

The Dostoyevskian Trope

Persistent Processes and State Incongruence in 19th Century Textual Cultural Heritage

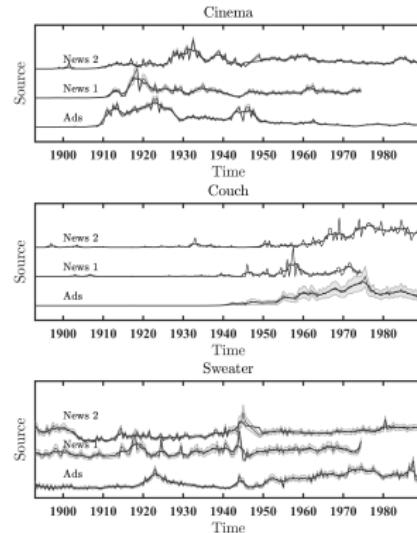
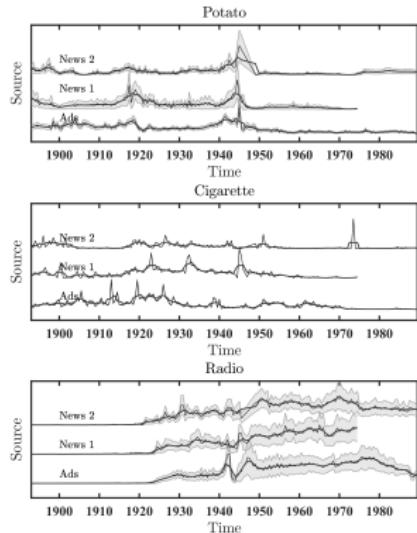
Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

Dept. of History & SDU eScience Center
University of Southern Denmark

PROGRAM

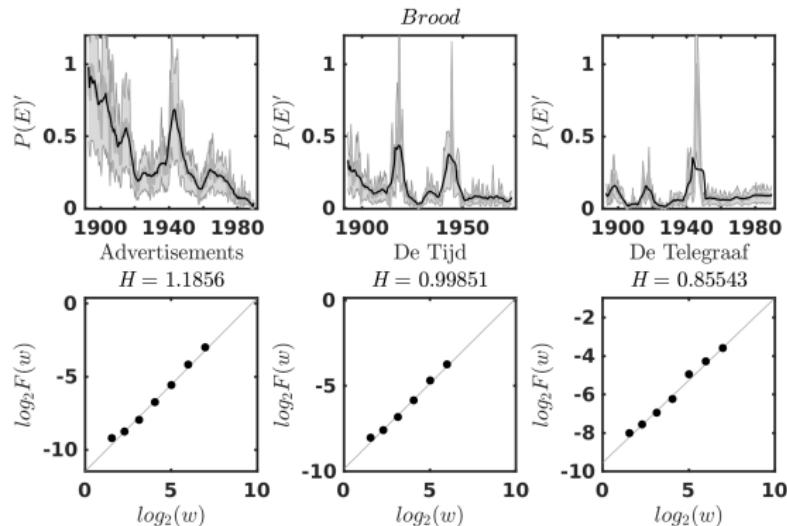
0	Motivation	history and dynamics
1	RQs and Data	Danish textual cultural heritage
2	Design	fractal analysis and causal dependencies
3	Kehre	author-specific cognitive dynamics
4	D-trope	cognitive-affective dependencies
*	Extra	literary optimality

Background



-
- dynamics are fundamental to cultural and historical objects of research
 - in history and anthropology “dynamics” almost has a status dogma

Background



-
- structural features can often be “read off” the historical process, but are not captured by simple statistics
 - risk of ignoring core features of the process (e.g. periodicity, fractal properties, long-range correlations).

Overview of Research Problems

Die Kehre

- “authors of fiction (and non-fiction) undergo personal paradigm shifts”

Dostoyevskian Trope

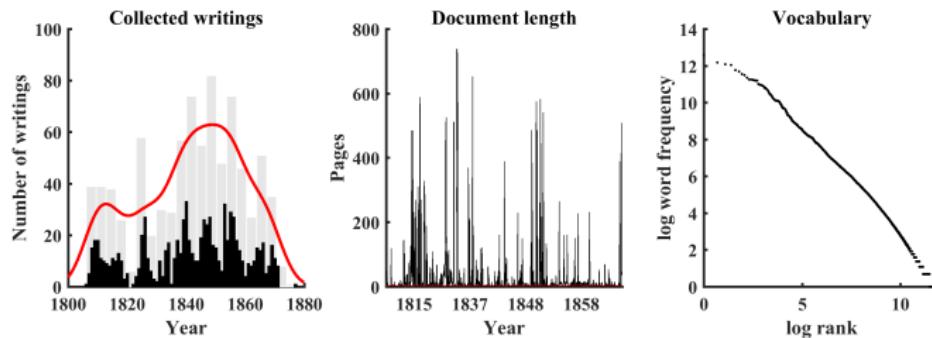
- “authors’ creative states are inversely related to their emotional states”
- “authors’ creative states *depend* on their emotional states” (state incongruent writers)

Geertzian v-function

- particular case where the *kehre* is reflected in a topicality shift for state congruent writers, which might characterize successful innovators

Sources of Danish Textual Cultural Heritage

- N.F.S. Grundtvig, active years 1804-1871 (born: 1783, dead: 1872)
- H.C. Andersen, active years 1829-1874 (born: 1805, dead: 1875)
- S.A. Kierkegaard, active years 1830-1855 (born: 1813, dead: 1855)



DATA|Danish Textual Cultural Heritage

- $N = 1329$ documents ($n_{NSFG} = 921$, $n_{HCA} = 194^*$, $n_{SAK} = 214$)
- language normalization (orthographic variation and casefolding)
- length normalization (slice len = $10^2 - 10^3$ words)

Normalization of Historical Texts

Orthographic Variation

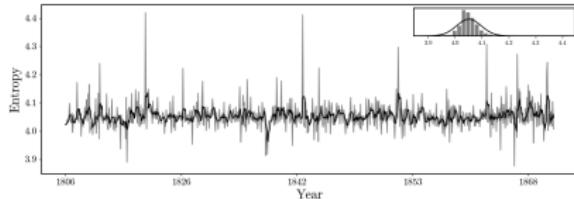
- automated analysis depends critically on existing tools and data (ex. sentiment dictionaries)
- NLP, IR and TM resources “suffer from presentism”
- projects often try to adapt the tool (ex. modify dictionary to historical data set)
- this solution scales badly due to lack of standardization

Solution

- statistical spelling corrector that treats historical variants as errors of contemporary Danish
- rule-based improvement

the dostoyevskian trope

- authors' creative states are **inversely** related to their affective states (*state incongruence*)
- authors' creative states **depend** on their affective states
- use average dictionay-based **sentiment** scores as a proxy for author's affective states
- use Shannon's (source) **entropy** h as a proxy for author's cognitive complexity

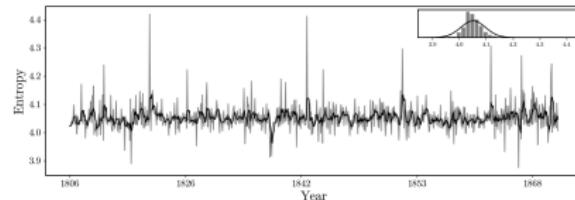


$$h = - \sum_{i=1}^K p_i \times \log_2(p_i)$$

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

the dostoyevskian trope

- authors' creative states are **inversely** related to their affective states (*state incongruence*)
- authors' creative states **depend** on their affective states
- use average dictionay-based **sentiment** scores as a proxy for author's affective states
- use Shannon's (source) **entropy** h as a proxy for author's cognitive complexity

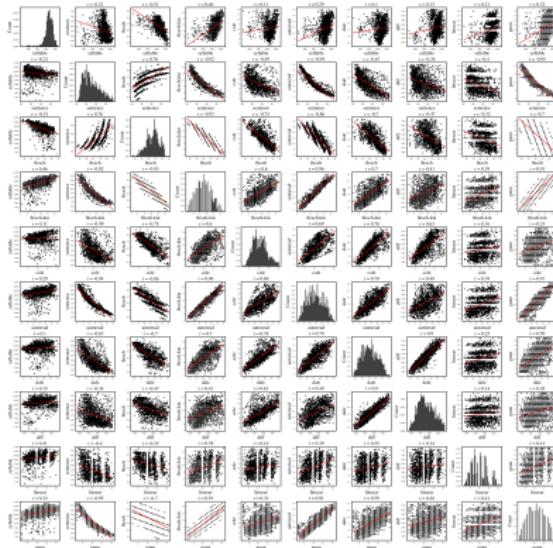


$$h = - \sum_{i=1}^K p_i \times \log_2(p_i)$$

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

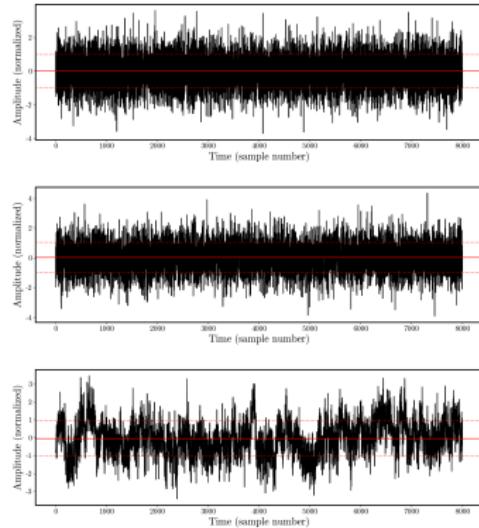
A comment on Shannon Entropy

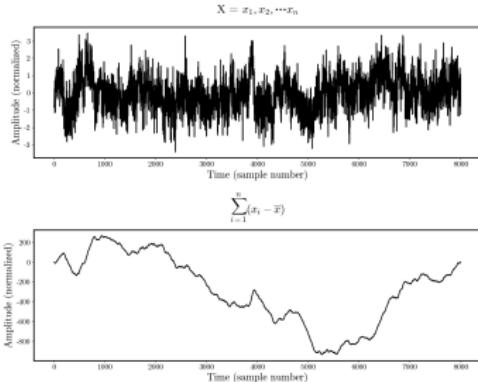
- axiomatic approach to information
- correlates with perceived complexity across modalities
- for natural language it is a direct measure of lexical variability
- associated with a wide range of readability* measures
- lost popularity due to semiotics and chaos theory



Fractal properties of entropy

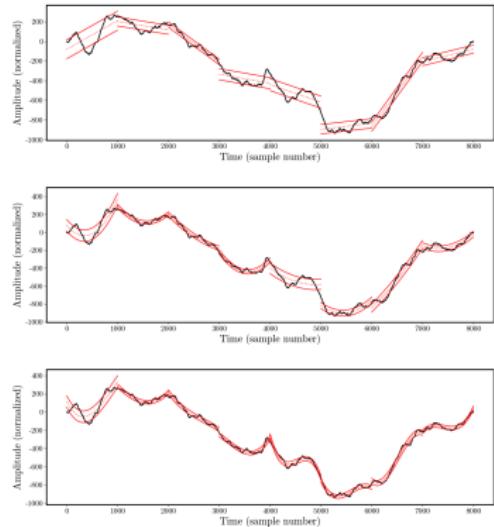
- not interested in an average state of complexity *per se*
- capture long-range correlations in h as a model of how complexity states persist
- *Detrended Fluctuation Analysis* or *Adaptive Fluctuation Analysis*
- estimate the power law relation between the overall variation for multiple time-windows (scales) as the Hurst exponent H





construct random walk process:

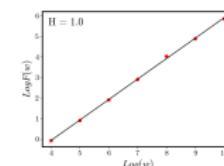
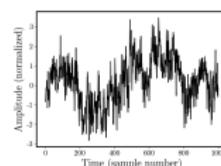
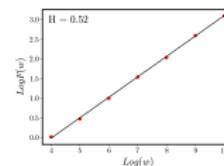
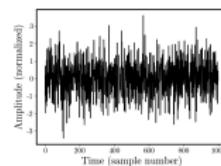
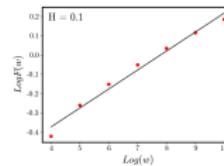
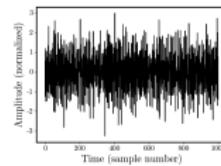
$$u(n) = \sum_{i=1}^n (x_i - \bar{x}), \quad n = 1, 2, \dots, N,$$



Estimation and interpretation

for $1/f^{2H+1}$ processes:

- anti-persistent process: $0 < H < 0.5$
- short-range correlations only $H = 0.5$
- $0.5 < H < 1$ **persistent process**



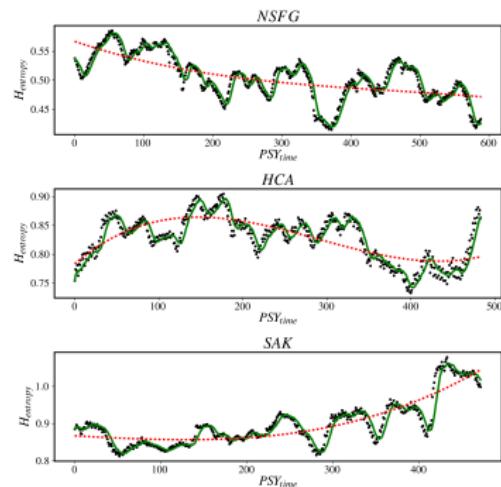
Finally, we estimate time-varying H for the data in order to build "creativity" processes resulting in an:

Author Typology

Kehre prototype (persistence → anti-persistence): **NSFG**

eternal child (persistence): **HCA**

extended 27 Club (persistence → intermittency): **SKS**



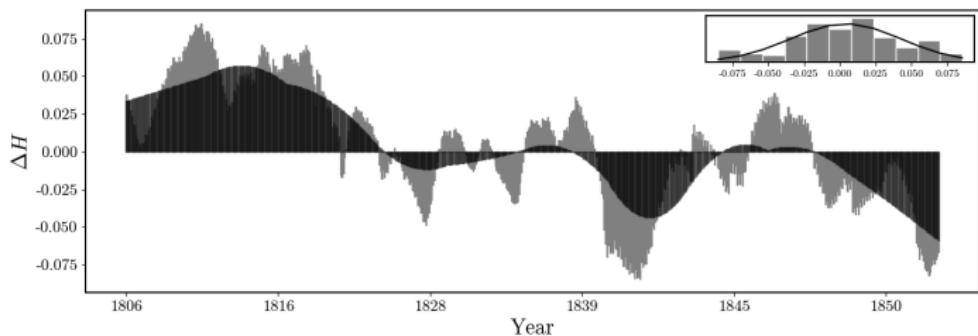
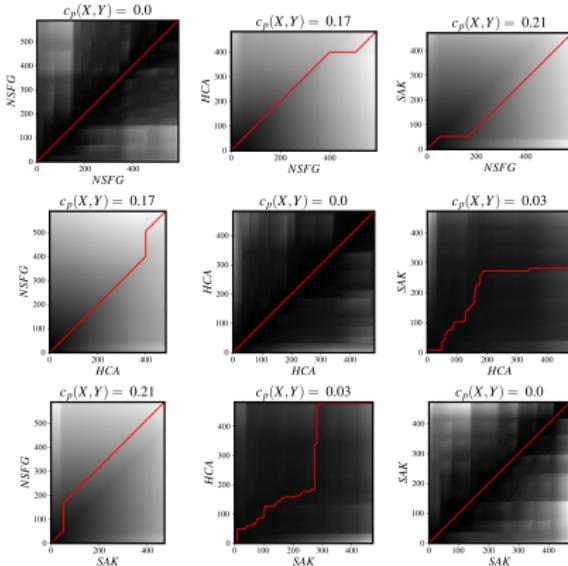


Table: Dominant Dynamic in the Phases of N.F.S. Grundtvig's Writings

Time period	Age of onset	Coarse	Fine	Behavior	Profile
1806-1826	23	$H > 0.5$	$H > 0.5$	<i>persistent</i>	theoretician
1826-1839	43	$H \leq 0.5$	$H \approx 0.5$	<i>short memory</i>	pragmatic
1839-1845	56	$H \leq 0.5$	$H < 0.5$	<i>anti-persistent</i>	breakthrough
1845-1848	62	$H \leq 0.5$	$H \approx 0.5$	<i>short memory</i>	disease
1849-1872	65	$H \leq 0.5$	$H < 0.5$	<i>anti-persistent</i>	politician

Author comparison

- with sequence alignment techniques (*DTW*), it becomes possible to compare dynamics between authors
- **SKS** is similar to **HCA** and dissimilar to **NFSG**, but that might have changed had **SKS** developed a late style

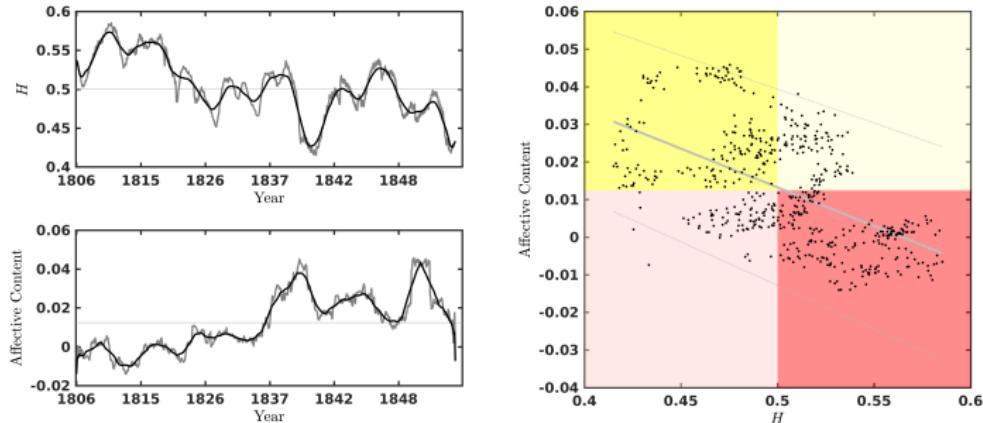


Affective state

- Lexical matching based on a contemporary Danish sentiment dictionary
- Technique is popular for author profiling (among other things)

Example

“Ingen veed, trods Halfreds Sange, Hvor Kong Olav fandt sin Grav;”
nobody knows, in spite of Halfred's songs, where king Olav found his grave
ingen vide trods Halfred sang hvor kong Olav finde sin **grav**
 $\langle -1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, -2 \rangle$



Creativity::affect

-
- Author's creative states are inversely related to his emotional states – NSFG is an example of a state incongruent author

Granger Causality

compare nested model:

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

with full model:

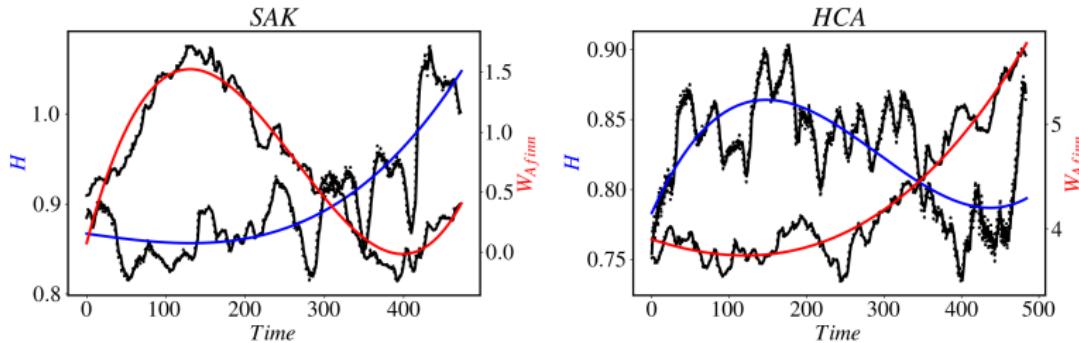
$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_k y_{t-k} + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_m x_{t-m} + \epsilon$$

to see which one does the better job at explaining y by testing:

$$H_0 : \alpha_i = 0 \text{ for each } i \text{ of } [1, m]$$

$$H_1 : \alpha_i \neq 0 \text{ for at least one } i \text{ of } [1, m]$$

- bidirectional test such that **dostoyevskian trope** finds support IF we can confirm that '*Sentiment Granger cause h*' AND reject that '*h Granger cause Sentiment*'



The Dostoyevskian Trope

state incongruent authors (creativity::affect): **NFSG, HCA, SAK**

writes on affective state (affect → creativity): **NFSG, SAK**

Conclusion

Kehre

- “No mathematician should ever allow himself to forget that mathematics, more than any other art or science, is a young man's game.” (G.H. Hardy, *A Mathematician's Apology*, 1940)
- 30-40 yrs
- change in scaling dynamics for creativity that separates early and late writer: NFSG, SAK*

Dostoyevskian trope

- support for the state incongruent author among authors from 19th century Denmark
- some authors seem to be driven by their affective state

Literary optimality

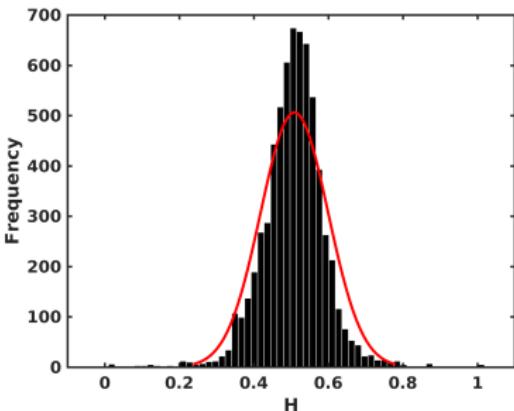
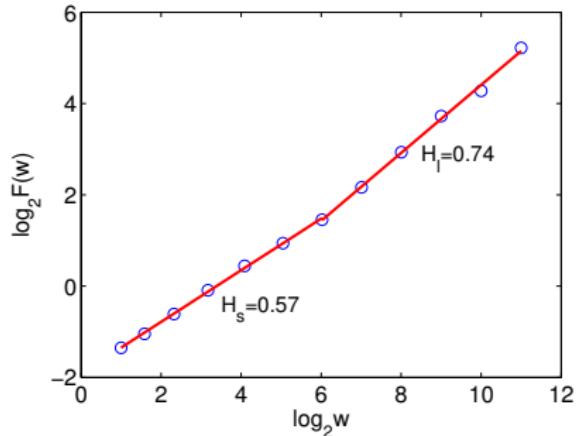
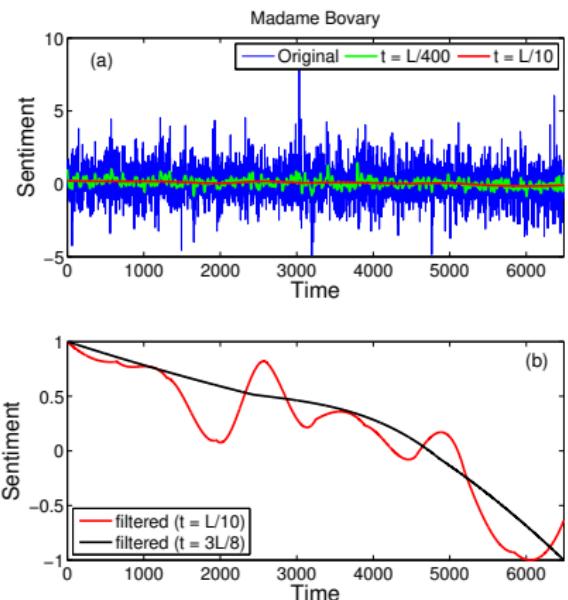


Figure: sentiment analysis and adaptive filtering reconstructs narrative vectors that reflect the reader experience. Particular fractal scaling-range, $0.6 < H \leq 0.8$, indicates literary optimality.

THANK YOU

**knielbo@sdu.dk
knielbo.github.io**

& credits to

Katrine F. Baunvig, datacube, University of Southern Denmark, DK
Jianbo Gao and Bin Liu, Institute of Complexity Science and Big Data, Guangxi University, CHN

data

adl.dk, grundtvigsværker.dk, sks.dk

funding

Institute of Pure and Applied Mathematics (UCLA)
Danish Agency for Science and Innovation