Data-intensive Research and eScience

A New Possibility Space for Historical and Cultural Research

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

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

May 31, 2018

Outline

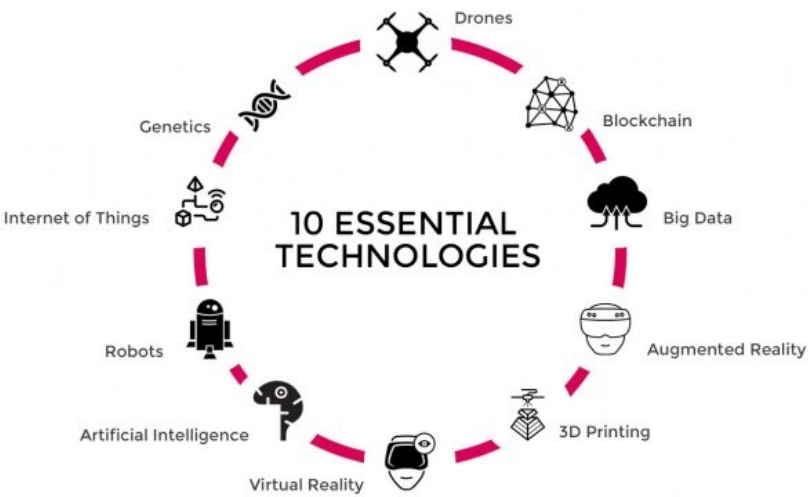
SDU

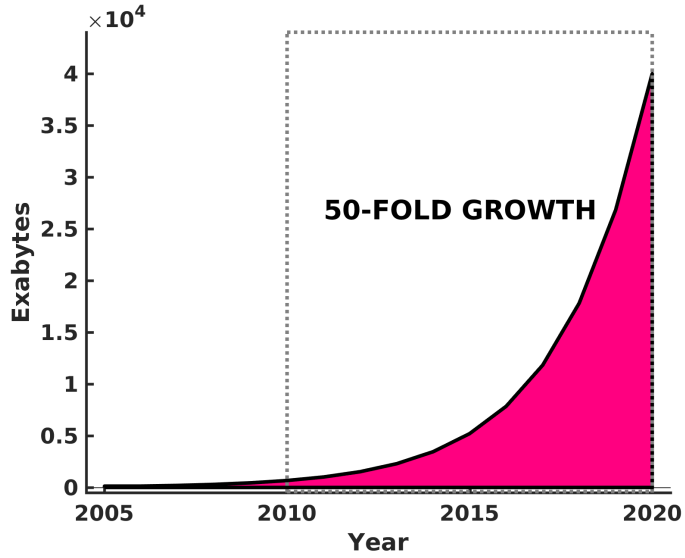
DATA-INTENSIVE RESEARCH

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

SDU

The data deluge





– the **data deluge** is transforming knowledge discovery and understanding in every domain of human inquiry

a large part of these data are soft and unstructured

– to get value from these data, humanities and social sciences must utilize advanced analytics solutions

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

SDU

Humanities need analytics

Figure 1: The increase in research publications & databases alone requires computational literacy. Publications related to Gospel of Marc (KJV) $>$ 50K, $\sim$ 16,500 words in 16 chp. on 11 p.

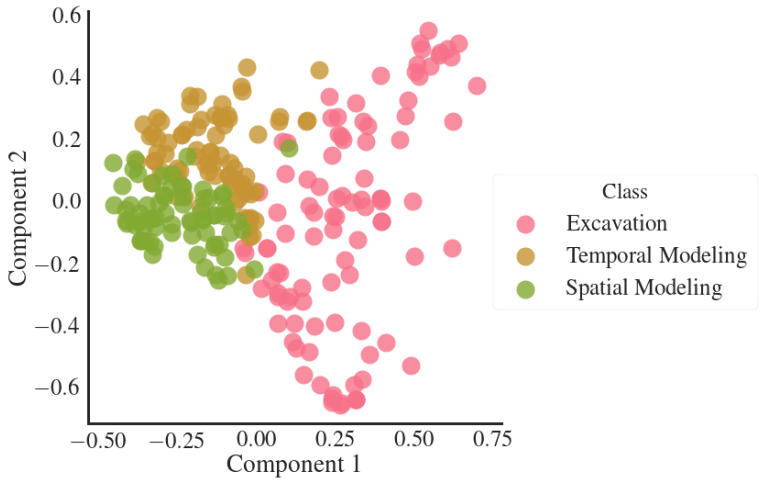


Figure 2: Advanced (human) analytics can merge, aggregate and project hetereogeneous data into lower dimensional spaces that allow visual manipulation

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

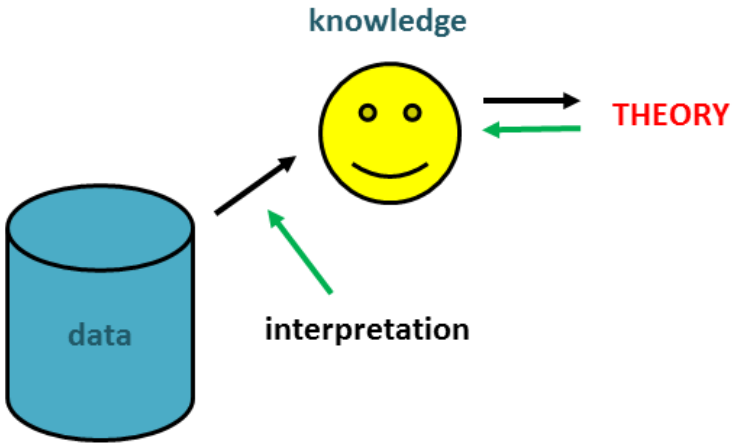
SDU

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

DATA ANALYTICS

SDU

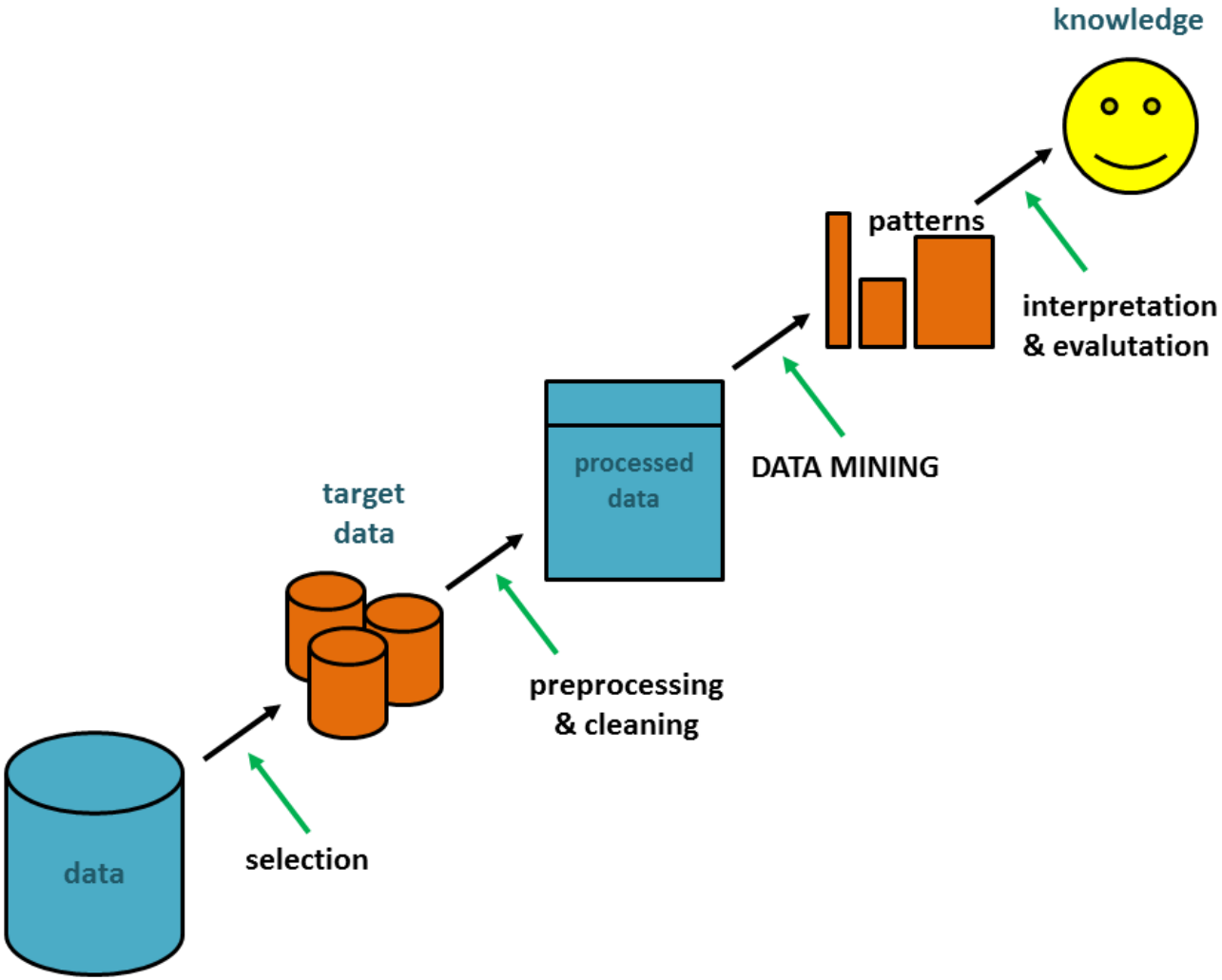
SSH default workflow



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

SDU

KDD workflow



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

SDU

Human analytics - computationally empowering humanities

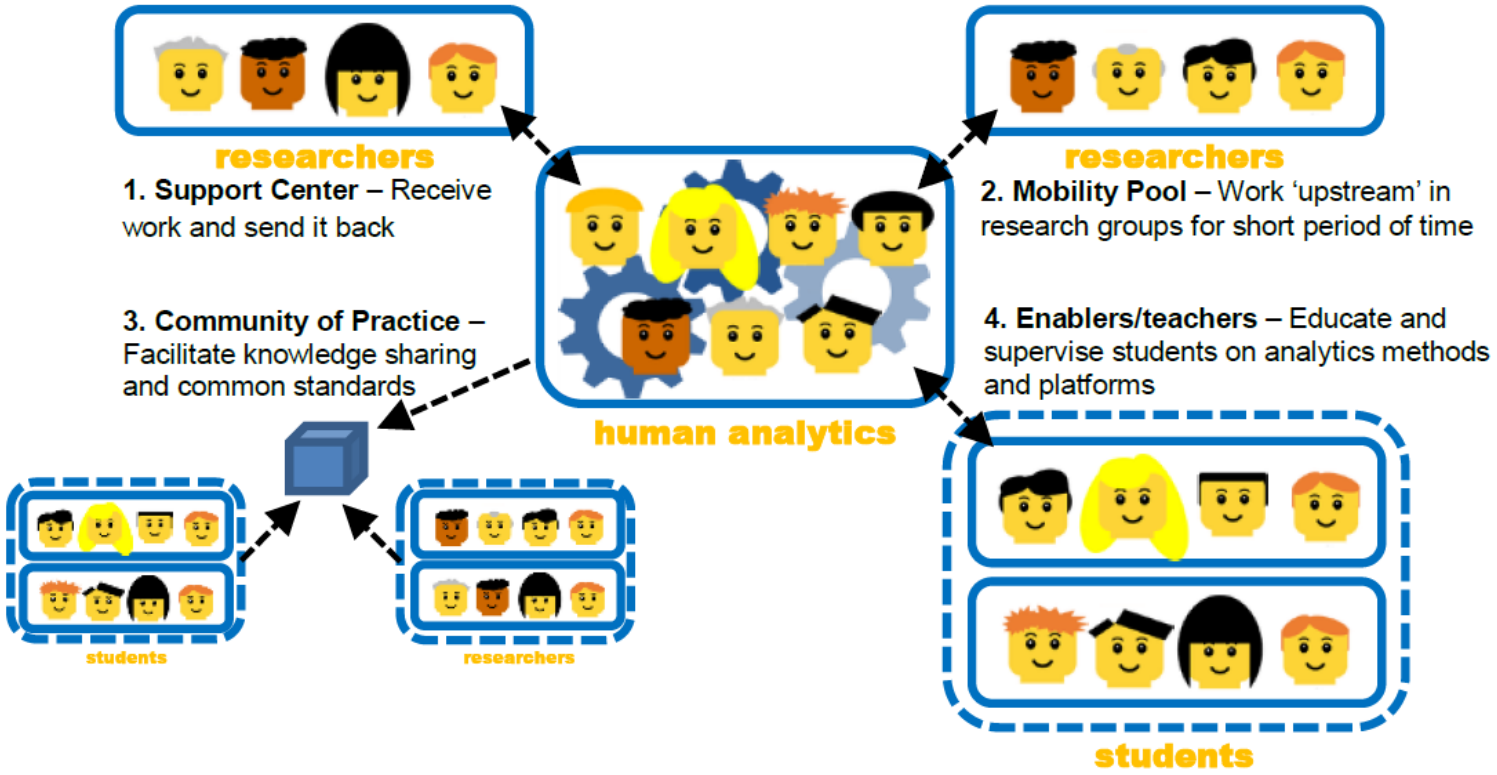


Figure 3: @ SDU we have established an analytics group (2.5 FTEs) with four modes of empowerment: **support**, **mobility**, **community of practice**, and **enablers**

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

SDU

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

APPLICATIONS

SDU

Chinese philosophy

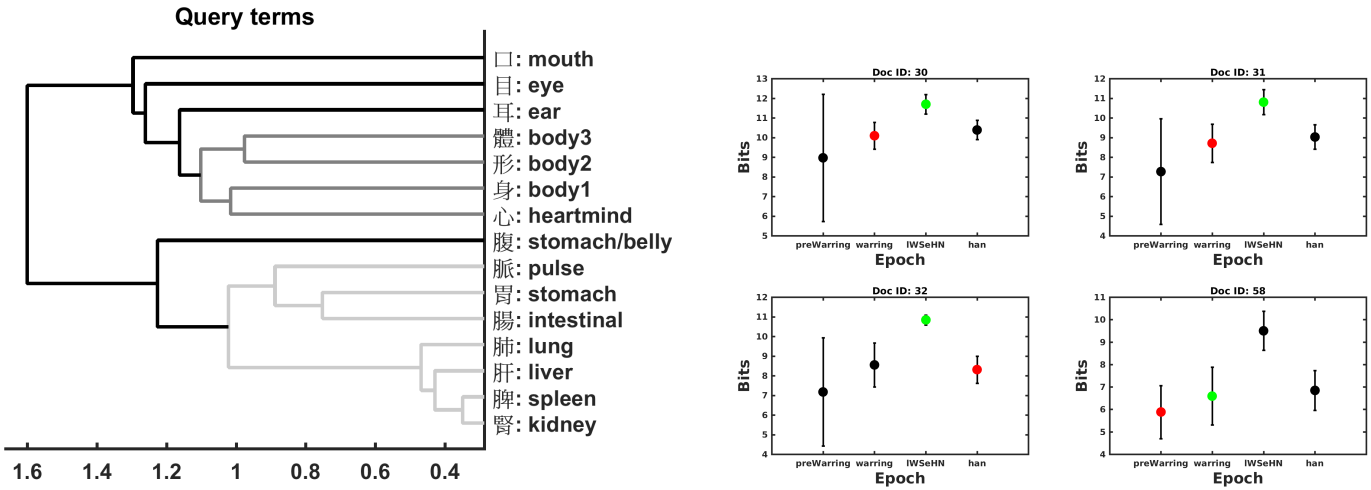


Figure 4: Philosphers and sinologists have been debating the existence of **mind-body dualism** in classical Chinese philosophy. With domain experts, we identified a **hierarchical dualistic space** based on latent semantic models. One model (LDA) was further utilized to predict class of origin for **controversial texts** slices.

Slingerland, E., Nichols, R., Nielbo, K., & Logan, C. (2017). The Distant Reading of Religious Texts: A Big Data Approach to Mind-Body Concepts in Early China. Journal of the American Academy of Religion, 85(4), 985–1016.

Nichols, R., Slingerland, E., Nielbo, K., Bergeton, U., Logan, C., & Kleinman, S. (2018). Modeling the Contested Relationship between Analects, Mencius, and Xunzi: Preliminary Evidence from a Machine-Learning Approach. The Journal of Asian Studies, 77(01), 19–57.

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

Medieval history and novelty detection

– historians debate historical transitions

– Saxo's *Gesta Danorum* c. 1200 CE history of the Danish royal dynasty

– transition between book 8 or 9?

– transition point or gradual?

– traditional word-level representation is ambivalent

– latent semantic model was trained over sentence windows

– change detection and recurrence plot used to identify phase transition centered in book 9

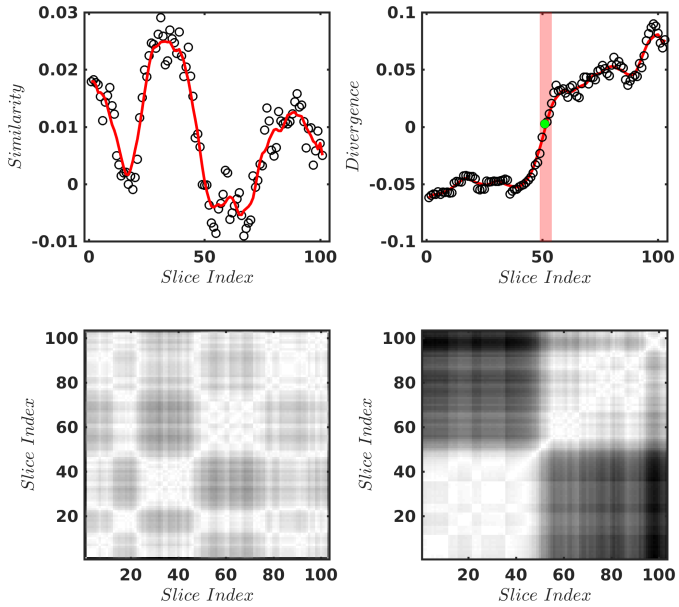
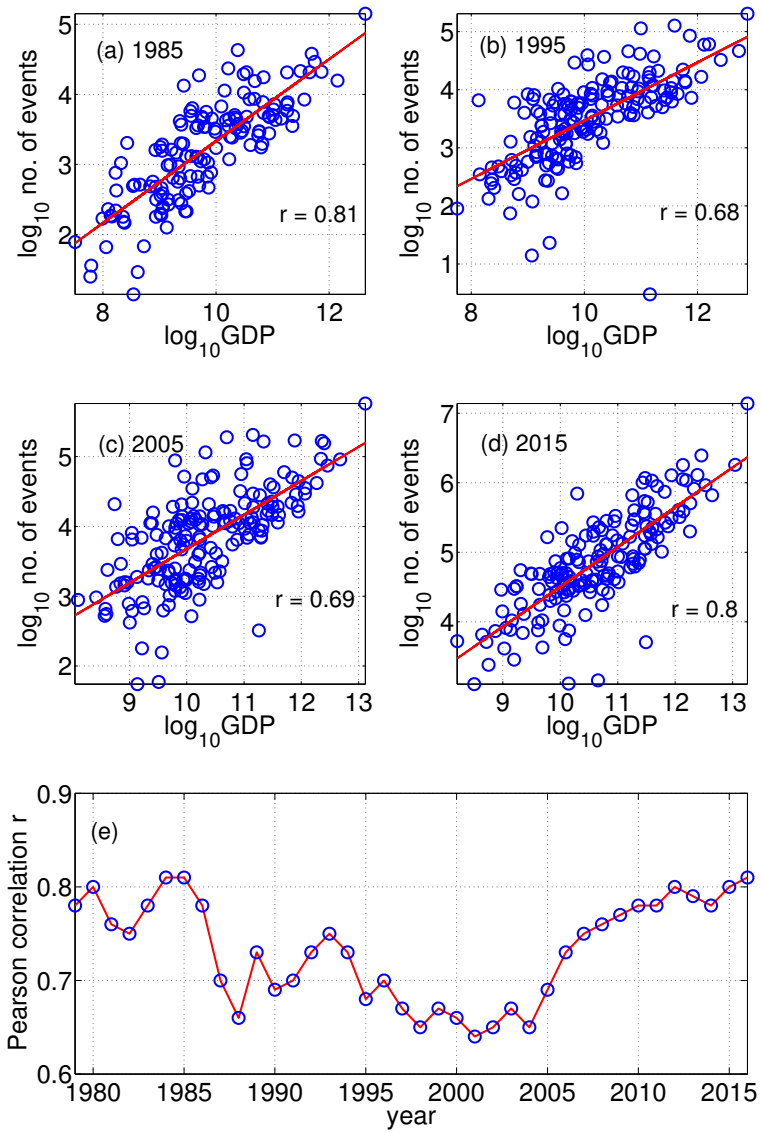


Figure 5: Cosine distance and KLD for TD high-rank vector space and guided LDA model respectively.

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

%domestic events indicates populism



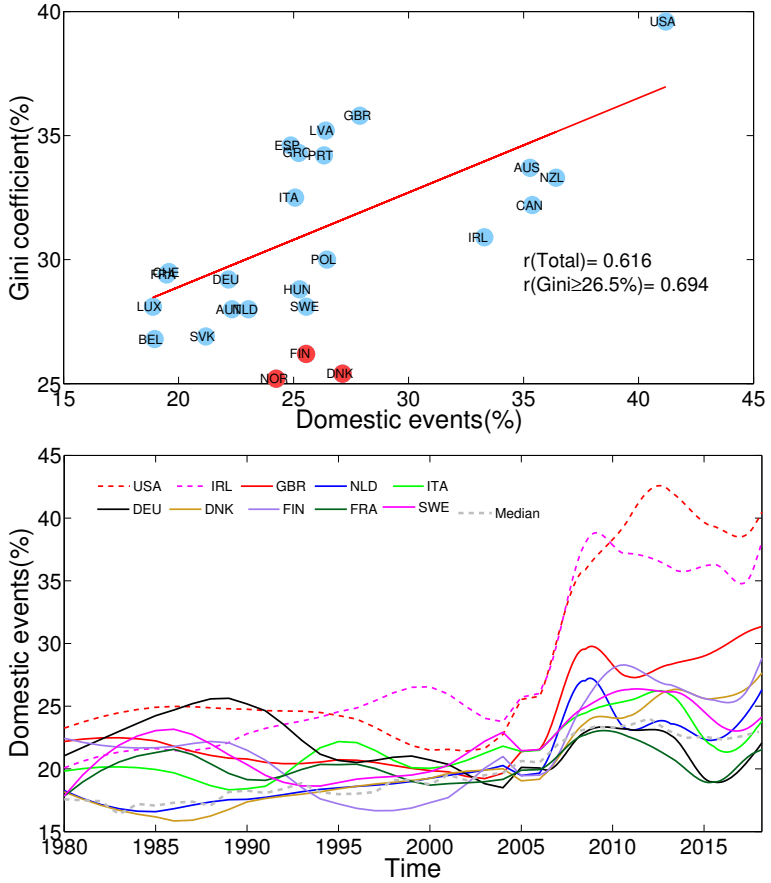


Figure 6: Event counts in the GDELT database reflect economic and political dynamics

Gao J., Ma M., Liu B., Nielbo K.L., Roepstorff A., Tangherlini T., Roychowdhury V. (in review) Brexit and Trump Presidency: were they black swan events or inevitable?

Consumer goods and cultural memory

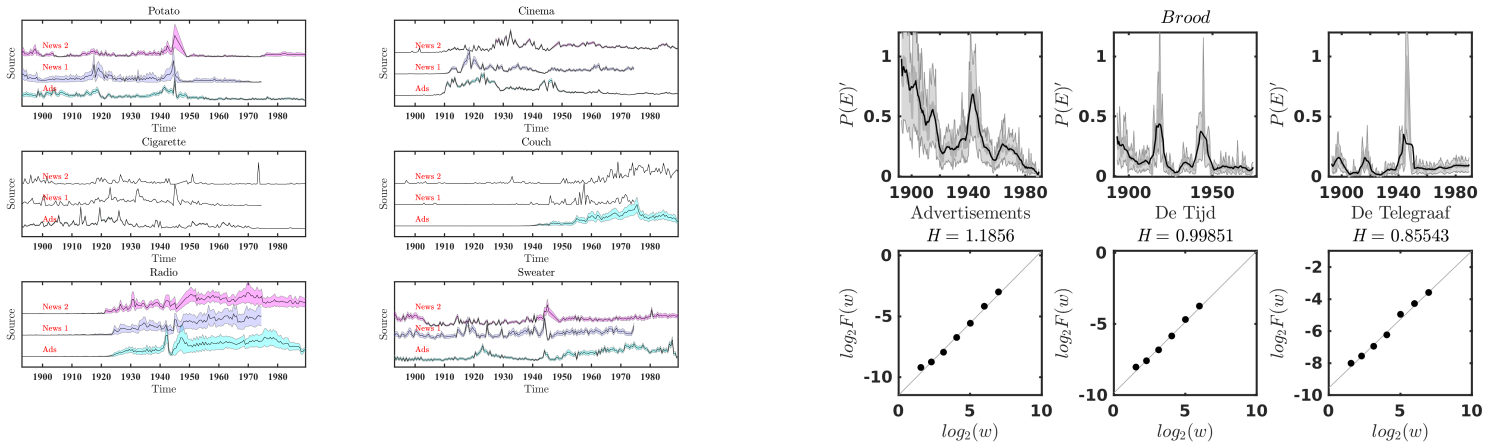
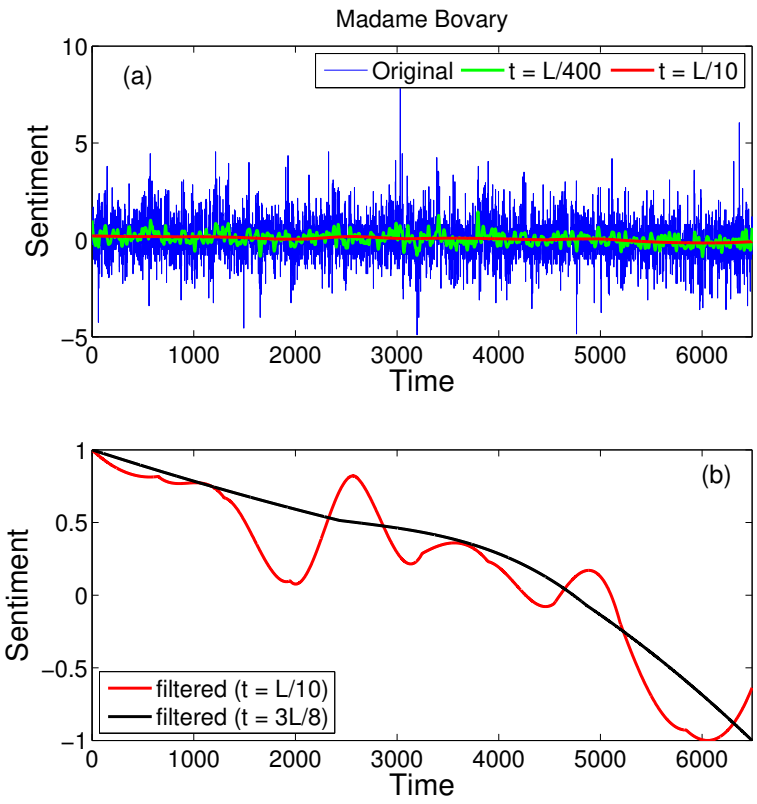


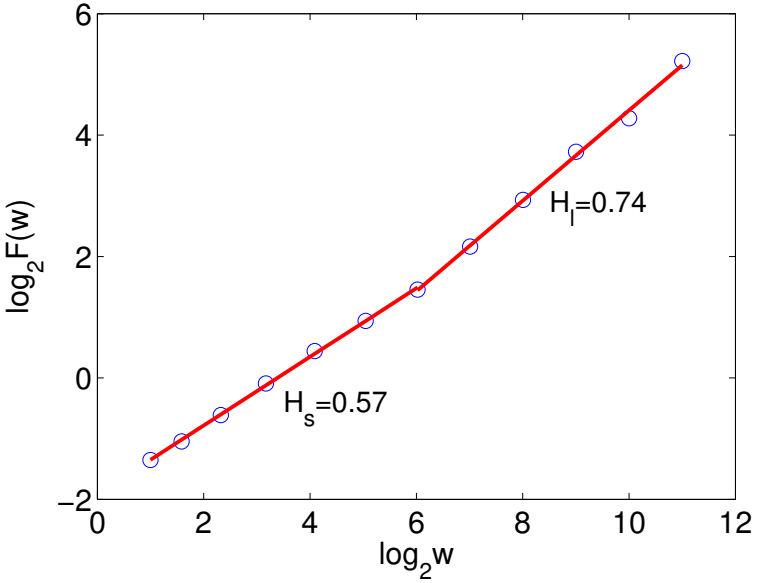
Figure 7: Historians and media researchers theorize about the causal dependencies between public discourse and advertisement. Causal modeling of keyword frequencies (from seedlists) indicated that for some categories 'ads shape society', while other categories merely 'reflect'. Advertisements show a faster decay (on-off intermittant behavior) than public discourse (long-range dependencies) - a proxy for cultural memory.

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

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

Literary optimality





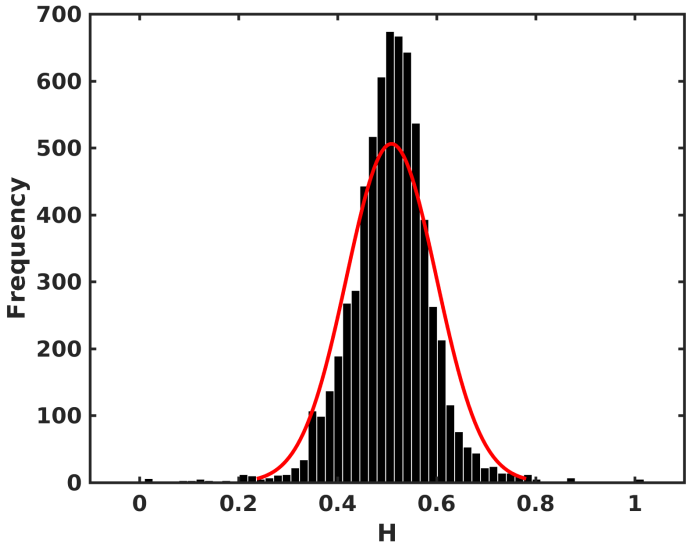


Figure 8: sentiment analysis and adaptive filtering reconstructs narrative vectors that reflect the reader experience. Particular fractal scaling-range, $0.6 < H \leq 0.8$, indicates literary optimality.

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

Author change points

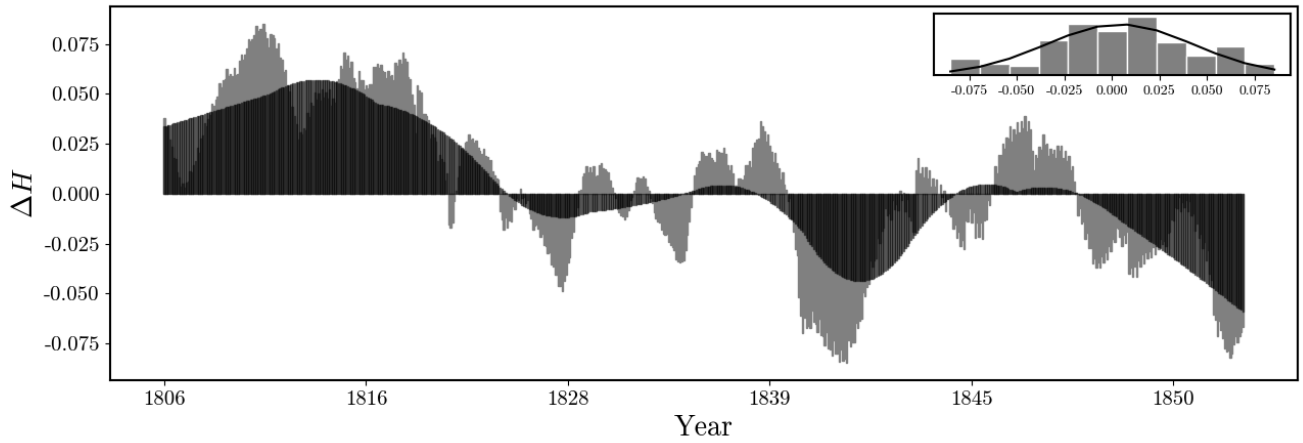


Table 1: Dominant Dynamic in the Phases of N.F.S. Grundtvig's Writings

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

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

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

Dynamic author profiling

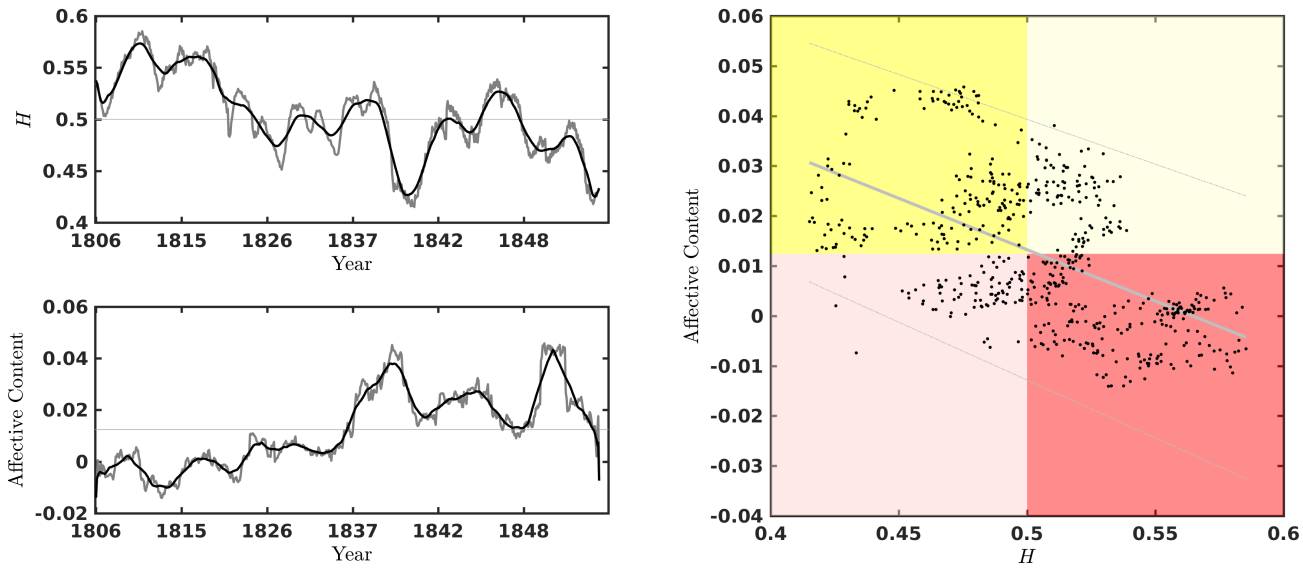


Figure 9: Combining persistent entropic trends with sentiment analysis and causal modeling, we can study **"the tormented artist"** phenomena in intellectual history.

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

SDU

Decision support for OCD

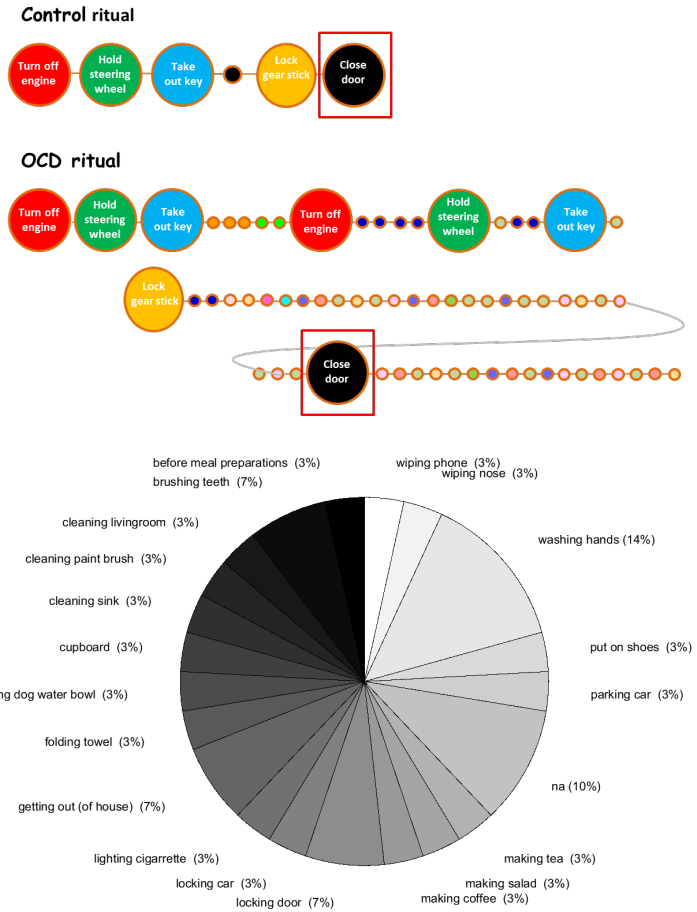


Figure 10: Event logging database annotated with Observer XT for OCD, comorbid, and control. [1]

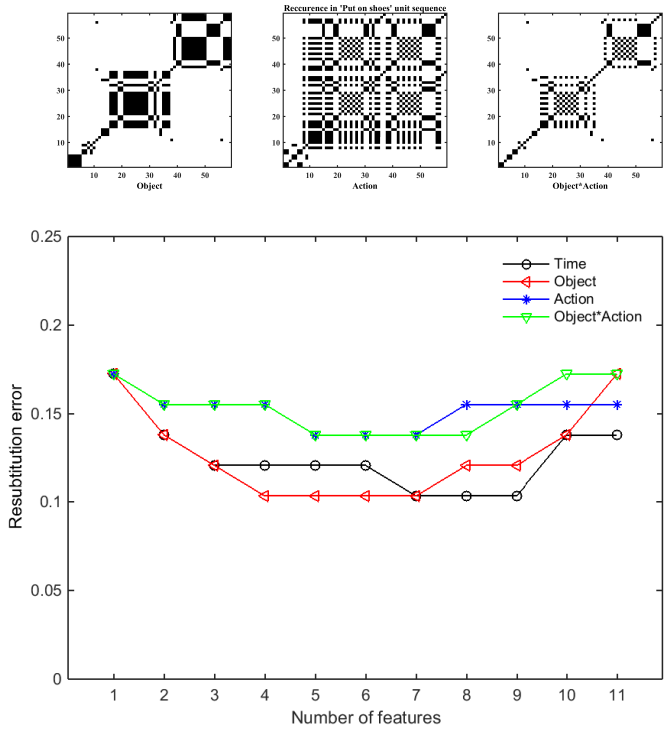


Figure 11: Binomial classifier (OCD vs. control) on unseen data.

[1] Zor, R., Hermesh, H., Szechtman, H., & Eilam, D. (2009). Turning order into chaos through repetition and addition of elementary acts in obsessive-compulsive disorder (OCD). World Journal of Biological Psychiatry, 10(4.2), 480–487.

The structure of dance

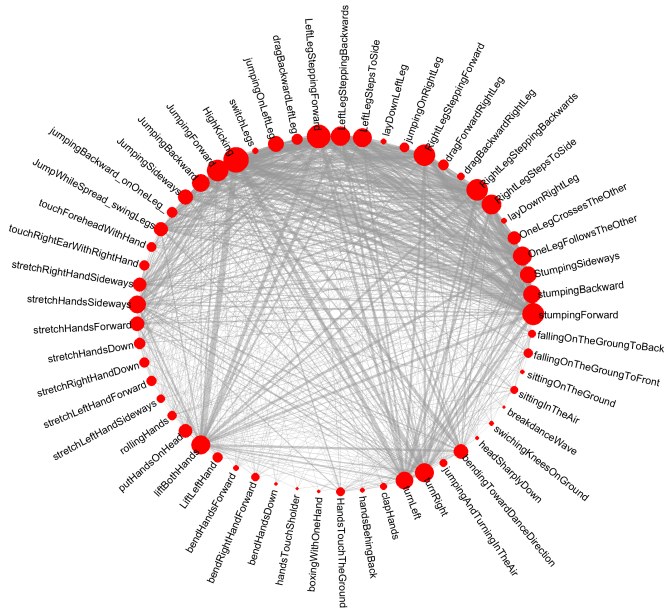


Figure 12: Event logging database annotated with Observer XT for ritualistic dance during **Zulu weddings** in two generations with male/female participants.

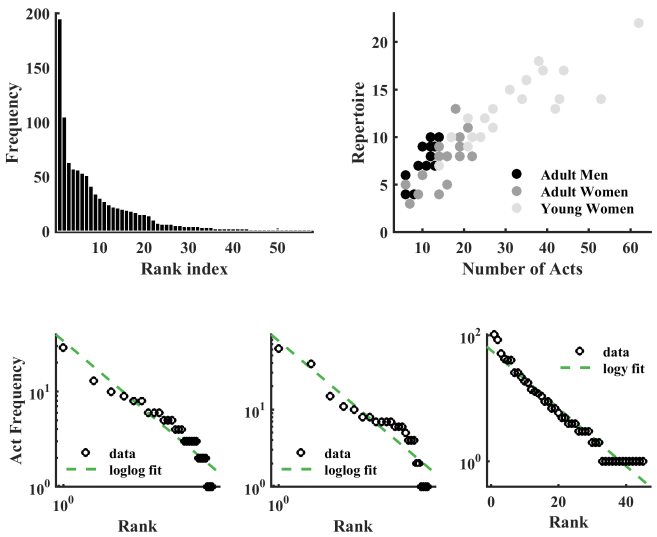


Figure 13: Movement inventory follow Heaps' and Zipf's laws known from other **communicative systems**.

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

Nielbo, K. L., Fux, M., Mort, J., Zamir, R., & Eilam, D. (2017). Structural differences among individuals, genders and generations as the key for ritual transmission, stereotypy and flexibility. Behaviour, 154(1), 93–114.

THANK YOU

knielbo@sdu.dk
knielbo.github.io

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

& credits to

Max R. Echardt and Katrine F. Baunvig, datakube, University of Southern Denmark, DK
Jianbo Gao and Bin Liu, Institute of Complexity Science and Big Data, Guangxi University, CHN
Melvin Wevers, DH Lab, KNAW Humanities Cluster, NL
Culture Analytics @ Institute of Pure and Applied Mathematics, UCLA, US

SDU