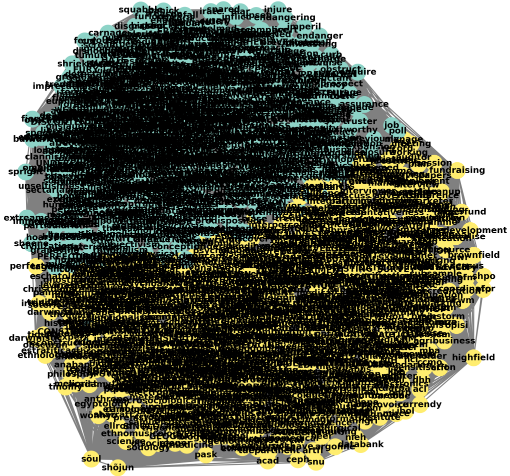
# on the inherent complexity of cultural information

**kristoffer nielbo**
kln@cas.dk
knielbo.github.io

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

CENTER FOR HUMANITIES
COMPUTING AARHUS




## cultural analysis/-tics

- ☑ noisy
- ☑ complex
- ☑ structure
- ☒ time-dependent

CENTER FOR HUMANITIES
COMPUTING AARHUS

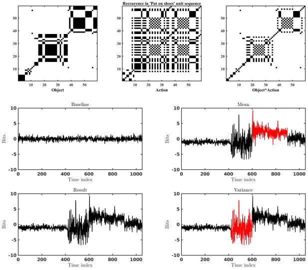
# outline



CENTER FOR HUMANITIES COMPUTING AARHUS

- kristoffer l. nielbo
- studied comparative religion & philosophy → cognitive science & mathematics
- worked in anthropology, history, literature, zoology, psychology, psychiatry, & ++interdisciplinary centers (e.g. IMC, IPAM, CSBD-Tech).

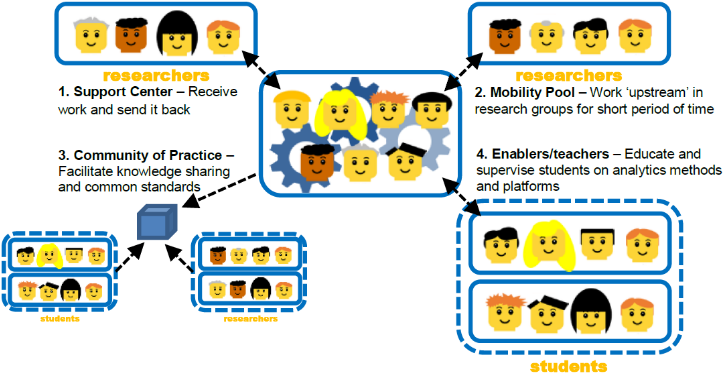
head center for humanities computing, aarhus university denmark




approximate the complexity of a large-scale cultural system by a small set of dynamic variables

- to understand complex system → continuously monitor its states, BUT inefficient to monitor all intrinsic variables of the system
- detailed dynamics of a system (that has an underlying attractor) can be studied by reconstructing a suitable phase space from a scalar time series recorded from the system
- characterize a system by only very few state variables instead of a random system with infinite numbers of degrees of freedom

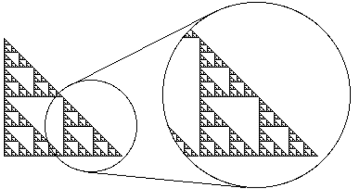
# center for humanities computing aarhus



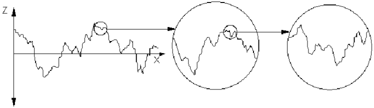

- research and development for SSH at aarhus university
- interactive high performance computing and eScience infrastructure
- group of researchers and RSEs genuinely interested in soft and unstructured data
- almost always looking for new colleagues and collaborators!

CENTER FOR HUMANITIES COMPUTING AARHUS

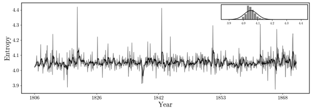
# self-affinity in cultural information



many complex dynamic systems are fractal $\rightarrow$ they display self-similar and scale-invariant behavior




fluctuation patterns at longer time-scales $\sim$ shorter time scales & measurement does not depend on the temporal resolution




a irregular information system represented by time-varying entropy

$$h = -\sum_{i=1}^{K} p_i \times \log_2(p_i)$$

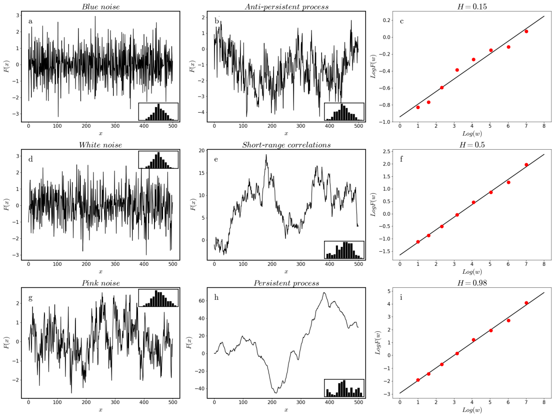
$$p_i = Fr(w_i) / \sum_{i=1}^{K} Fr(w_i)$$

.

- investigate relationship between the measurement and time scale
- display power-law scaling?
- detect changes in relationship as a function of external or internal events
- compare $n+1$ system dynamics on this relationship

CENTER FOR HUMANITIES COMPUTING AARHUS

introduction
whoami
CHCAA
self-affinity in cultural information
self-affinity
fractal analysis
author profiling
consumer history
computational narratology
trend reservoirs
author change points
pandemic information dynamics
information dynamics
news-media baseline
left-omnibus
change-point detection
right-omnibus
predictive model
affective dynamics
summary

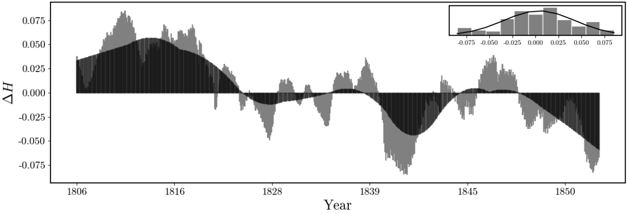



– **construct a random walk** $u(n) = \sum_{k=1}^{n}(x_k - \overline{x}), \quad n = 1, 2, \cdots, N,$
– divide the random walk process into **non-overlapping segments**
– determine the **local trends** of each segment as the best polynomial fit
– determine the average variance over all the segments and residual $u(i) - v(i)$ of the fit is fluctuations around global trend and its variance is the **Hurst parameter** $(H)$

$\Rightarrow$ $H$ **quantifies persistence** in time series: $0 < H < 0.5$ is an anti-persistent process, $H = 0.5$ is a short-memory process, and $0.5 < H < 1$ is a persistent process

---

Nielbo, K.L., Baunvig, K.F., Liu, B. & Gao, J. (2018). A Curious Case of Entropic Decay: Persistent Complexity in Textual Cultural Heritage, DSH.

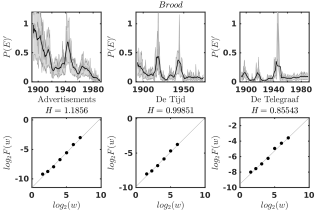


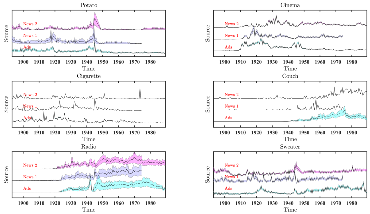

time-varying hurst difference parameter for n.f.s. grundtvig

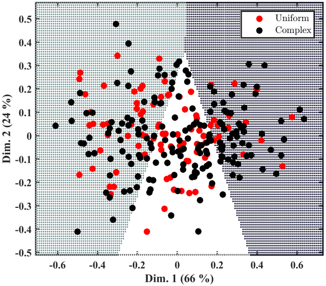
| Time period | Age of onset | Coarse | Fine | Behavior |
|---|---|---|---|---|
| 1806-1826 | 23 | $H > 0.5$ | $H > 0.5$ | *persistent* |
| 1826-1839 | 43 | $H \leq 0.5$ | $H \approx 0.5$ | *short memory* |
| 1839-1845 | 56 | $H \leq 0.5$ | $H < 0.5$ | *anti-persistent* |
| 1845-1848 | 62 | $H \leq 0.5$ | $H \approx 0.5$ | *short memory* |
| 1849-1872 | 65 | $H \leq 0.5$ | $H < 0.5$ | *anti-persistent* |

developmental phases, their age of onset, and dominating dynamic as reflected by grundtvig's writings at a coarse and a fine resolution level. the general pattern indicates decay in the persistence of entropic states as a function of age.

Nielbo, K.L., Baunvig, K.F., Liu, B. & Gao, J. (2018). A Curious Case of Entropic Decay: Persistent Complexity in Textual Cultural Heritage, DSH.

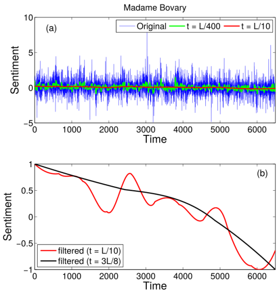


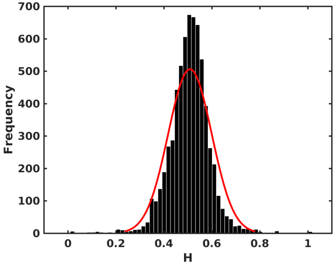



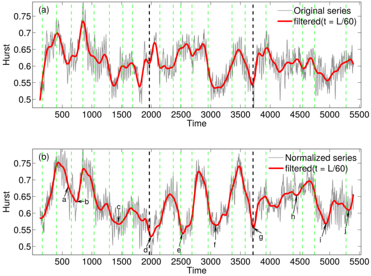




- media class impacts information dynamics – advertisements show on-off intermittent behavior, while newspapers are very persistent for *natural produce* and *food*
- product group-specific causal patterns
  *advertisement* $\rightarrow$ *article* : 0.2
  *article* $\rightarrow$ *advertisement* : 0.17
  *advertisement* $\leftrightarrow$ *article* : 0.49
- fashion shapes, while produce and energy dominate

Wevers, M., Gao, J., & Nielbo, K.L. (2020). Tracking the Consumption Junction: Temporal Dependencies between Articles and Advertisements in Dutch Newspapers, DHQ.

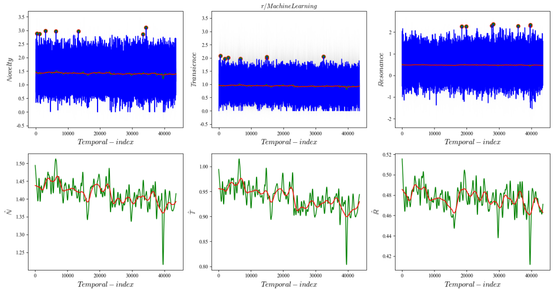


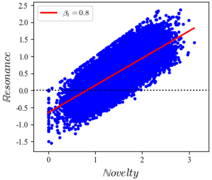



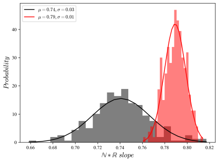



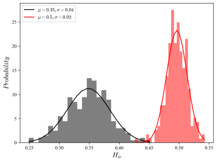

- global $H$ of story arcs provide an index of narrative coherence
- narrative coherence: $0.5 < H \leq 1.0$
- local $H$ detects changes in the narrative
- good narratives balances coherence to optimize reader motivation

Hu, Q., Liu, B, Thomsen, M.R., Gao, J. & Nielbo, K.L. (2020). Dynamic evolution of sentiments in Never Let Me Go: Insights from multifractal theory and its implications for literary analysis, DSH.



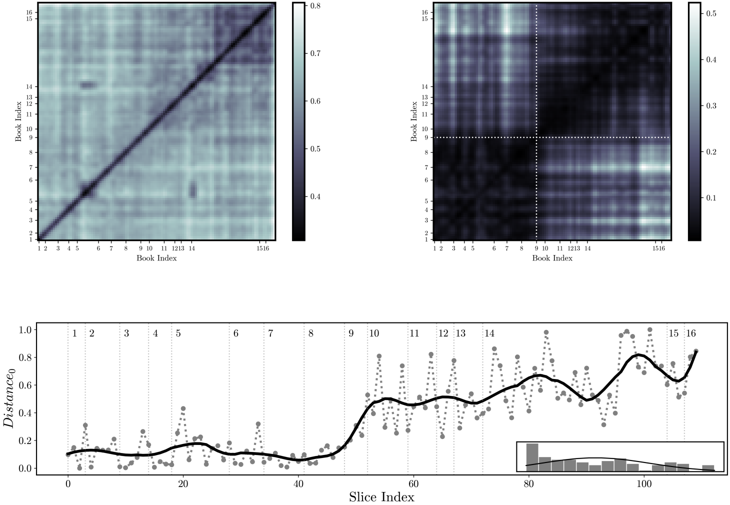




*trend reservoirs* (i.e., social media signals that display high trend potential) can be identified by their relationship between novel and resonant behavior, and their minimal persistence.

Nielbo, K.L., Vahlstrup, P.B., Gao, J. & Bechmann A. (accepted) Trend Reservoir Detection: Minimal Persistence and Resonant Behavior of Trends in Social Media. Proceedings of CHR 2020.




reliable change point in latent variable in book nine, starting in book eight, and ending in book 10.

Nielbo, K.L., Perner, M.L., Larsen, C. , Nielsen, J. & Laursen, D. (2019). Automated Compositional Change Detection in Saxo Grammaticus' Gesta Danorum. DHN Proceedings.

# pandemic information dynamics - an example

## News Information Decoupling

As the first wave of Covid-19 virus spread across the world, content alignment of news stories could be observed both within and between media sources. During December 2019 and January 2020, Covid-19 news stories were, outside China, interspersed with news coverage of other events (e.g., Hong Kong protests, Iranian–American confrontation, Trump impeachment). As the virus spread across Europe and America, news media front pages focused almost exclusively on the pandemic, all news sections (politics, business, sports, and arts) related to Covid-19, and breaking news became *corona news* in continuously updated media. 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 the dynamics of news media's information at an unprecedented level of detail.

$\mathbb{N}$: 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 $s^{(j)}$:

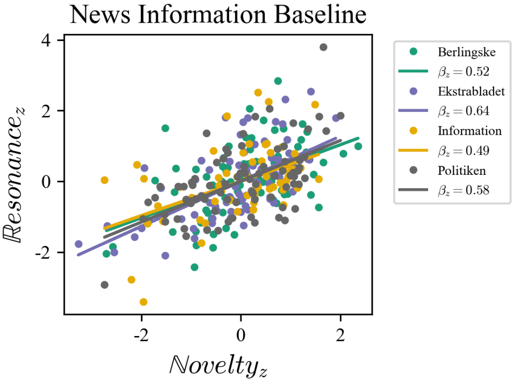
$$\mathbb{T}_w(j) = \frac{1}{w} \sum_{d=1}^{w} JSD(s^{(j)} \mid s^{(j+d)})$$

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

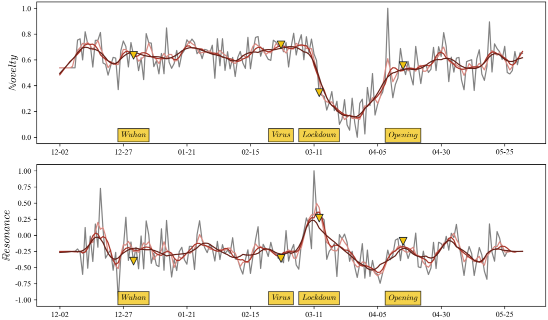
$$JSD(s^{(j)} \mid s^{(k)}) = \frac{1}{2} D(s^{(j)} \mid M) + \frac{1}{2} D(s^{(k)} \mid M)$$

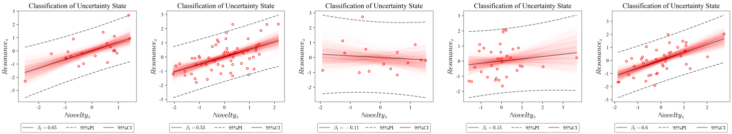
with $M = \frac{1}{2}(s^{(j)} + s^{(k)})$ and $D$ is the Kullback-Leibler divergence:

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



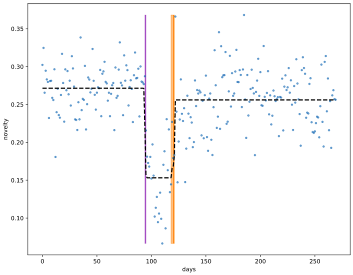

$\mathbb{N} \times \mathbb{R}$ basline models for danish legacy media



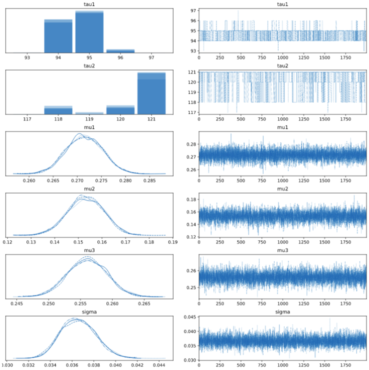




front pages from danish legacy print media *politiken*

introduction
whoami
CHCAA
self-affinity in cultural information
self-affinity
fractal analysis
author profiling
consumer history
computational narratology
trend reservoirs
author change points
pandemic information dynamics
information dynamics
news-media baseline
left-omnibus
change-point detection
right-omnibus
predictive model
affective dynamics
summary




validation of observed change points for politiken




trace plots for politiken change points

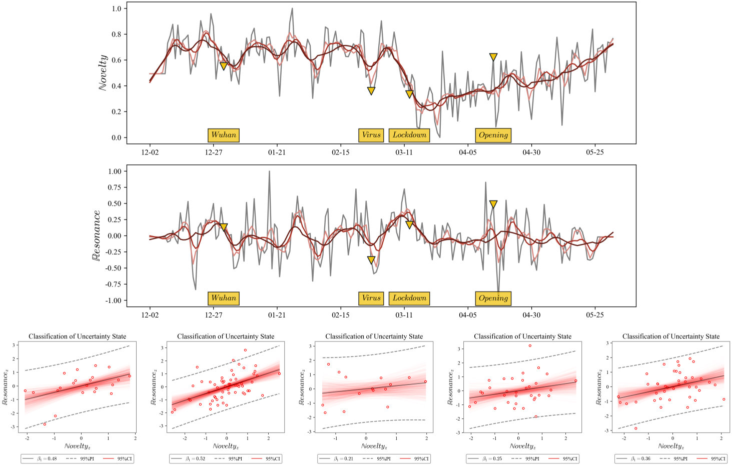
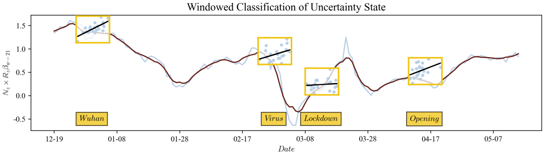
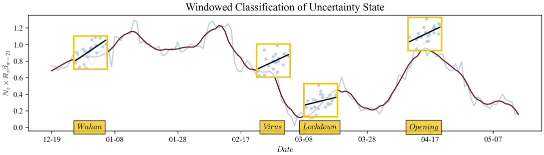



Figure: Berlingske front pages



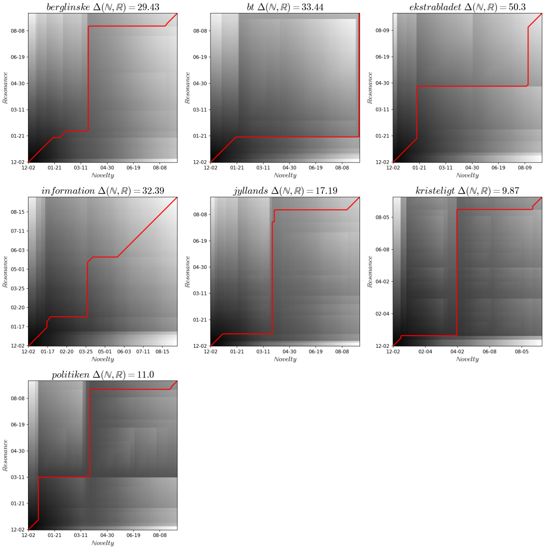

Slope $\beta$ of $\mathbb{N} \times \mathbb{R}$ model windows of length $w = 21$ with maximal overlap

## News Information Decoupling

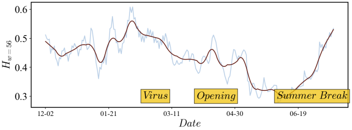
News Information Decoupling (NID) principle states that as in response to unexpected and dangerous temporally extended events, the ordinary information dynamics of news media are (initially) decoupled such that the content novelty decreases as media focus monotonically on the catastrophic event, but the resonant property of said content increases as its continued relevance propagate throughout the news information system.

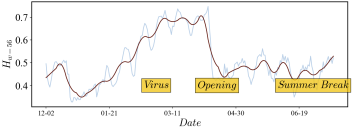


CENTER FOR HUMANITIES
COMPUTING AARHUS

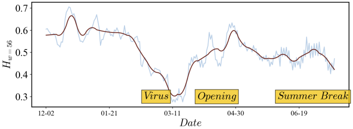



differences in NID for classes of newspapers (omnibus vs. tabloid)




- in contrast to omnibus newspapers, tabloids change their affective dynamics in response to the pandemic events unfolding
- BT loses persistence, while EB stabilizes in persistence of sentiment state(-s).



- data are noisy and complex
- in order to understand to data-generating mechanisms, we need to characterize the system with a small set of temporal indices
- dynamics are the key to answering questions about the fine-grained patterns of variation
- domain knowledge of the system should be decisive in choice of formalism

CENTER FOR HUMANITIES
COMPUTING AARHUS

## THANKS

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

## SLIDES

knielbo.github.io/files/kln_inhercmplx.pdf

CENTER FOR HUMANITIES
COMPUTING AARHUS