

When a few data points are not enough

DATALAB special

Kristoffer L Nielbo

`kln@cas.dk`

`knielbo.github.io`

`chcaa.io`

Center for Humanities Computing
Aarhus University, Denmark

Outline

When a few data
points are not enough

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

1 Tracking the consumption junction

- Effects of advertisements
- Shaping or reflecting
- Long-range dependencies
- Fractal scaling in media

Tracking the
consumption junction

- Effects of advertisements
- Shaping or reflecting
- Long-range dependencies
- Fractal scaling in media

The shape of
innovation

- Trend-detection in social
media
- Resonant information
- Resonant subreddits
- Innovation indicators

2 The shape of innovation

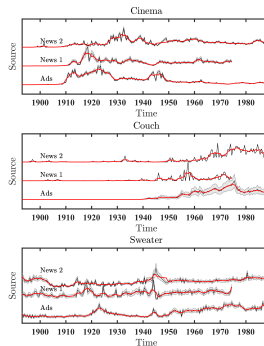
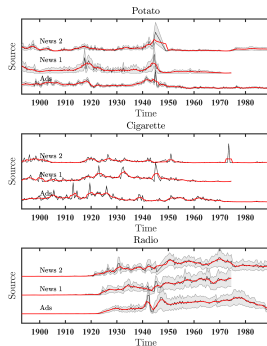
- Trend-detection in social media
- Resonant information
- Resonant subreddits
- Innovation indicators



Effects of advertisements

When a few data points are not enough

Kristoffer L. Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io



Tracking the consumption junction

Effects of advertisements

Shaping or reflecting

Long-range dependencies

Fractal scaling in media

The shape of innovation

Trend-detection in social media

Resonant information

Resonant subreddits

Innovation indicators

Articles and advertisements from *De Tijd* (1890-1974) and *De Telegraaf* 1893-1989, $N \simeq 30E^6$.



Shaping or reflecting

When a few data points are not enough

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

We test for *X Granger cause Y*, by comparing the performance of the nested 'newspaper discourse only' model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_k y_{t-k} + \epsilon$$

with the full 'newspaper and advertisement discourses' model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_k y_{t-k} + \alpha_1 x_{t-1} + \dots + \alpha_m x_{t-m} + \epsilon$$

to identify which one does the better job at explaining y_t based on the residuals. The zero-model for the hypothesis then is $H_0 : \alpha_i = 0$ for each i of the element $[1, m]$ with the alternative hypothesis being $H_1 : \alpha_i \neq 0$ for at least one i of the element $[1, m]$. We applied the test bi-directionally such that a shaping relation finds support if we can confirm that '*X Granger cause Y*' and reject that '*Y Granger cause X*' in case of a reflecting relationship (the inverse of shaping). Finally, if both '*X Granger cause Y*' and '*Y Granger cause X*' find support this is viewed as support for a more complex relationship between the two time series.

Shaping: *advertisements* \rightarrow *articles*

Reflecting: *articles* \rightarrow *advertisements*

Complex: *advertisements* \leftrightarrow *articles*

Tracking the consumption junction

Effects of advertisements

Shaping or reflecting

Long-range dependencies

Fractal scaling in media

The shape of innovation

Trend-detection in social media

Resonant information

Resonant subreddits

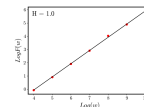
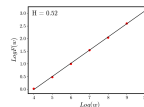
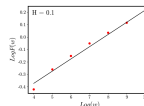
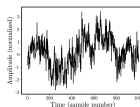
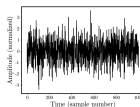
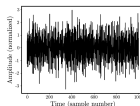
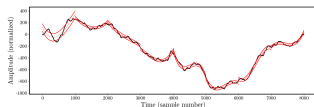
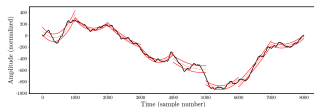
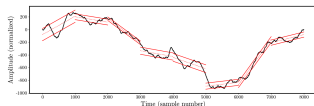
Innovation indicators



Long-range dependencies

When a few data points are not enough

Kristoffer L. Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io



Tracking the consumption junction
Effects of advertisements
Shaping or reflecting
Long-range dependencies
Fractal scaling in media

The shape of innovation
Trend-detection in social media
Resonant information
Resonant subreddits
Innovation indicators

Computation of local fluctuations around linear, quadratic, and cubic trends

Estimation of Hurst parameter using Adaptive Fractal Analysis

K. L. Nielbo, K. F. Baunvig, B. Liu, and J. Gao, "A curious case of entropic decay: Persistent complexity in textual cultural heritage," Digital Scholarship in the Humanities, 2018.

Software library: <https://github.com/knielbo/saffine>

When a few data points are not enough

Kristoffer L. Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Tracking the consumption junction

Effects of advertisements

Shaping or reflecting

Long-range dependencies

Fractal scaling in media

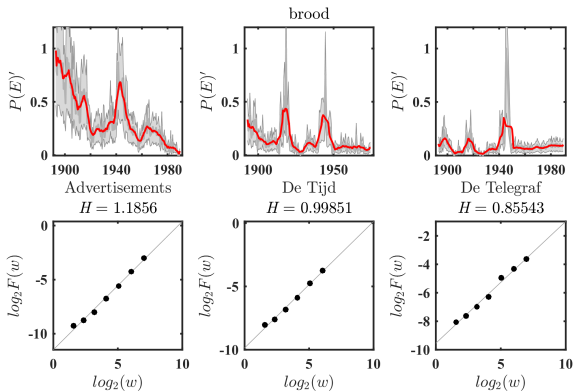
The shape of innovation

Trend-detection in social media

Resonant information

Resonant subreddits

Innovation indicators



Adaptive Fractal Analysis to estimate Hurst exponent: *antipersistent correlations*: $0 < H < \frac{1}{2}$, *memoryless*: $H = \frac{1}{2}$, *persistent correlations*: $\frac{1}{2} < H < 1$



When a few data points are not enough

Kristoffer L. Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Tracking the consumption junction

Effects of advertisements

Shaping or reflecting

Long-range dependencies

Fractal scaling in media

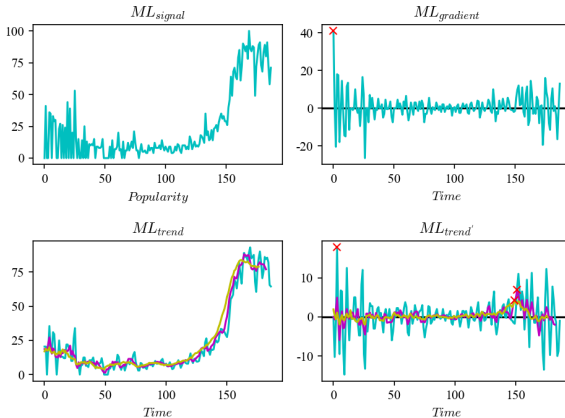
The shape of innovation

Trend-detection in social media

Resonant information

Resonant subreddits

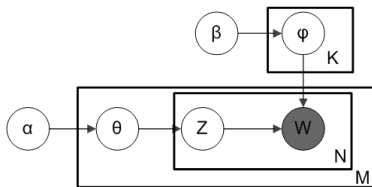
Innovation indicators



Trend detection in social media is modeled on natural catastrophes and epidemics → point-like events.



Resonant information



Novelty over window w :

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

with \mathbb{T} ransience:

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

for \mathbb{R} esonance

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

When a few data points are not enough

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Tracking the consumption junction

Effects of advertisements

Shaping or reflecting

Long-range dependencies

Fractal scaling in media

The shape of innovation

Trend-detection in social media

Resonant information

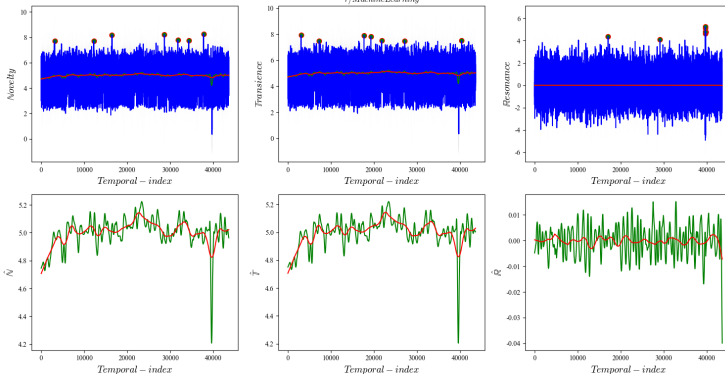
Resonant subreddits

Innovation indicators

Resonant subreddits

When a few data points are not enough

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io



Tracking the consumption junction
Effects of advertisements
Shaping or reflecting
Long-range dependencies
Fractal scaling in media

The shape of innovation

Trend-detection in social media

Resonant information

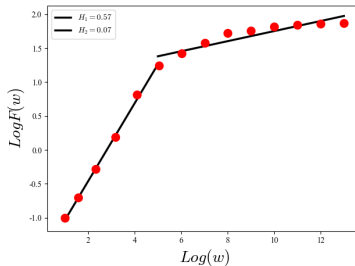
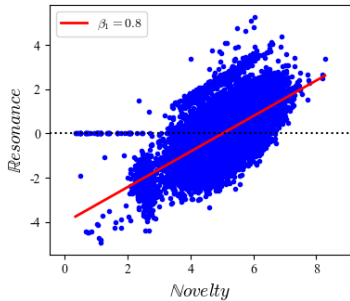
Resonant subreddits

Innovation indicators

Trend detection for social media works well for point-like events (e.g., natural catastrophes and epidemics), but what are the signature(-s) of social trends? Analyzing 7TB+ data from reddit.com, we find that in certain subreddits, novelty resonates more with the future and that content display long-term memory at short and intermediary time-scales.



Innovation indicators



When a few data points are not enough

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Tracking the consumption junction

Effects of advertisements

Shaping or reflecting

Long-range dependencies

Fractal scaling in media

The shape of innovation

Trend-detection in social media

Resonant information

Resonant subreddits

Innovation indicators



When a few data
points are not enough

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io
chcaa.io

Tracking the
consumption junction

Effects of advertisements

Shaping or reflecting

Long-range dependencies

Fractal scaling in media

The shape of
innovation

Trend-detection in social
media

Resonant information

Resonant subreddits

Innovation indicators

Thank you for your attention

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

slides: http://knielbo.github.io/files/datalab_special.pdf

