When no news is bad news

Detection of negative events from news media content

Kristoffer L. Nielbo, Frida Haestrup, Kenneth C. Enevoldsen, Peter B. Vahlstrup, Rebekah B. Baglini, and Andreas Roepstorff

center for humanities computing aarhus|chcaa.io
aarhus university, denmark

SOLIDUM PETIT IN PROFUNDIS · UNIVERSITAS ARHUSIENSIS ·

CENTER FOR HUMANITIES
COMPUTING AARHUS

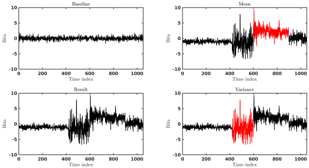
outline

CENTER FOR HUMANITIES
COMPUTING AARHUS

background

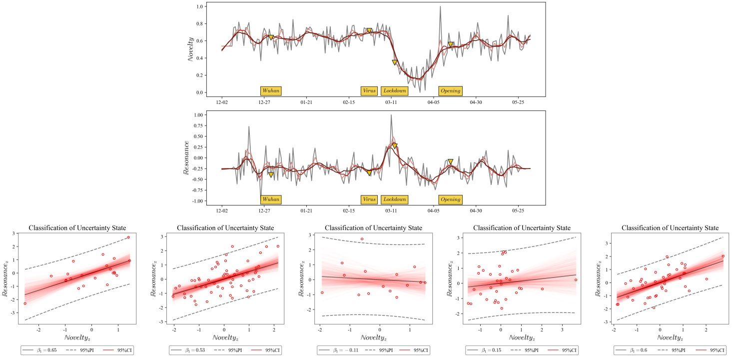
HOPE

how democracies cope with COVID-19 a data-driven approach is an national research project that is part of the (DK) national pandemic monitoring program.



research team interested in cultural dynamics, in particular **how events impact cultural information systems**

> use news media coverage of COVID-19 as a proxy for how cultural information systems respond to **unexpected and dangerous temporally extended events**.

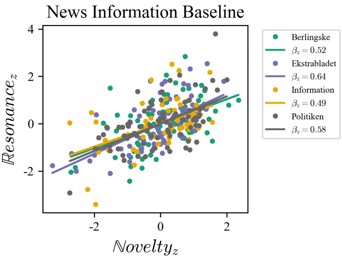


front page information from broadsheet newspaper *politiken* during COVID-19 phase 1.

in response to unexpected and dangerous temporally extended events, the ordinary information dynamics of news media are (initially) decoupled 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**

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

when no news is bad news



$\mathbb{N} \times \mathbb{R}$ baseline models for danish legacy media

- **validate NID observations** with a more formal approach to change detection
- **compare national newspapers**
  - np-type: *broadsheet//tabloid*
  - np-political: *left//right*
- ultimate goal: media monitoring system

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," arXiv:2102.06505 [cs]

data, normalization, and representation

DATA
linguistic content (title and body text) from **front pages of six DK national newspapers** (2x*tabloid*, 4x*broadsheet*).
sampled during COVID-19 phase 1 (december 1, 2019 to july 1 2020)
NORMALIZATION
advertisements and metadata removed
lemmatization, tf-idf weighting, casefolding
REPRESENTATION
bag-of-words model (LDA*) to generate low-rank representations of front pages
variables were estimated for **windows of one week** ($w = 7$).

$\mathbb{N}$: novelty as article $s^{(j)}$'s reliable difference from past articles $s^{(j-1)}, s^{(j-2)}, \ldots, s^{(j-w)}$ in window w:

$$\mathbb{N}_w(j) = \frac{1}{w} \sum_{d=1}^{w} JSD(s^{(j)} \mid s^{(j-d)})$$

$\mathbb{R}$: resonance as the degree to which future articles $s^{(j+1)}, s^{(j+2)}, \ldots, s^{(j+w)}$ conforms to article $s^{(j)}$'s novelty:

$$\mathbb{R}_w(j) = \mathbb{N}_w(j) - \mathbb{T}_w(j)$$

where $\mathbb{T}$ is the transience of $s^{(j)}$:

$$\mathbb{T}_w(j) = \frac{1}{w} \sum_{d=1}^{w} JSD(s^{(j)} \mid s^{(j+d)})$$

we propose a symmetrized and smooth version by using the Jensen–Shannon divergence (JSD):

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

with $M = \frac{1}{2}(s^{(j)} + s^{(k)})$ and D is the Kullback-Leibler divergence:

$$D(s^{(j)} \mid s^{(k)}) = \sum_{i=1}^{K} s_i^{(j)} \times \log_2 \frac{s_i^{(j)}}{s_i^{(k)}}$$

CENTER FOR HUMANITIES COMPUTING AARHUS

Assume two change points, τ_1 and τ_2 and an otherwise stable series that follow a normal distribution with varied mean, μ_i, and singular variance, σ. This gives us the following model given the observed Novelty, $\mathbb{N}_i$:

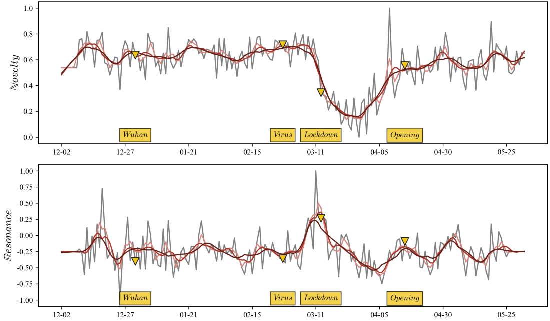
$$\mathbb{N}_t = \begin{cases} \mathcal{N}(\mu_1, \sigma) \text{ for } t < \tau_1 \\ \mathcal{N}(\mu_2, \sigma) \text{ for } \tau_1 \leq t < \tau_2 \\ \mathcal{N}(\mu_3, \sigma) \text{ for } t \geq \tau_2 \end{cases}$$

Estimate the location of τ_i, means μ_i and variance σ, i.e. the following posterior:

$$P(\mu_i, \sigma, \tau_i | \mathbb{N}_t) = P(\mu_1, \mu_2, \mu_3, \sigma, \tau_1, \tau_2 | \mathbb{N}_t)$$

Estimation was carried out with NUTS and the assumptions were modelled using the following priors:

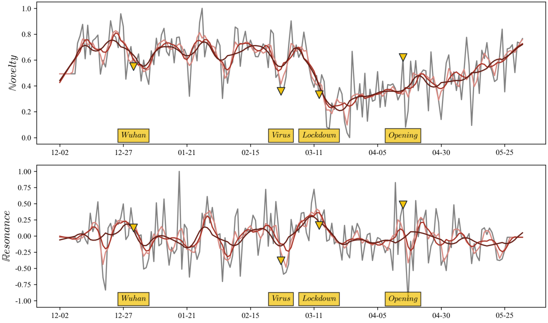
$$\begin{aligned} \mu_i &\sim \mathcal{N}(0, 0.5) \\ \sigma &\sim \text{Half Cauchy}(0.5) \\ \tau_1 &\sim \text{Uniform}(0, \max(\mathbb{N}_t)) \\ \tau_2 &\sim \text{Uniform}(\tau_1, \max(\mathbb{N}_t)) \end{aligned}$$



Novelty (upper panel) and resonance (lower panel) for the center-left newspaper *Politiken* before and during Covid-19 phase 1.

Source	Class	*NID* Start	*NID* End	*NID*
Berlingske	*B*	03.07 [03.03, 03.09]	04.28 [04.09, 05.08]	*True*
BT	*T*	04.10 [12.30, 09.01]	07.25 [04.22, 09.03]	*False*
Ekstrabladet	*T*	01.28 [01.02, 03.17]	05.08 [01.16, 07.22]	*False*
Jyllands-Posten	*B*	03.10 [03.08, 03.14]	05.25 [05.21, 06.06]	*True*
Kristligt Dagblad	*B*	03.07 [03.05, 03.12]	04.15 [04.11, 04.17]	*True*
Politiken	*B*	03.13 [03.12, 03.13]	04.08 [04.05, 04.08]	*True*

Estimated temporal change points at 94% HDIs for novelty. Column one contains the name of the newspaper, columns two its class (*B*roadsheet or *T*abloid).



Novelty (upper panel) and resonance (lower panel) for the center-right newspaper *Berlingske* before and during Covid-19 phase 1.

Source	$\mathbb{N}_{pre}$	$\mathbb{N}_{NID}$	$\mathbb{N}_{post}$
Berlingske	0.36 [0.35, 0.37]	0.29 [0.27, 0.31]	0.34 [0.34, 0.35]
Jyllands-Posten	0.29 [0.28, 0.30]	0.23 [0.22, 0.24]	0.27 [0.26, 0.28]
Kristligt Dagblad	0.27 [0.26, 0.28]	0.19 [0.18, 0.21]	0.26 [0.25, 0.27]
Politiken	0.27 [0.26, 0.28]	0.15 [0.14, 0.17]	0.26 [0.25, 0.26]

Novelty values at 94% HDIs before during and after the lockdown for the four broadsheet newspapers that supported the NID principle..

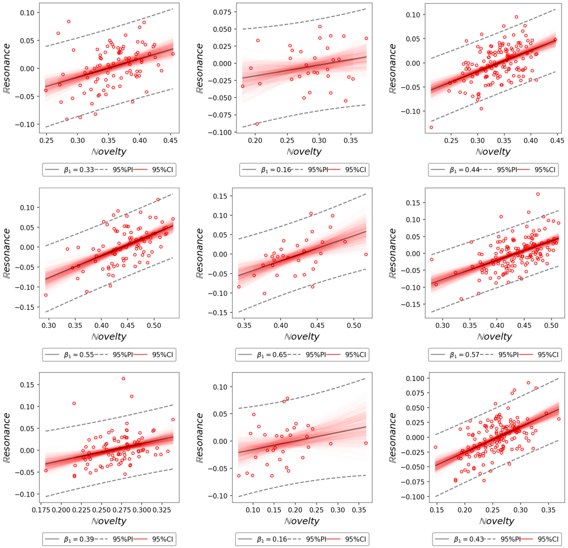


Figure: $\mathbb{N} \times \mathbb{R}$ slopes before during and after the lockdown for Berlingske (upper row), Ekstrabladet (middle row), and Politiken (lower row) during Covid-19 phase 1.

in conclusion...

"Nothing travels faster than the speed of light with the possible exception of bad news, which obeys its own special laws." (D. Adams – Hitchhiker's Guide)

in the case of pandemic information dynamics,

- variation in $\mathbb{N}$*reliably* detected *lockdown* and *opening*
- $\mathbb{N} \times \mathbb{R}$ slopes indicated a decoupling of resonance from novelty during the lockdown
- lockdown interval indicated that lockdown can be be predicted from the first incident
- opening interval may reveal political observation
- tabloids follow different dynamics

no news is bad news, when the lack of novel content persists!

THANKS

kln@cas.au.dk
knielbo.github.io
chcaa.io

SLIDES

knielbo.github.io/files/kln_dhb21.pdf

ACKNOWLEDGEMENTS

This research was supported the "HOPE - How Democracies Cope with COVID-19"-project funded by The Carlsberg Foundation with grant CF20-0044, NeiC's Nordic Digital Humanities Laboratory project, and DeiC Type-1 HPC with project DeiC-AU1-L-000001. The authors would like to thank Berlingske Media, JP/Politkens Hus, and Kristeligt Dagblad for providing access to proprietary data.

CENTER FOR HUMANITIES COMPUTING AARHUS