Cultural Disruption and Time-Dependency in Systems of Symbolic Production

CAARE, London 2016

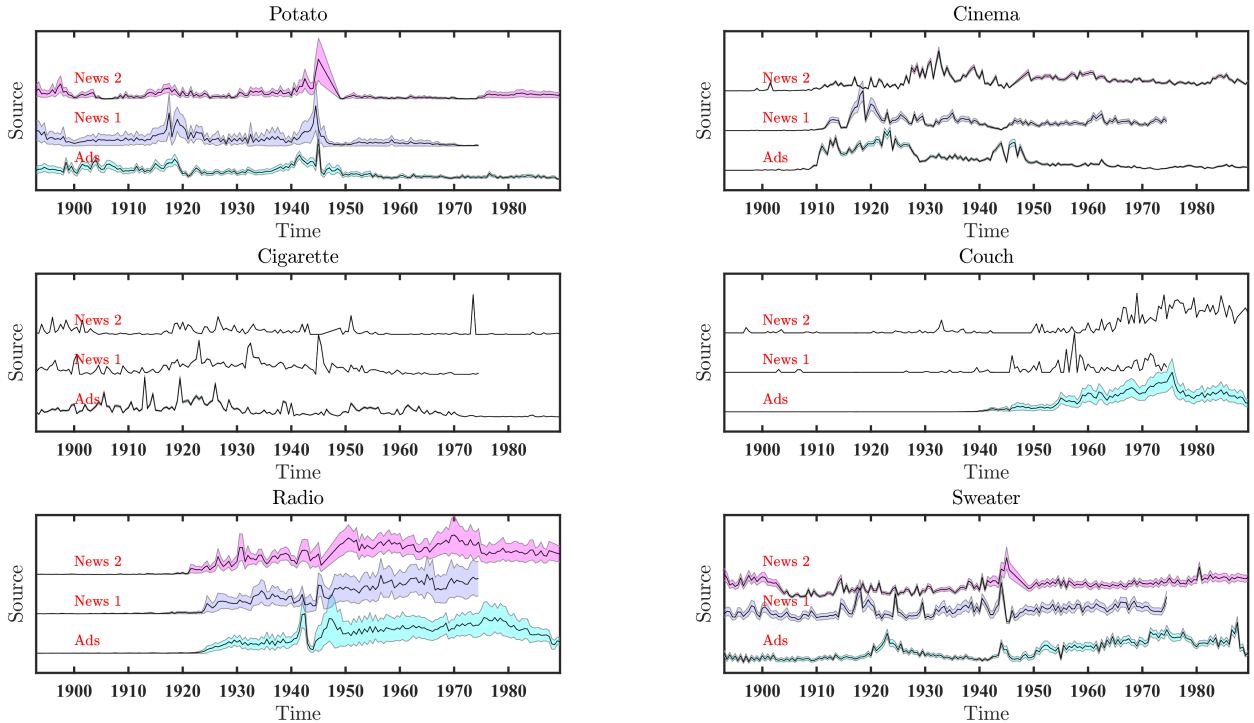
Kristoffer L Nielbo
kln@cas.au.dk

DIGITAL TEXT LAB|INTERACTING MINDS CENTRE|AARHUS UNIVERSITY

IMC
INTERACTING MINDS CENTRE

Cultural Disruption and Time-Dependency in Systems of Symbolic Production

CHALLENGES to a valid study of human culture
– deep temporality|THEORY
– data surge|EMPIRICAL BASE
– black box solutions|METHODS implies THEORY

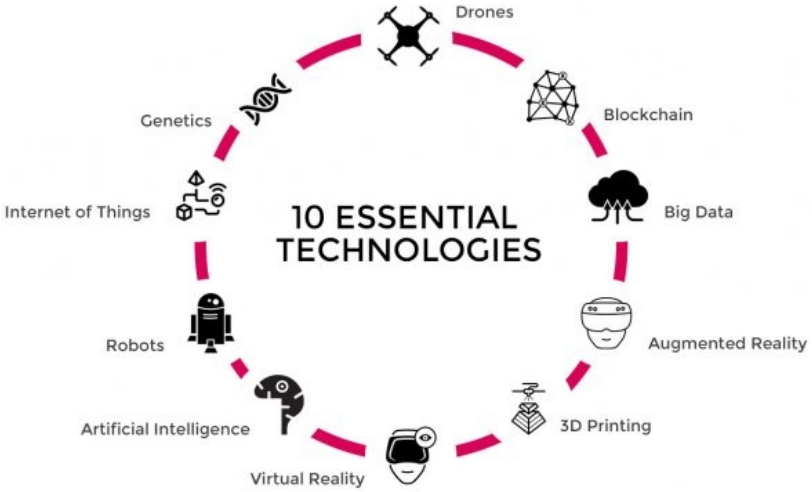


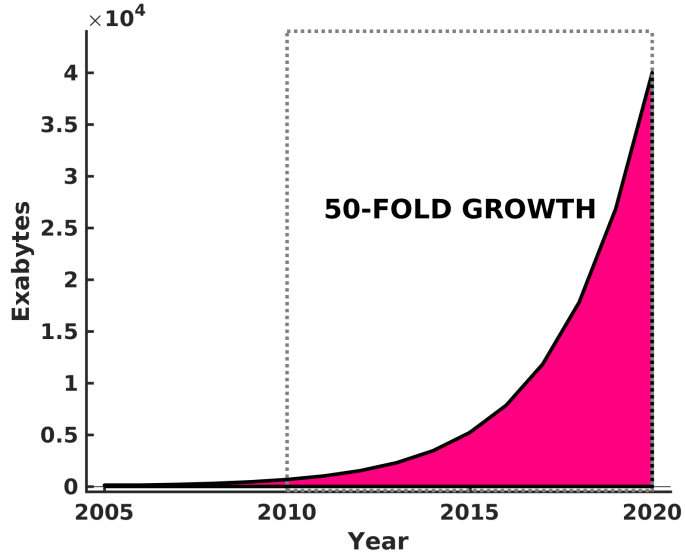
DEEP TEMPORALITY

– human culture is characterized by a species-unique capacity for long-term planning and future-oriented information processing

– fundamental features of human culture are revealed in (persistent) temporal dynamics (e.g., cultural memory)

$\models$ leverage tools from non-linear time series analysis



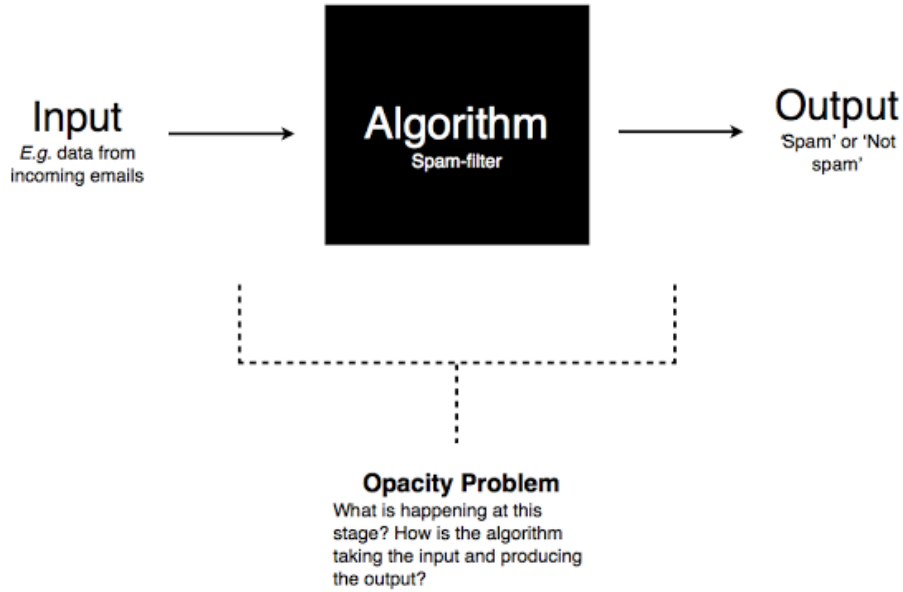


DATA SURGE

– digitization and digital media have generated a rapid proliferation of data that is unprecedented in the history of man

– transforming knowledge discovery and understanding in every domain of human inquiry

$\models$ build tools that can learn from data and automatically make the right decisions



BLACK BOX

– the majority of algorithms available in machine learning libraries *black box* the solutions

– solve predictive tasks, $P(spam \mid email)$, but not conducive to the understanding of culture

$\models$ combine techniques from ML with a more transparent theoretical framework that allow for time series analysis

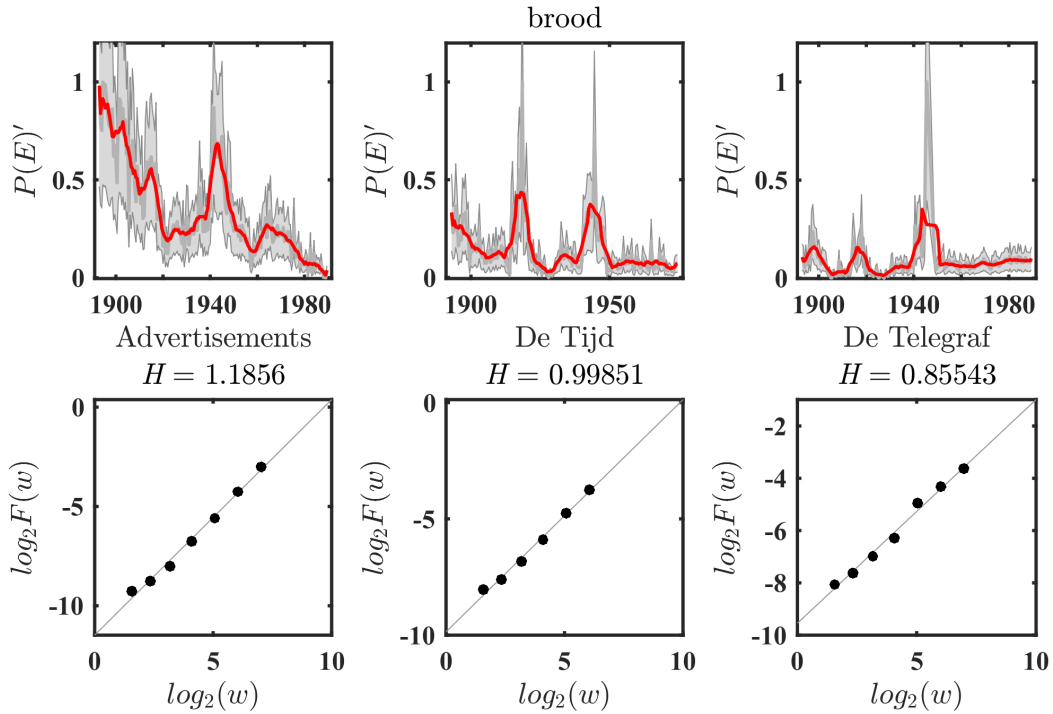
INFORMATION THEORY

– model culture as internal and external transmission of information $\Rightarrow$ rely on a well-established theoretical framework for quantifying information

$$H = -\sum_{i=1}^{n} p_i \times \log_2(p_i)$$

$$p_i = \frac{Fr(w_i)}{\sum_i^n Fr(w_i)}$$

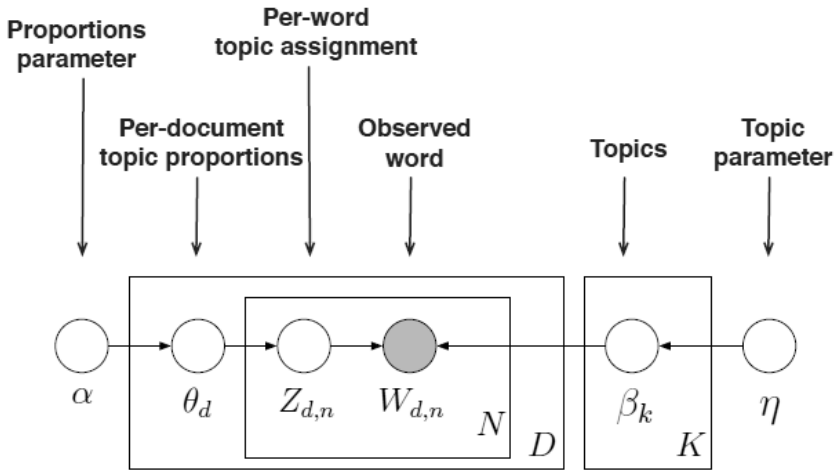
– measures average amount of information in a text string as the (un-)predictability of the string in terms of its units (e.g., characters, words, ngrams)



LONG RANGE DEPENDENCIES

– Hurst exponent is an index of the degree of persistence/long-term memory in time series (trends exiting over a long time scale)

– $H_{exp} = 0.5$ indicates lack of long-term memory, $0.5 < H_{exp} \leq 1$ indicates persistent behavior (i.e., long-term memory), $H_{exp} < 0.5$ indicates anti-persistent behavior



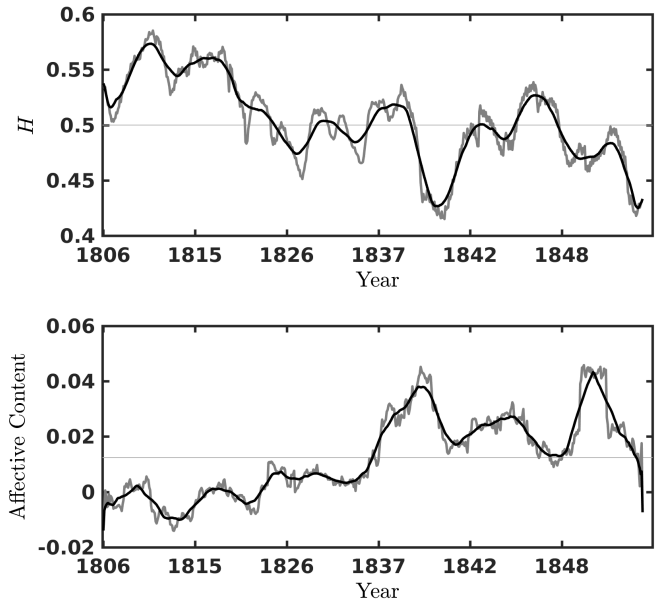
LATENT DIRICHLET ALLOCATION

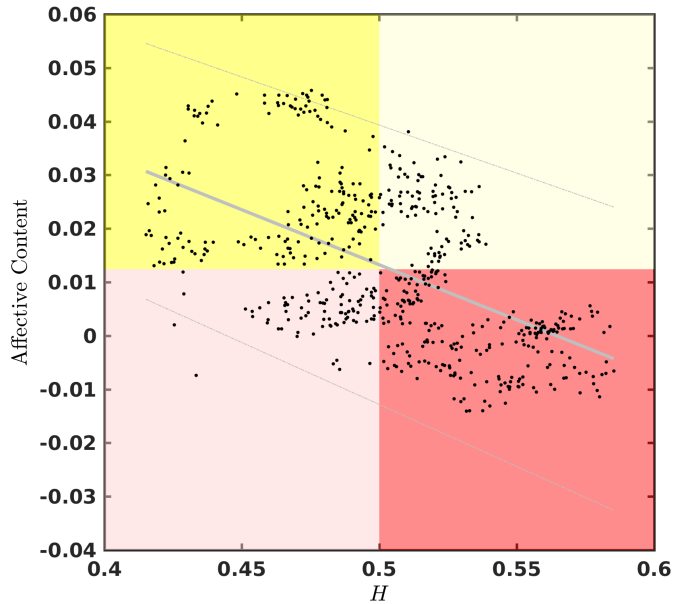
– culture is not only about information, it is about semantics ⇒ use a simple Bayesian mixed model to capture semantics at the lexical level

– model each document as a distribution on lexical topics (e.g., $P_1 = [0\ .09\ .78\ .11\ .2]$), where each 'topic' is a distribution on words, and compare document similarity

$$D_{KL}(P \parallel Q) = \sum_{i=1}^{n} P(i) \times \log_2 \frac{P(i)}{Q(i)}$$

– bracket concrete semantics (∼ reduce interpretive load) and only compare relative documents on topics (need to validate the model)



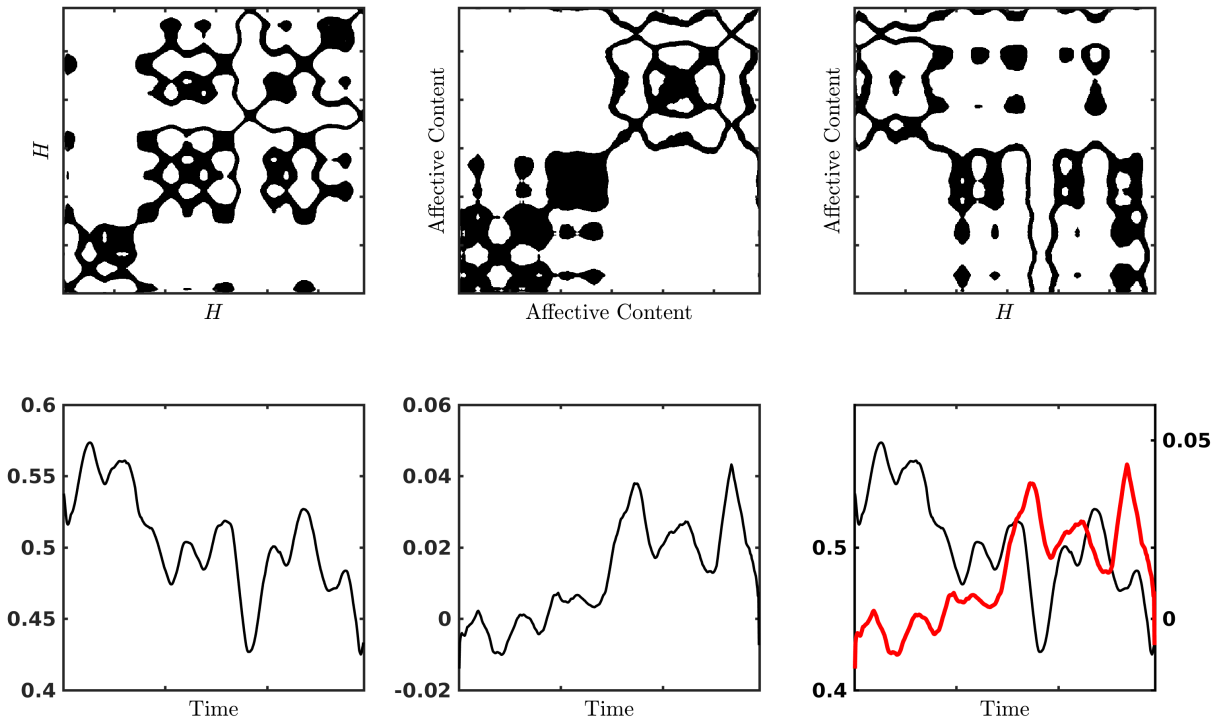


PERSISTENCE in INDIVIDUAL COGNITION

– contours of cognitive profile (creativity and memory) of specific author

– use information to reconstruct the cognitive and affective mind of NFS Grundtvig

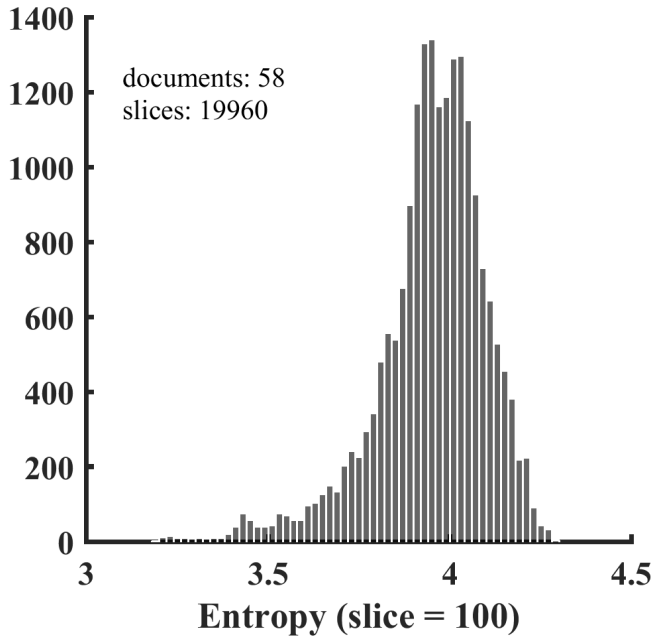
IMC
INTERACTING MINDS CENTRE

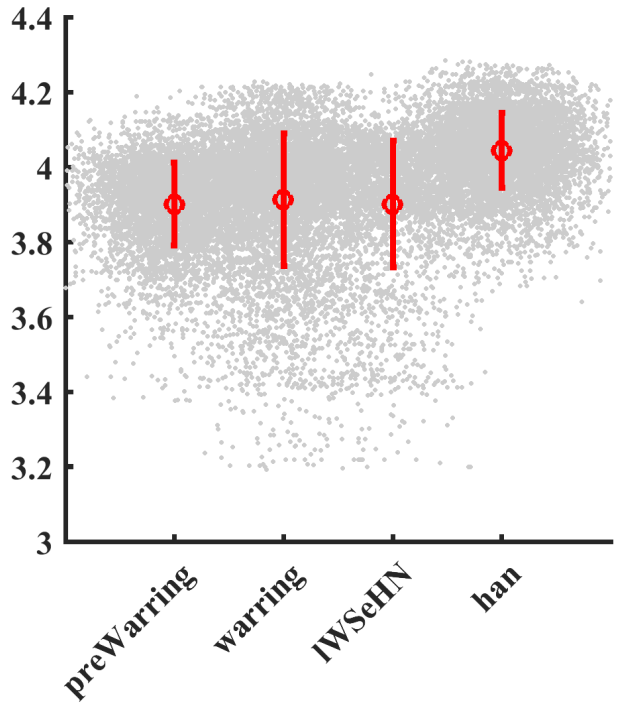


INTERDEPENDENCIES and PREDICTIVE CAUSALITY

– coupling between creativity and affective content

– affective content *granger-cause* creativity



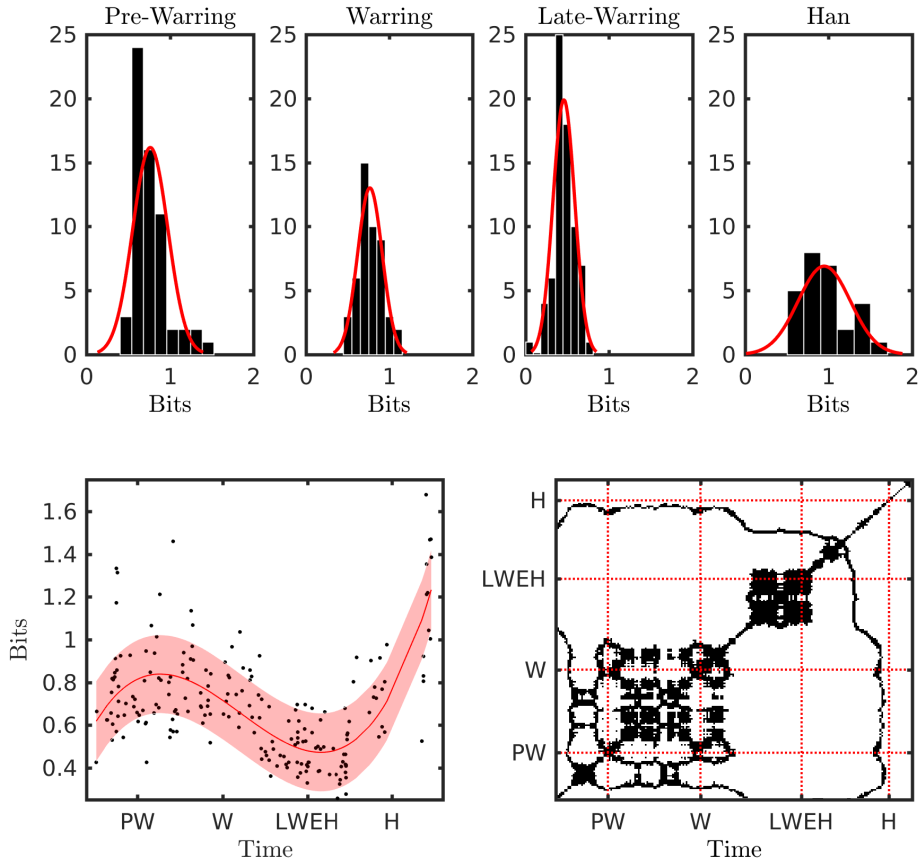


LEXICAL DENSITY in SYMBOLIC PRODUCTION

$$H(\text{a rose is a rose is a rose}) < H(\text{a rose is red and thorny})$$

$$H(\text{a rose is a rose is a rose}) = H(\text{a a a is is rose rose rose})$$

$$H(\text{a rose is a rose is a rose}) = H(\text{erea oisasessar oiors})$$



DISRUPTION in COLLECTIVE COGNITION

– trace evolution of ideas in classical Chinese literature

– use information to date controversial chapters of the Shangshu

THANK YOU

kln@cas.au.dk