

from humanities computing to human analytics

analytics initiatives at university of southern denmark

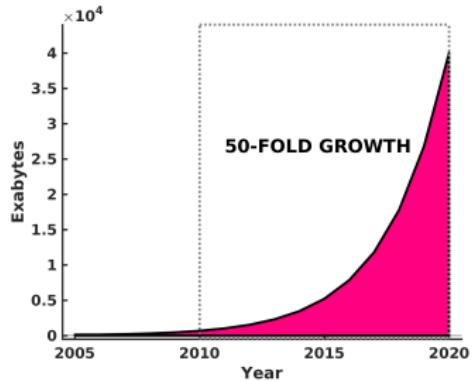
Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

datacube | Department of History | University of Southern Denmark

PROGRAM

0.00	digitization or transformation	a bit of history
0.10	human analytics at sdu	four modes of empowerment
0.35	applications	existing projects
0.40	discussion	...

DIGITIZATION or TRANSFORMATION



-
- the **data deluge** is transforming knowledge discovery and understanding in every domain of human inquiry
 - knowledge discovery depends critically on advanced computing capabilities
 - a large part of these data are unstructured and fundamentally cultural and normative

Strategi for Danmarks digitale vækst

Erhvervsministeriet

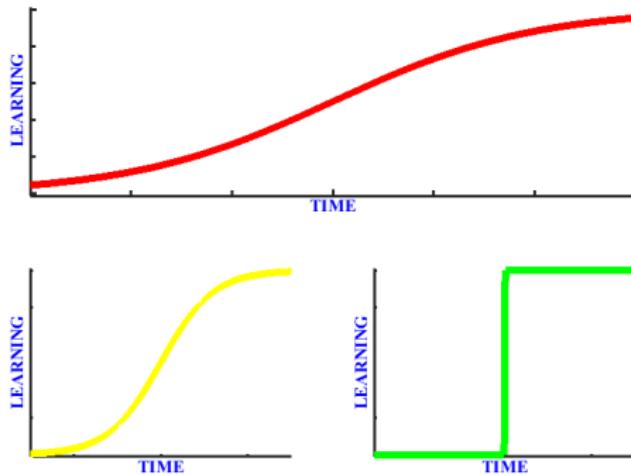


- typical response to the data deluge is **digitization**
- instead we should be focusing on a domain sensitive analytics **transformation** of research
- but the majority of the humanities lacks computational and data literacy



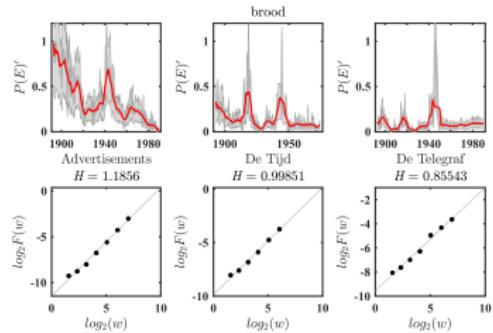
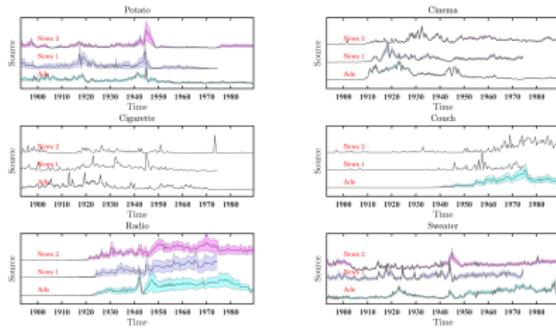
Human Analytics

empowerment of humanities domain knowledge through development and application of large scale data analysis for human history and culture



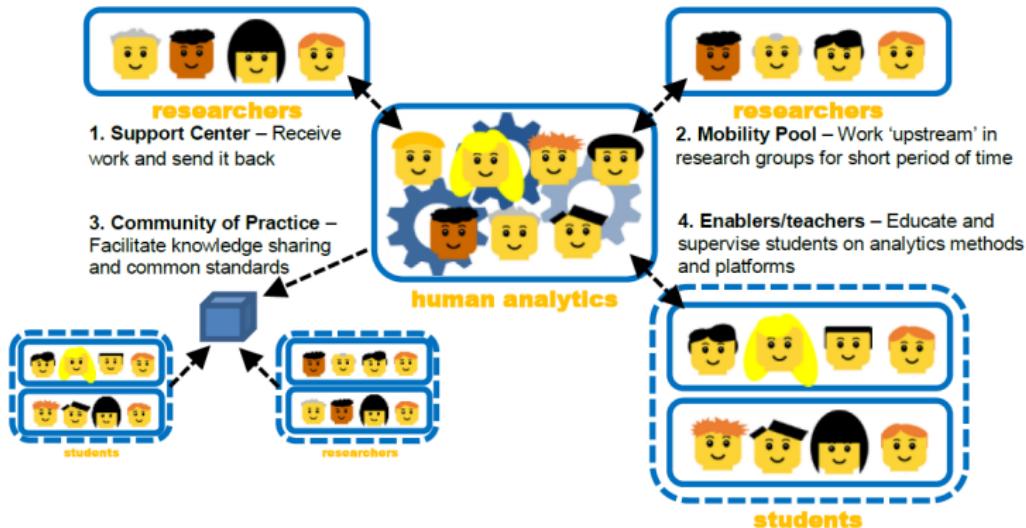
- in order to build a **culture of analytics**, every research institution has to “break” the learning curve
- at SDU|Humanities we have established a small research and development team that is affiliated with all departments and the eScience center

history case|predictive causality & slow decay



- historians and media researchers theorize about causal dependencies between public and advertisement discourses
- time series analysis 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 intermittent behavior) than public discourse (long-range dependencies)

HUMAN ANALYTICS at SDU



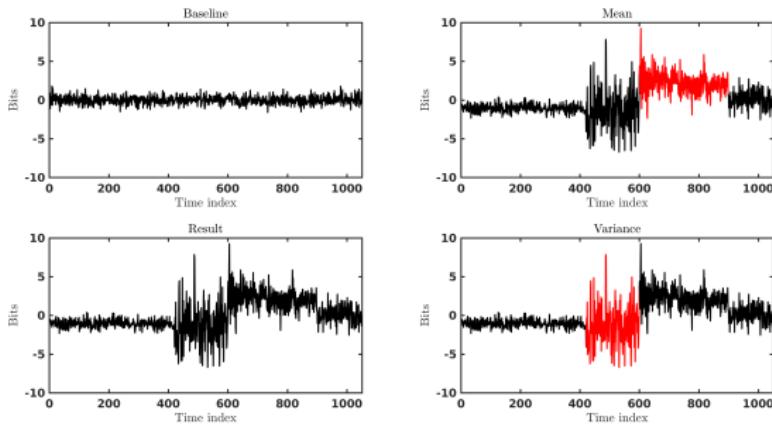
-
- our **vision** is an analytics team that has four modes of empowerment: support, mobility, community of practice, enablers
 - currently we have ~ 2.5 FTEs: Max R. Eckardt (development), Katrine F. Baunvig (outreach), Kristoffer L Nielbo (development and management), and Zehui Li (student programmer)



research support

-
- support for research design, data management, and analysis, PROJECTS, FUNDING APPLICATIONS
 - *first come, first served* with constraints → SCALE, IMPACT, RE-USE
 - business model: CO-AUTHORSHIP and, depending on project, JOINT FUNDING
 - current focus is NATURAL LANGUAGE PROCESSING, INFORMATION RETRIEVAL, but if you need other skills, we will acquire them

media studies example|novelty detection

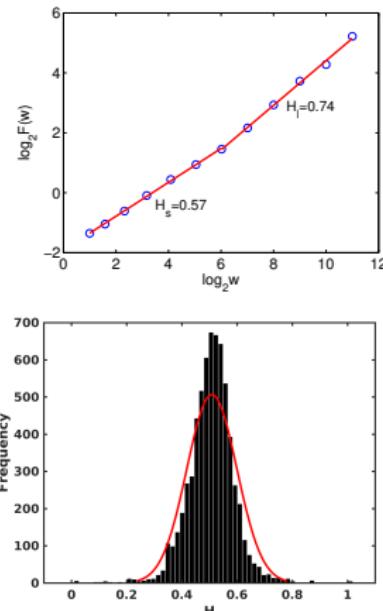
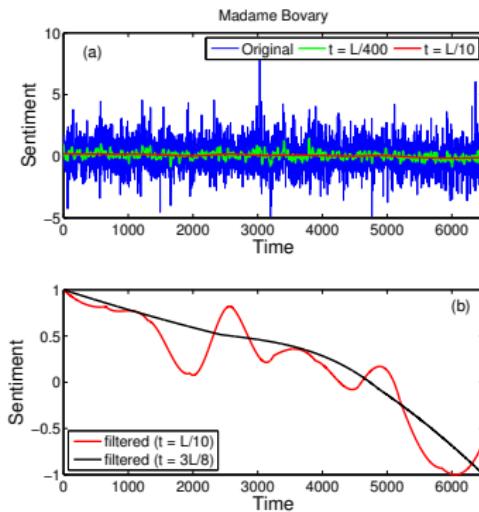


-
- change point detection in topicality space applies to “a change in the media tone”
 - train model on 200 years of newspapers in a comparative study between DK and NL
 - collaboration between historians, media studies and information science with a predictive scope

mobility pool

- work in research groups for a short period of time using the same model as general research support
- offer research commons at Odense campus currently, if you need to work together in a physical computer laboratory
- supply the necessary student programmers, PhD students or postdocs for your project

literary studies example|sentiment analysis



-
- dictionary-based sentiment analysis can reconstruct narrative/plot vectors that reflect human reading
 - basic insights from structural linguistics and narratology can be captured by this approach
 - a particular scaling-range, $0.6 < H \leq 0.8$, indicates literary optimality



community of practice

-
- all things being equal, our tools that are open and available to all researchers at SDU
 - develop apps for SDU cloud both generic and tailored to specific projects
 - Explorative Document Operator

enablers

- we are not an educational unit, but do offer workshops for researchers and students
- currently we offer an NLP workshop for PhD students and a study group on Humanities Programming
- contribute with educational modules to existing courses
- members of Software & Data Carpentry and are currently developing lessons targeting humanities

APPLICATIONS of HUMAN ANALYTICS

Human Analytics

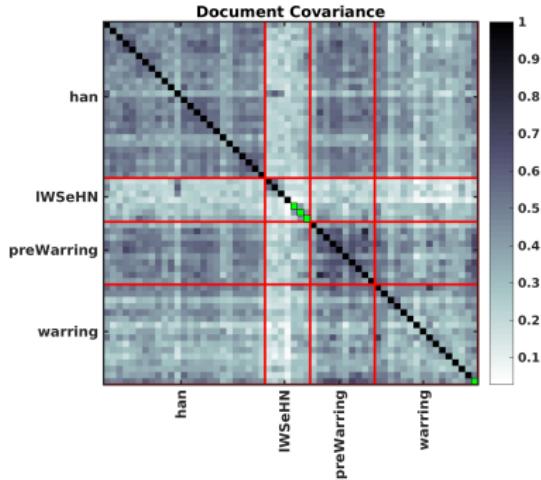
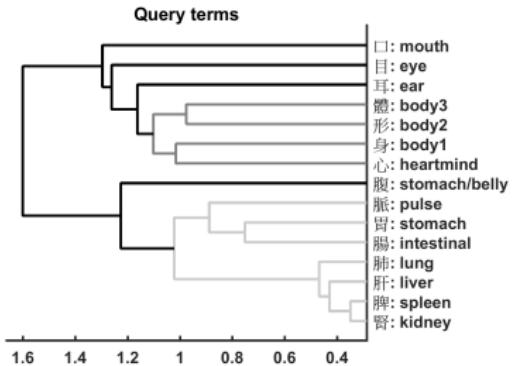
in comparison to analytics proper

- descriptive not predictive
- neither side of the interdisciplinary divide is conceptualized as service
- preference for micro-scale analysis
- predominantly unstructured data
- low-resource varieties/historical perspective (cultural heritage data)
- reliance on qualitative assessment (e.g., hyper-parameters and validation procedures)

... to similar trends (e.g., culturomics, cliodynamics, culture analytics)

- multi-scale/multi-resolution
- data-intensive ethos (scalability matters)

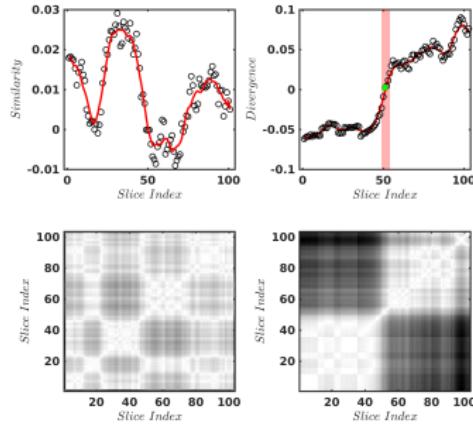
Philosophy|Latent Semantic Variables



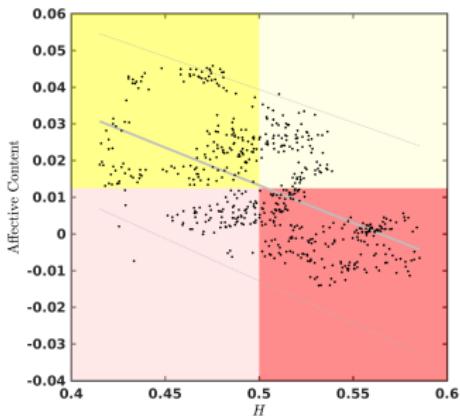
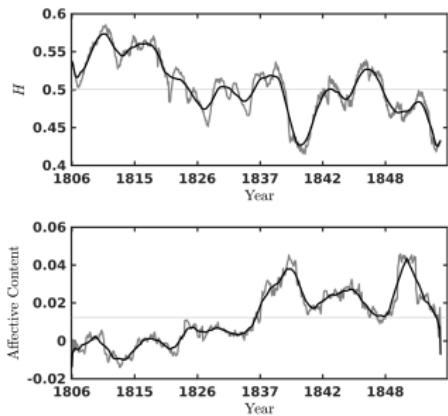
-
- philosophers and sinologists have been debating the existence of mind-body dualism in classical Chinese philosophy
 - with domain experts, unsupervised learning was used to identify a multi-level dualistic semantic space
 - one model (LDA) was further utilized to predict class of origin for controversial texts slices

Medieval History|novelty detection

- historians debate historical transitions
- Saxo's *Gesta Danorum* c. 1200 AD.
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 focusd
in book 9



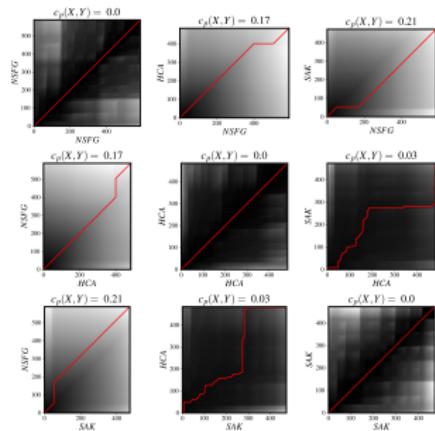
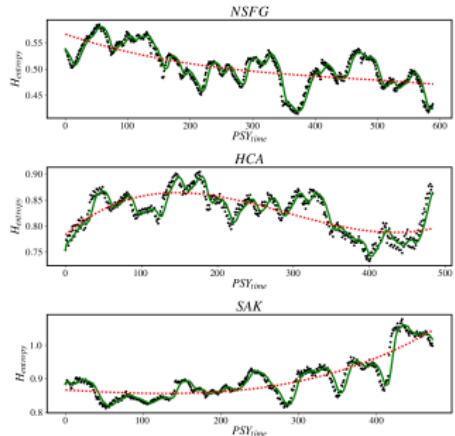
Literary History|Lexical Density



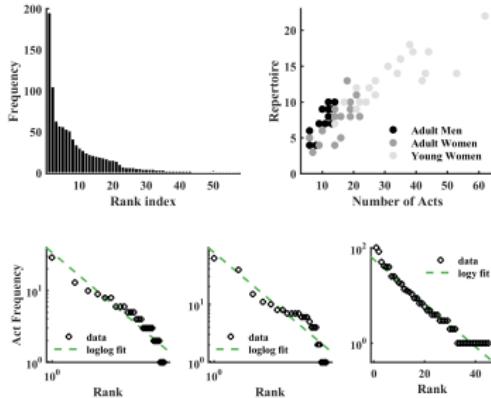
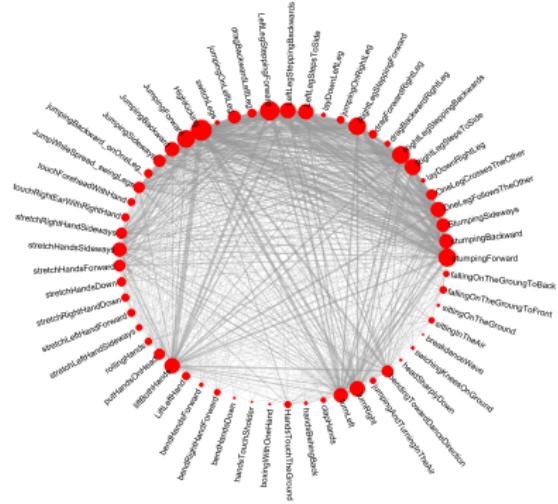
Literary scholars and creativity researchers argue for the “tortured artist”

- “writers’ creative state is inversely related to their emotional state”
- “writers’ creative state *depends* on their emotional state”
- look for dependencies in lexical density and sentiment scores for highly prolific writers to identify state incongruences

Literary History|Sequence Alignment



-
- there is a new biographical trend in literary history
 - using lexical density and sequence alignment, we can compare creative trajectories of authors



-
- anthropologists discuss why rituals appear rigid, while they seem to maintain behavioral variability
 - manual annotation of ritual dance applied to ethnographic video archives from multiple generations
 - very few behavioral units are transmitted between generations (compulsory), allowing for both flexibility and rigidity

THANK YOU

knielbo@sdu.dk

https://knielbo.github.io/files/kln_ha_021918.pdf

& THANKS TO

Jianbo Gao, Melvin Wevers, & David Eilam