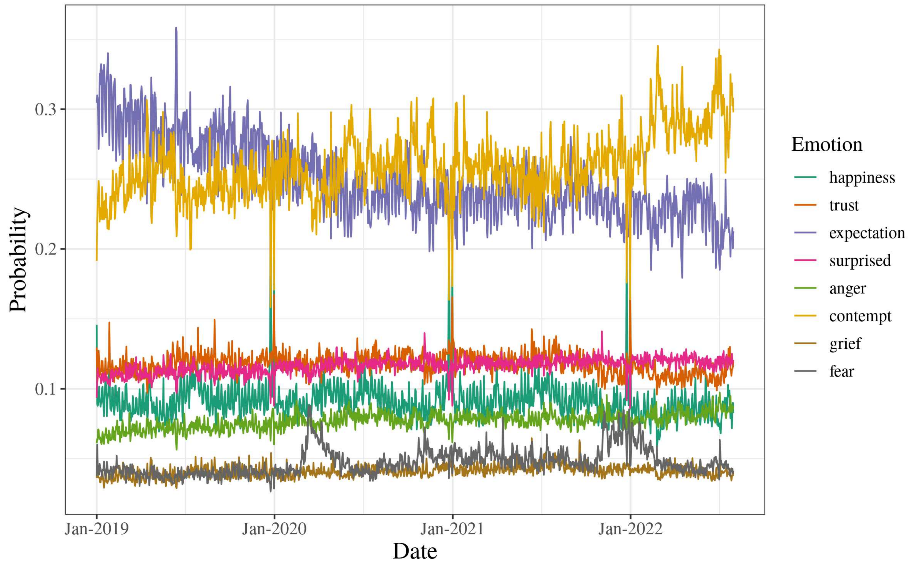

Figure 2: Mean daily value for each of the summarized emotion categories.

$$JSD(s^{(j)}|s^{(k)}) = \frac{1}{2}D(s^{(j)}|M) + \frac{1}{2}D(s^{(k)}|M)$$

Here, $s^{(j)}$ is the probability distribution at the j'th day (and similarly for $s^{(k)}$), $M = \frac{1}{2}(s^{(j)} + s^{(k)})$, and D is KLD, which is defined as

$$D(s^{(j)}|s^{k}) = \sum_{i=1}^{K} s_i^{(j)} log \frac{s_i^{(j)}}{s_i^{(j)}}$$

K corresponds to the number of labels in the probability distribution $s^{(j)}$.

The emotion probability distributions of the BERT models were used as latent variables for the information dynamics measures *novelty*, *transience*, and *resonance*. These measures were calculated following previous definitions [1, 15]. Novelty of the j'th distribution was calculated as

$$N_w(j) = \frac{1}{w} \sum_{d=1}^{w} JSD\left(s^{(j)}|s^{(j-d)}\right)$$

Here, w is the window size. Novelty of the probability distribution of a given day is thus the mean of the entropy between that distribution and the w previous distributions.

Similarly, transience for the jth distribution was calculated as

$$T_w(j) = \frac{1}{w} \sum_{d=1}^{w} JSD\left(s^{(j)}|s^{(j+d)}\right)$$

Transience of the probability distribution of a given day is thus the mean of entropy between that distribution and the w subsequent distributions. Finally, resonance was calculated as

$$R_w(j) = N_w(j) - T_w(j)$$

2.3. Nonlinear Adaptive Filtering

Nonlinear adaptive filtering is applied to the information signals because of the their inherent noisiness, [4]. First, the signal is partitioned into segments (or windows) of length $w = 2n + 1$ points, where neighboring segments overlap by $n + 1$. The time scale is $n + 1$ points, which ensures symmetry. Then, for each segment, a polynomial of order D is fitted. Note that $D = 0$ means a piece-wise constant, and $D = 1$ a linear fit. The fitted polynomial for *ith* and $(i + 1)th$ is denoted as $y^{(i)}(l_1), y^{(i+1)}(l_2)$, where $l_1, l_2 = 1, 2, ..., 2n + 1$. Note the length of the last segment may be shorter than w. We use the following weights for the overlap of two segments.

$$y^{(c)}(l_1) = w_1 y^{(i)}(l + n) + w_2 y^{(i)}(l), l = 1, 2, \ldots, n + 1 \tag{1}$$

where $w_1 = (1 - \frac{l-1}{n}), w_2 = 1 - w_1$ can be written as $(1 - \frac{d_j}{n}), j = 1, 2$, where d_j denotes the distance between the point of overlapping segments and the center of $y^{(i)}, y^{(i+1)}$. The weights decrease linearly with the distance between point and center of the segment. This ensures that the filter is continuous everywhere, which ensures that non-boundary points are smooth.

A window of three days ($w = 3$) was chosen for the analysis, meaning that resonance for each day was calculated relative to three previous and three following days. The chosen window can be thought of as deciding the granularity of the analysis. The longer the window, the less fine-grained the analysis. In newspapers, a typical cycle is seven days [15], but dynamics change much more quickly on social media [21]. By setting a window of three days, we can capture the main fluctuations in emotional dynamics related to external events.

2.4. Change Point Detection

The search method Pruned Exact Linear Time (PELT) was used to identify the change points. This method not only finds the relevant change points but also determines the number of change points. By using a linear penalization on the number of change points, the PELT algorithm identifies the number of change points while aiming to minimize overfitting [24]. PELT is an optimal search method, meaning that it is guaranteed to find the optimal segmentation of the signal given the cost function and penalization [24].

We used the radial basis function (rbf) as the cost function, which is a cost function based on a Gaussian kernel. The kernel k for rbf is defined as

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \tag{2}$$

Here $\| \cdot \|$ is the Euclidian norm and $\gamma > 0$ is the bandwidth parameter which is defined as the inverse of the median of all pairwise distances [24]. When fitting the model, we used a smoothing parameter $\beta = 4$. A low β would result in an increased segmentation of the signal while a higher value for β would make the algorithm disregard more change points. Thus,

Table 2
Coefficients of the $N \times R$ slopes for each change point (CP) period together with the 95% confidence intervals.

CP period	$N \times R$ slope
1	0.826 [0.693,0.959]
2	0.885 [0.756,1.015]
3	2.215 [1.819,2.610]
4	0.859 [0.750,0.968]
5	2.235 [1.767,2.703]
6	0.973 [0.823,1.122]

setting this parameter can be thought of as a trade-off between complexity and goodness-of-fit for the model. The model was fitted using the `ruptures` python package [24].

2.4.1. Resonance-novelty coupling

To describe the changes in the signal between the different change point periods, we investigated the coupling between resonance and novelty following [15]. This was implemented as a linear regression model predicting resonance from novelty within each of the change point periods,

$$R_i = \beta_0 + \beta_1 N_i + \epsilon_i \tag{3}$$

where R_i and N_i refer to resonance and novelty at the i'th day, respectively. β_0 is the intercept, β_1 the $N \times R$ slope, and ϵ_i the error term. To make the estimate of the $N \times R$ slope more interpretable, both resonance and novelty were z-scored before fitting the model.

3. Results

Resonance and novelty signals using window $w = 3$ calculated from probability distributions of emotions from the Danish BERT Emotion model are visualized in Figure 3. We observe clear and easily detectable tendencies in the signals: rather than decoupled [14], major peaks in novelty and resonance appear to be strongly correlated and spaced at regular intervals corresponding to the Christmas/winter holiday period.

To better observe variance potentially related to Covid-19 pandemic events, Figure 4 shows the unsmoothed resonance time series together with the change points periods and selected events related to COVID-19 from a timeline published by Statens Serum Institut (SSI) [22] and reproduced in Table 3 in Appendix A.2. The changes in variance between the change point periods are visually apparent when inspecting the signal.

Table 2 below shows $N \times R$ slopes from the linear regression models predicting resonance from novelty, with a threshold filter 0.01 applied to both novelty and resonance values. In change point period 1, the estimated coefficient was $\beta_1 = 0.826$ and was the smallest out of all of the four time periods, while the estimate of the $N \times R$ slope in the fifth change point period was the largest with a coefficient of $\beta_1 = 2.235$.

The $N \times R$ slopes of the linear models are visualized in Figure 5. Notice that the scale of both axes differ between change point periods. This is due to variations in the distribution of the

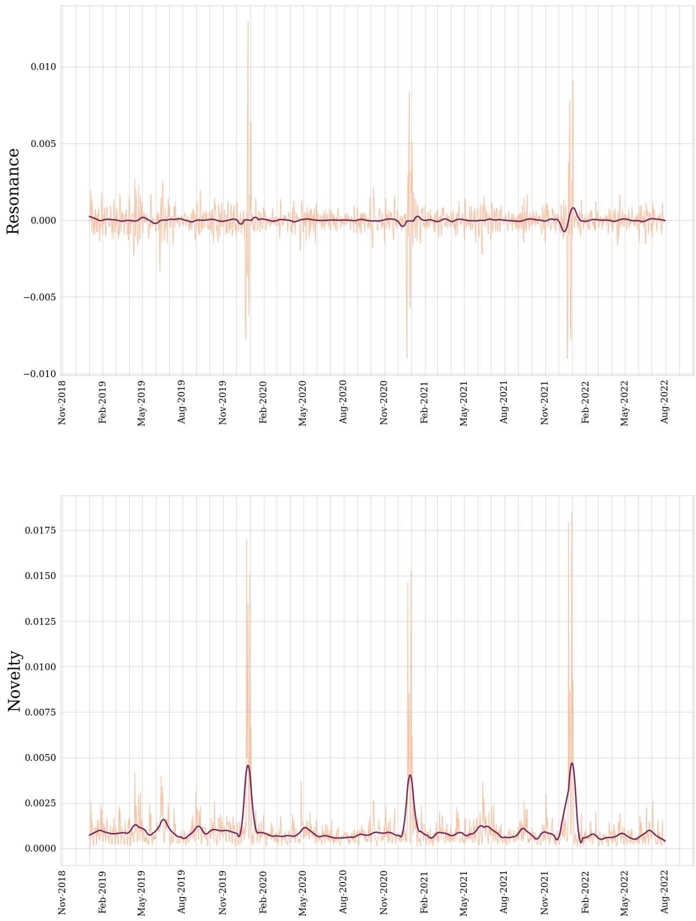

Figure 3: Resonance and novelty calculated with a window $w = 3$. The orange line represents the raw signal, and the black line is the signal smoothed using a non-linear adaptive filter [3].

data points, which makes visualizations using the same scales difficult to interpret. The figure shows a general positive coupling between resonance and novelty. Moreover, it can be seen that $N \times R$ slope in the holiday change point periods 3 and 5 are significantly steeper than all other periods.

Matrices showing the correlation between the individual emotion time series signals in each change point period are in Figure 6. The correlations between the emotions are generally strongest in change point periods 3 and 5, representing the 2020 and 2021 Christmas holidays, respectively. These periods contains very clear clusters: *happiness* and *trust* are positively correlated with each other while negatively correlated with *surprise*, *anger*, *contempt*, and *fear*. The four latter emotions are all positively correlated with each other. All of these observed correlations have a correlation coefficient $r > .5$.

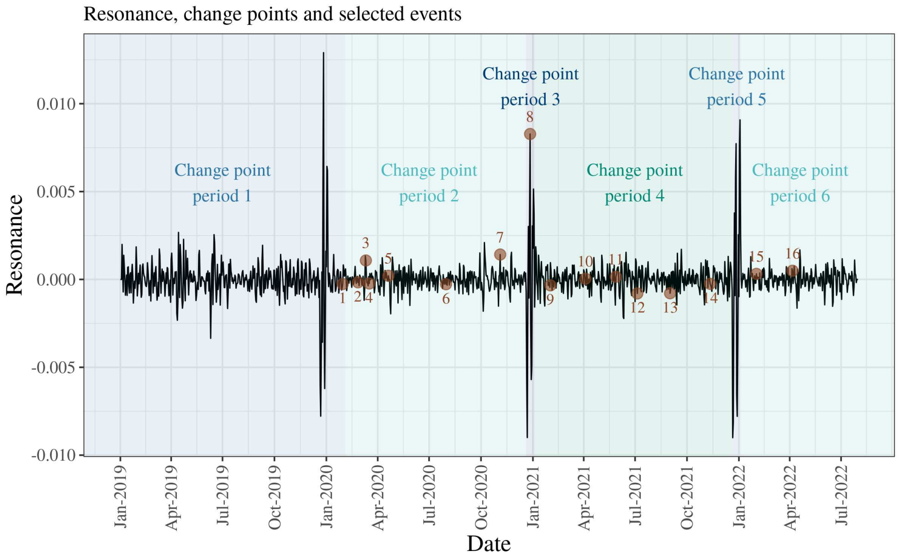

Figure 4: Resonance signal with background colors separating the change point periods. Event descriptions can be found in table 3, and change point start- and end dates in table 3

4. Discussion

Recall that novelty is a measure of the average amount of relative surprise between the probability distribution at a given time point and the probability distributions in a window with the w previous time point (three days, in our case), while transience is comparing the probability distribution at that time point with the w following probability distributions. Resonance is high if novelty is high while transience is low, meaning the documents are very surprising compared to previous documents but not the following. In other words, a stronger correlation between novelty and resonance means that as more information enters the system, more of it "sticks" and remains relevant. Based on the slope of $N \times R$ for 2019 (change point period 1)—representing the pre-pandemic baseline—the normal state of Danish Twitter is one of high emotional entropy: new information is regularly entering the discourse, producing a novel distribution of emotional responses, but resonance is relatively weak (Figure 5). On an annual basis around the Christmas holiday, we see a marked shift to a lower-entropy state where the emotion distribution is more predictable: new information floods in, but resonance is high. This state persists only for a short time, until the holiday ends, people return to work, and the news media cycle returns to normal. Somewhat surprisingly, despite the occurrence of a major catastrophe in early 2020—the onset of the Covid-19 pandemic—our change point detection model does not distinguish this period as abnormal with respect to the baseline emotional dynamics of Twitter. This is because even as the onset of a major shock event floods the system with new information with a high variety of different emotional reactions (novelty) the persistence of these patterns overtime is not significantly different from the normal baseline resonance rate; i.e. the high entropy conditions produced by even a major disaster are not so dissimilar from the normal state of affairs on Twitter, which is a high entropy system. Thus,

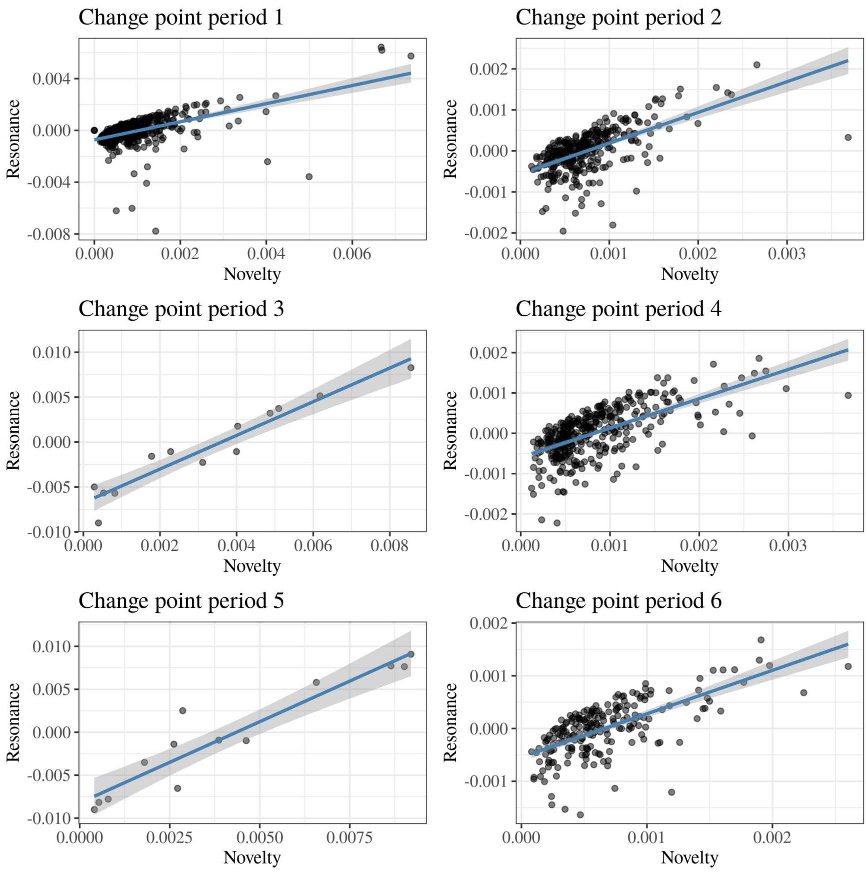

Figure 5: $N \times R$ slopes in the six change point periods. Notice that both axes vary between plots.

change point detection based on emotions are only tuned to discern more systematic seasonal shifts from high-to-low emotional entropy.

To see the fingerprint of Covid-19 events and other unpredictable local events in the time series, we must look to the dynamics of the individual emotion scores (Figure 3). Variations in the dynamics of the different signals can be visually detected, e.g. the pandemic period in Denmark, starting in early 2020, has been marked by an overall shift towards *contempt* being the predominant emotion on Twitter, and a gradual drop-off of *expectation*, while *grief* appears the most stationary as well as the least likely emotion. The figure also depicts sudden changes in values for some of the emotions, for example, the marked spikes in *fear* in late February-March of 2020 and December 2021-February 2022, corresponding to the period of Denmark's first lockdown and the surge of infection due to the omicron variant, respectively (cf. the timeline in Table 3). The same period also shows a decrease in *trust*, which has not recovered to baseline as of August 2022.

In our continuing work, we will experiment with different parameters and change point detection methods which might show higher sensitivity to micro-disruptions of the novelty and

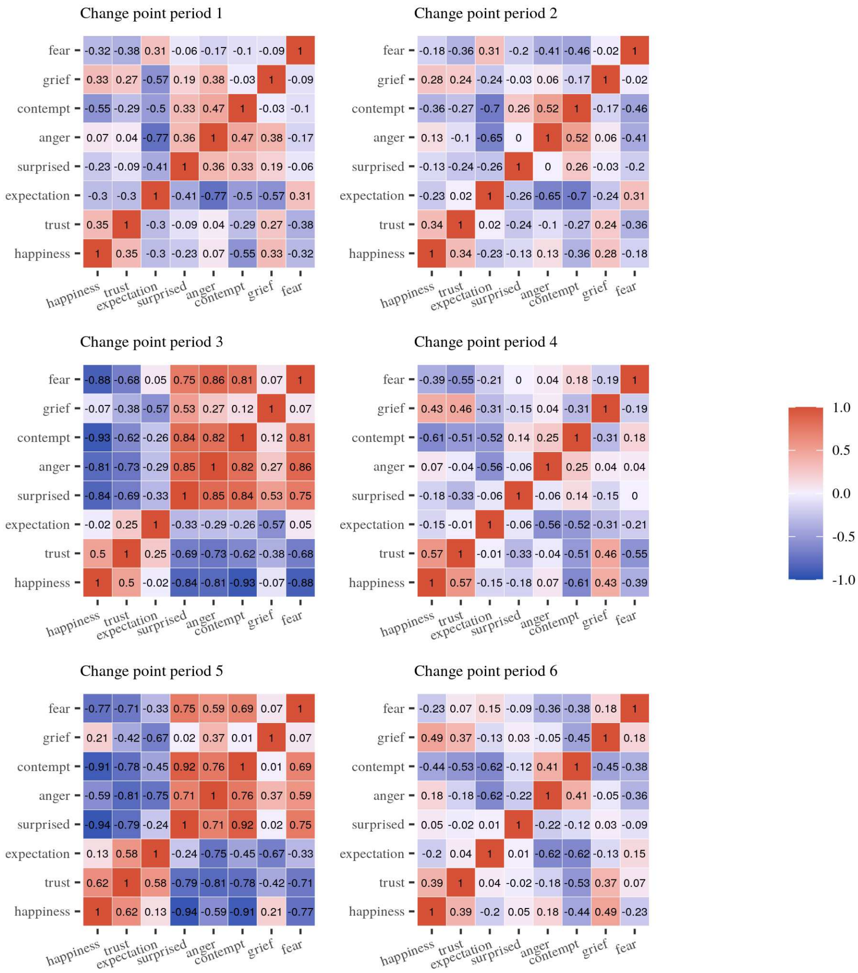

Figure 6: Correlation matrices for the eight emotion time series signals within the periods defined by change points.

resonance signal triggered by external events. We will also experiment with coupling emotion distributions with other representations of document content in extracting information signals.

References

[1] A. T. J. Barron, J. Huang, R. L. Spang, and S. DeDeo. "Individuals, institutions, and innovation in the debates of the French Revolution". In: *Proceedings of the National Academy of Sciences* 115.18 (2018), pp. 4607–4612. DOI: 10.1073/pnas.1717729115. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1717729115.

[2] C. Duncombe. "The Politics of Twitter: Emotions and the Power of Social Media". In: *International Political Sociology* 13.4 (2019), pp. 409–429. DOI: 10.1093/ips/olz013. URL: https://doi.org/10.1093/ips/olz013.

[3] J. Gao, J. Hu, and W.-w. Tung. "Facilitating Joint Chaos and Fractal Analysis of Biosignals through Nonlinear Adaptive Filtering". In: *Plos One* 6.9 (2011), e24331. DOI: 10.1371/journal.pone.0024331. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0024331.

[4] K. L. Gray. "Comparison of Trend Detection Methods". In: *Graduate Student Theses, Dissertations, & Professional Papers* 228 (2007), p. 98.

[5] J. Guldi. "The Measures of Modernity: The New Quantitative Metrics of Historical Change Over Time and Their Critical Interpretation". In: *International Journal for History, Culture and Modernity* 7.1 (2019), pp. 899–939. DOI: 10.18352/hcm.589.

[6] P. R. Johnson and S. Yang. "Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use". In: *Communication Technology Division of the annual convention of the Association for Education in Journalism and Mass Communication in Boston, MA*. Vol. 54. 2009.

[7] P. Lison and J. Tiedemann. "Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles". In: (2016).

[8] A. Masciantonio, D. Bourguignon, P. Bouchat, M. Balty, and B. Rimé. "Don't put all social network sites in one basket: Facebook, Instagram, Twitter, TikTok, and their relations with well-being during the COVID-19 pandemic". In: *PloS one* 16.3 (2021), e0248384.

[9] J. Murdock, C. Allen, and S. DeDeo. "Exploration and Exploitation of Victorian Science in Darwin's Reading Notebooks". In: *arXiv:1509.07175* (2015).

[10] D. Nguyen, M. Liakata, S. DeDeo, J. Eisenstein, D. Mimno, R. Tromble, and J. Winters. "How we do things with words: Analyzing text as social and cultural data". In: *Frontiers in Artificial Intelligence* 3 (2020), p. 62. DOI: 10.3389/frai.2020.00062.

[11] K. Nielbo, P. Vahlstrup, and A. Bechmann. "Trend Reservoir Detection: Minimal Persistence and Resonant Behavior of Trends in Social Media". In: *Proceedings of Computational Humanities Research* 1 (2021).

[12] K. L. Nielbo, K. F. Baunvig, B. Liu, and J. Gao. "A curious case of entropic decay: Persistent complexity in textual cultural heritage". In: *Digital Scholarship in the Humanities* 34.3 (2019). DOI: 10.1093/llc/fqy054.

[13] K. L. Nielbo, R. B. Baglini, P. B. Vahlstrup, K. C. Enevoldsen, A. Bechmann, and A. Roepstorff. "News Information Decoupling: An Information Signature of Catastrophes in Legacy News Media". In: *arXiv:2101.02956 [cs]* (2021).

[14] K. L. Nielbo, K. Enevoldsen, R. Baglini, E. Fano, and J. Gao. "Pandemic news information uncertainty - News dynamics mirror differential response strategies to COVID-19". In: (2021), p. 16.

[15] K. L. Nielbo, F. Haestrup, K. C. Enevoldsen, P. B. Vahlstrup, R. B. Baglini, and A. Roepstorff. “When no news is bad news – Detection of negative events from news media content”. In: *arXiv:2102.06505 [cs]* (2021). URL: http://arxiv.org/abs/2102.06505.

[16] K. L. Nielbo, M. L. Perner, C. P. Larsen, J. Nielsen, and D. Laursen. “Automated Compositional Change Detection in Saxo Grammaticus’ Gesta Danorum”. In: *Dhn.* 2019, pp. 320–332.

[17] C. S. Park. “Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement”. In: *Computers in human behavior* 29.4 (2013), pp. 1641–1648.

[18] A. B. Pauli, M. Barrett, O. Lacroix, and R. Hvingelby. “DaNLP: An open-source toolkit for Danish Natural Language Processing”. In: *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa).* Reykjavik, Iceland (Online): Linköping University Electronic Press, Sweden, 2021, pp. 460–466. URL: https://aclanthology.org/2021.nodalida-main.53.

[19] J. Qiang, Z. Qian, Y. Li, Y. Yuan, and X. Wu. “Short text topic modeling techniques, applications, and performance: a survey”. In: *IEEE Transactions on Knowledge and Data Engineering* 34.3 (2020), pp. 1427–1445.

[20] T. R. d. Reis. *Measuring the statistical similarity between two samples using Jensen-Shannon and Kullback-Leibler.* 2020. URL: https://tinyurl.com/34jhn7xd.

[21] Z. Saeed, R. A. Abbasi, O. Maqbool, A. Sadaf, I. Razzak, A. Daud, N. R. Aljohani, and G. Xu. “What’s Happening Around the World? A Survey and Framework on Event Detection Techniques on Twitter”. In: *Journal of Grid Computing* 17.2 (2019), pp. 279–312. DOI: 10.1007/s10723-019-09482-2. URL: http://link.springer.com/10.1007/s10723-019-09482-2.

[22] Statens Serum Institut. *Tidslinje for covid-19.* 2022. URL: https://www.ssi.dk/-/media/arkiv/subsites/covid19/presse/tidslinje-over-covid-19/covid-19-tidslinje-for-2020-2022-lang-version---version-1---april-2022.pdf?la=da.

[23] M. Thelwall, K. Buckley, and G. Paltoglou. “Sentiment in Twitter events”. In: *Journal of the American Society for Information Science and Technology* 62.2 (2011), pp. 406–418. DOI: 10.1002/asi.21462. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21462.

[24] C. Truong, L. Oudre, and N. Vayatis. “Selective review of offline change point detection methods”. In: *Signal Processing* 167 (2020), p. 107299. DOI: 10.1016/j.sigpro.2019.107299. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165168419303494.

[25] E. Vrangbæk and K. Nielbo. “Composition and Change in De Civitate Dei: A Case Study of Computationally Assisted Methods”. In: *Studia Patristica* (2021).

[26] K. Welbers and M. Opgenhaffen. “Presenting News on Social Media”. In: *Digital Journalism* 7.1 (2019), pp. 45–62. DOI: 10.1080/21670811.2018.1493939. URL: https://doi.org/10.1080/21670811.2018.1493939.

A. Appendix 1

A.1. Scraping keywords

Below follows the list of top words from Danish, Swedish and Norwegian used to scrape Twitter:

> *aften, aldrig, alltid, altid, andet, arbejde, bedste, behöver, behøver, beklager, berätta, betyr, blev, blevet, blir, blitt, blive, bliver, bruge, burde, bättre, båe, bør, deim, deires, ditt, drar, drepe, dykk, dykkar, där, död, döda, død, døde, efter, elsker, endnu, faen, fandt, feil, fikk, finner, flere, forstår, fortelle, fortfarande, fortsatt, fortælle, från, få, fået, får, fått, förlåt, första, försöker, før, først, første, gick, gikk, gillar, gjennom, gjerne, gjorde, gjort, gjør, gjøre, godt, gå, gång, går, göra, gør, gøre, hadde, hallå, havde, hedder, helt, helvete, hende, hendes, hennes, herregud, hjelp, hjelpe, hjem, hjälp, hjå, hjælp, hjælpe, honom, hossen, hvem, hvis, hvordan, hvorfor, händer, här, håll, håller, hør, høre, hører, igjen, ikkje, ingenting, inkje, inte, intet, jeres, jävla, kanske, kanskje, kender, kjenner, korleis, kvarhelst, kveld, kven, kvifor, känner, ledsen, lenger, lidt, livet, längre, låt, låter, længe, meget, menar, mycket, mykje, må, måde, många, mår, måske, måste, måtte, navn, nogen, noget, nogle, noko, nokon, nokor, nokre, någon, något, några, nån, når, nåt, nødt, också, også, pengar, penger, pratar, prøver, på, redan, rundt, rätt, sagde, saker, samma, sammen, selv, selvfølgelig, sidan, sidste, siger, sikker, sikkert, själv, skete, skjedde, skjer, skulle, sluta, slutt, snakke, snakker, snill, snälla, somt, stadig, stanna, sted, står, synes, säger, sätt, så, sådan, såg, sånn, tager, tiden, tilbage, tilbake, tillbaka, titta, trenger, trodde, troede, tror, två, tycker, tänker, uden, undskyld, unnskyld, ursäkta, uten, varför, varit, varte, veldig, venner, verkligen, vidste, vilken, virkelig, visste, väg, väl, väldigt, vän, vår, våra, våre, væk, vær, være, været, älskar, åh, år, åt, över.*

A.2. Timeline of major Covid-19 events in Denmark from SSI

Table 3
Date and translated description for the relevant events from the timeline published by SSI [22]

No.	Period	Date	Description
1	1	2020-01-30	The outbreak of the virus is declared a threat to global health by WHO.
2	2	2020-02-26	The first Danish citizen is tested positive for COVID-19.
3	2	2020-03-11	The Danish prime minister has her second press conference and she announces a two-week lockdown in Denmark. All schools, daycares and institutions are closing. Assembly ban for more than 100 people is introduced. Public employees with a non-critical functionality are sent home.
4	2	2020-03-17	The Queen of Denmark speaks to the public about the COVID-19 crisis.
5	2	2020-04-20	Partial reopening. Driving schools, hair dressers, research laboratories, and certain other liberal professions together with youngest grade levels and outdoor sport activities without body contact is allowed to reopen.
6	2	2020-07-31	Danish health authority recommend wearing face masks in public transportations if there are many people.
7	2	2020-11-04	The government decides to put down all mink on Danish mink farms due to an outbreak of a COVID-19 mutation.
8	3	2020-12-27	The first Danish citizens are vaccinated using the Pfizer/BioNTech vaccine.
9	4	2021-01-28	The lockdown in Denmark, which was introduced in December 2020, is prolonged until February 28, 2021.
10	4	2021-04-14	The AstraZeneca vaccine is withdrawn completely from the Danish vaccination program.
11	4	2021-05-28	The Danish coronapas app can now be downloaded in the App store.
12	4	2021-07-05	The Delta variant now dominates in Denmark. Before this, the alpha variant dominated.
13	4	2021-09-10	Covid-19 is no longer described as a socially critical disease in Denmark.
14	4	2021-11-11	Covid-19 is again a disease critical to society in Denmark.
15	6	2022-02-01	The restrictions are lifted and covid-19 is changed to no longer be a critical illness. Requirements for tests upon entry to Denmark are retained.
16	6	2022-06-03	The rapid test centers close.