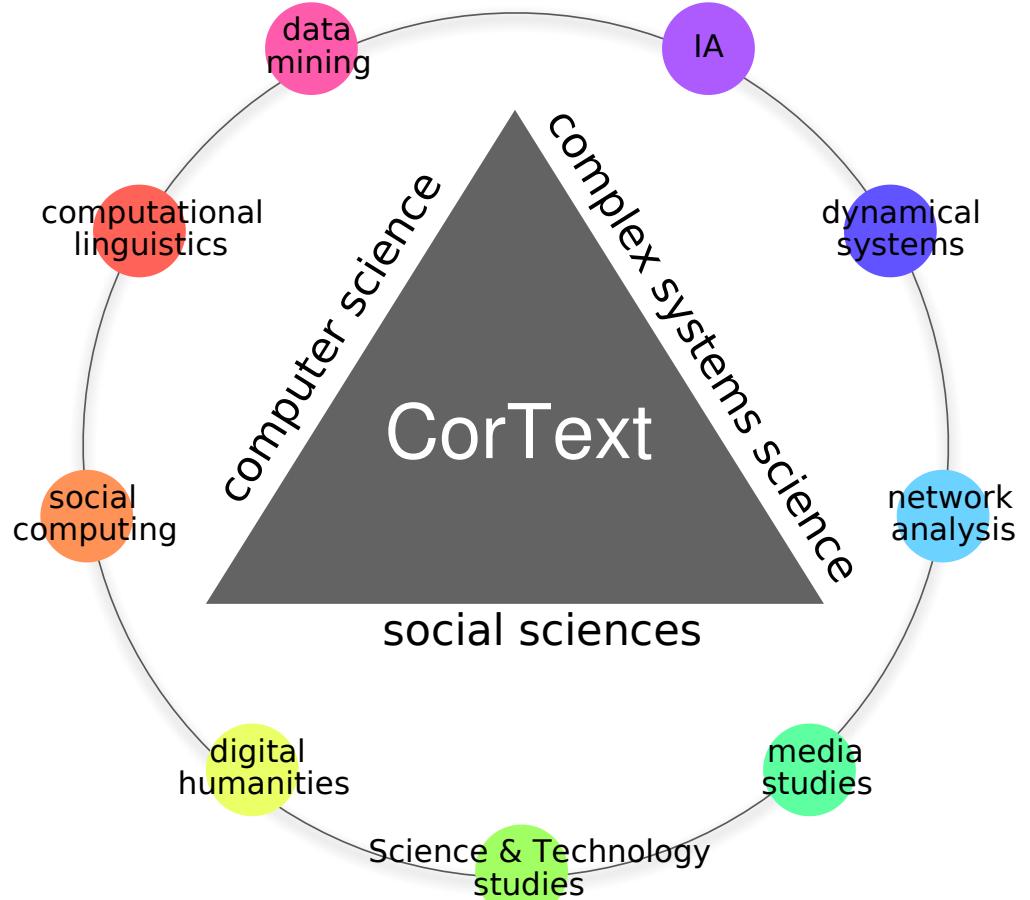


Análise automática de documentos
em ciências sociais, da saúde e humanidades:
progressos e ferramentas

Automated document analysis
for social and health sciences and humanities:
progresses and tools

CorText

- A web based digital platform for social scientists conducting *empirical studies* in the fields of Media Studies, Science and Technology Studies, Digital Humanities, etc.
- CorText is an open online service for *heterogeneous data analysis, modeling and visualization*
- open registration:
<http://managerv2.cortext.net>



Abordagens

- Citation analysis
↓ Network (projections)
- Collaboration networks
- Co-word analysis
- Topic-models
↓ Vector (projections)
- Word embeddings
- SASHIMI
↓ Sashimi

Co-word: o contexto é que conta

« Somente a proposição tem um sentido ;
é apenas ligado a uma proposição que o nome tem um significado »

Wittgenstein, "Tractatus logico-philosophicus" 1921, prop 3.3.

« Uma palavra é caracterizada pela sua companhia »

Firth, John R. "A synopsis of linguistic theory, 1930-1955."
Studies in linguistic analysis (1957) - 11.

Perspectiva multi-nível

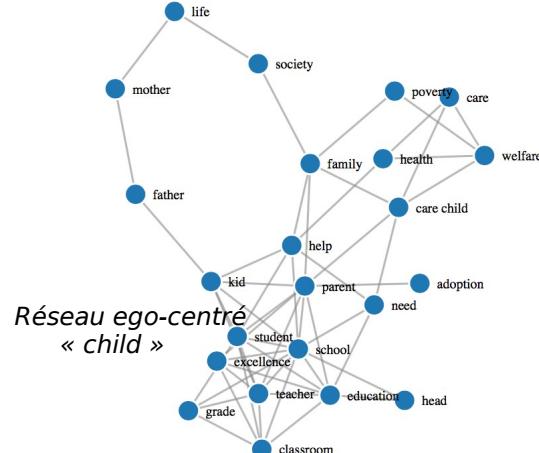
Textos

*Extraction
terminologique*

label	forms
abundance	abundance
acres	acres& acre
aliens	aliens& alien
armed forces	armed forces & armed force & force of arms & arms and force
business men	business men & business man & men of business & man of business & businessman
children	children& child
commerce and navigation	commerce and navigation & navigation and commerce & navigation or commerce
crews	crews & crew
crime	crime & crimes
crisis	crisis & crises
crops	crops & crop
cruisers	cruisers & cruiser
democracy	democracy & democracies
diplomatic relations	diplomatic relations
farm products	farm products & products of the farm & products of farm
great importance	great importance & greater importance & greatest importance
health care costs	health care costs & cost of health care & health care cost
income tax	income tax & tax the income & tax on the income
peace and freedom	peace and freedom & freedom and peace & peace with freedom
property rights	property rights & property right & rights of property & right of property & rights and property

Analyse textuelle

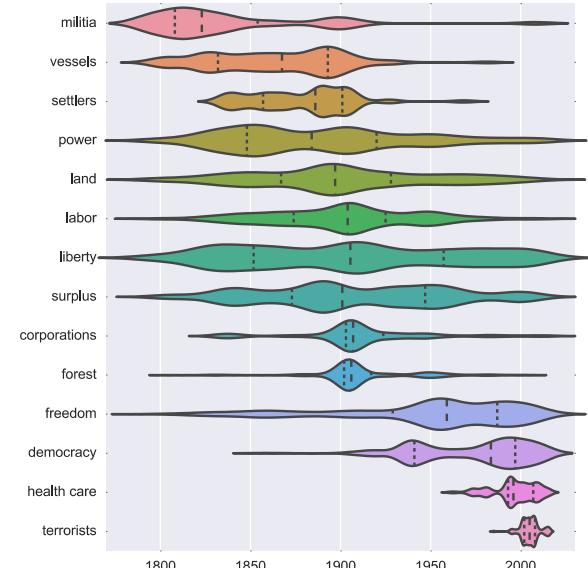
Relações



Structure locale

Dinâmicas

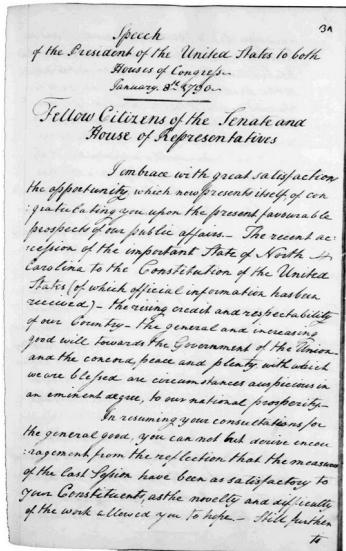
*Profils
d'évolution des
quelques termes*



Perspectiva multi-nível

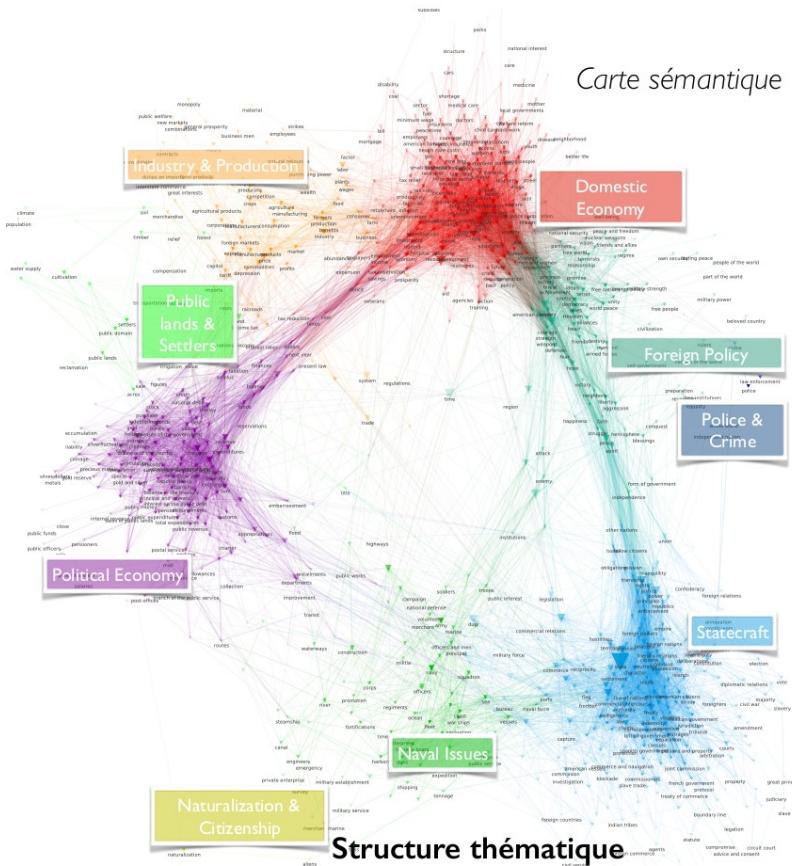
Textos

Discours de
Georges
Washington



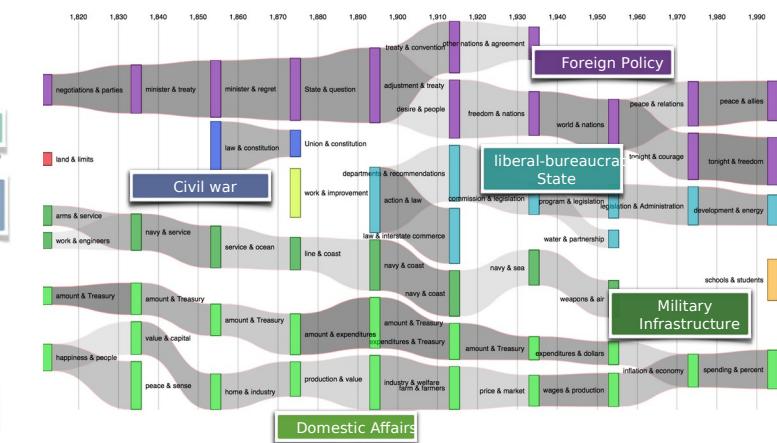
Documents

Relações



Dinâmicas

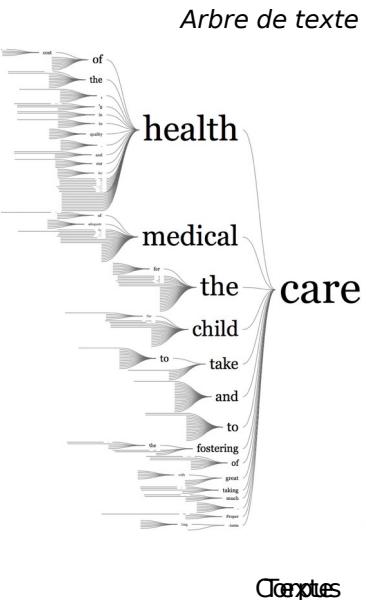
Diagramme alluvial
de l'évolution des
thématisques



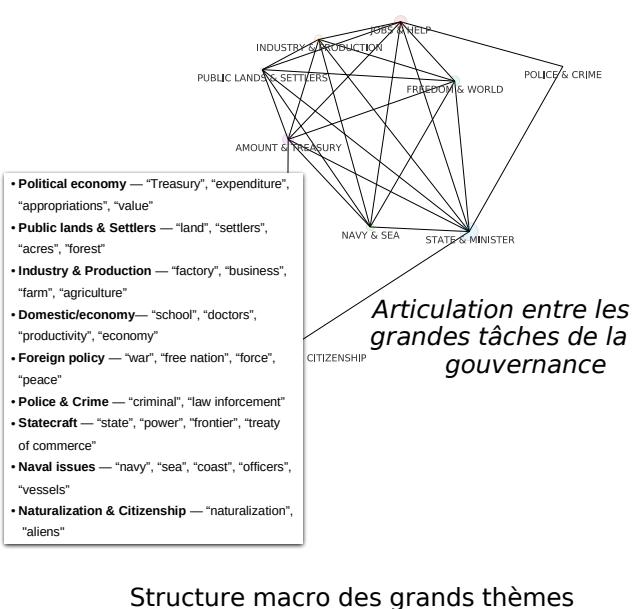
Dynamique des grandes thématiques

Perspectiva multi-nível

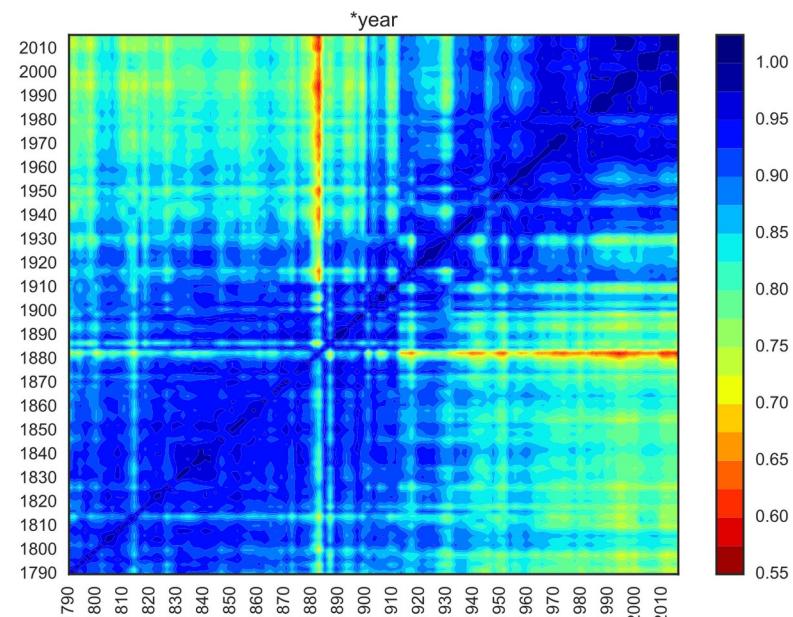
Textos



Relações



Dinâmicas



Occurrence matrices

- *Occurrence matrix O : $O_{ij} = 1$ iff item i is used in document j , 0 otherwise*
- The cooccurrence matrix enumerates every joint appearances of two items in the same document. $C = O^t O$

documents items \ documents	A	B	C	D
C1		1		
C2	1	1		
C3			1	1
C4	1	1		
C5				1

Cooccurrence matrices

- Occurrence matrix $O : O_{ij} = 1$ iff item i is used in document j , 0 otherwise
- The cooccurrence matrix enumerates every joint appearances of two items in the same document. $C = O^t O$

	C1	C2	C3	C4	C5
C1	x	1		1	
C2	1	x	1	2	
C3			x		1
C4	1	2		x	
C5			1		x

Cooccurrence matrices

Vamos dar um pulinho
numa outra apresentação...

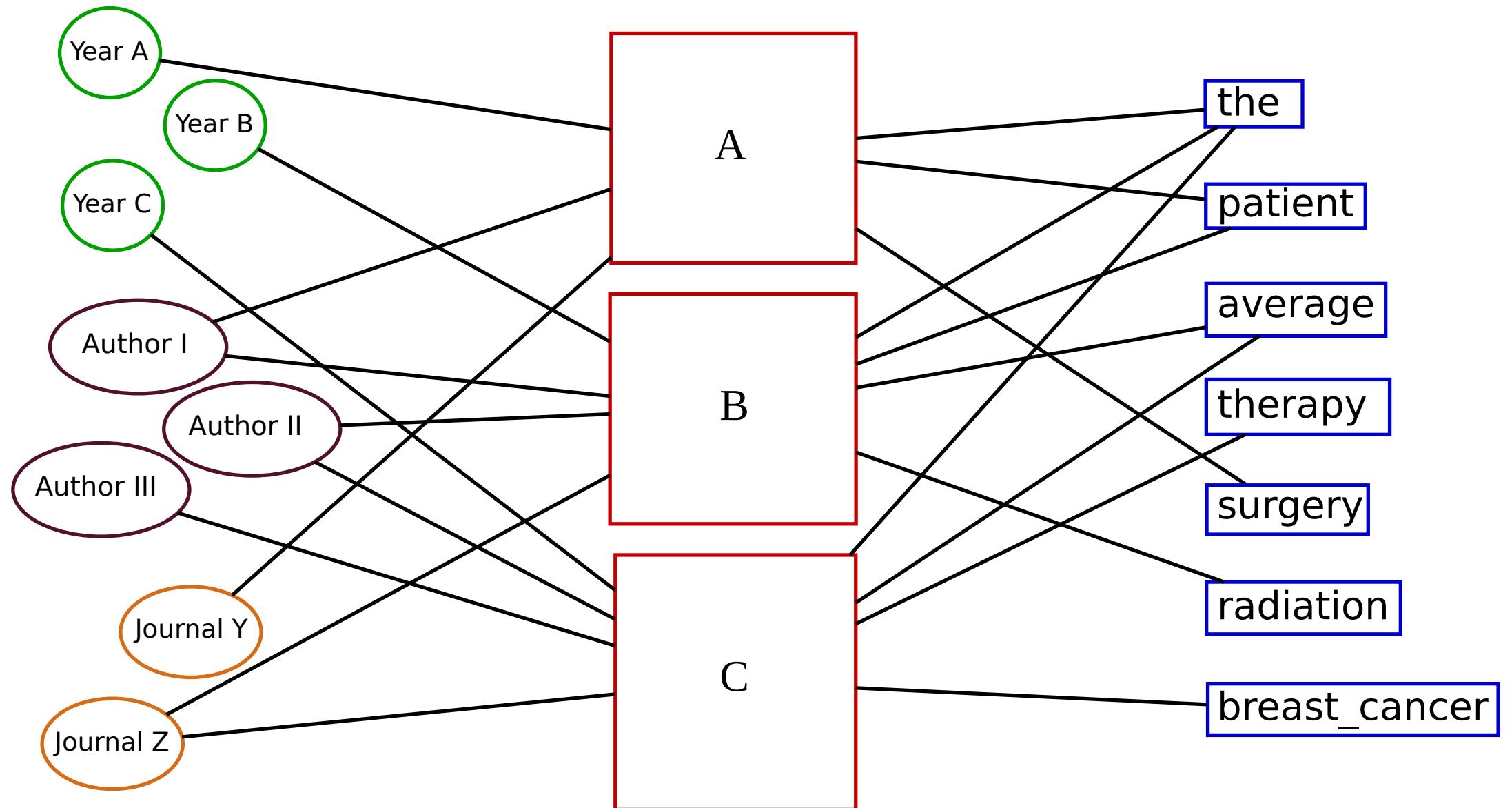
<https://gitlab.com/cortex/workshops/2017-06-marne/-/blob/master/presentations/cortex-training-2017.pdf>

Occurrence matrices

- Occurrence matrix $O : O_{ij} = 1$ iff item i is used in document j , 0 otherwise
- The cooccurrence matrix enumerates every joint appearances of two items in the same document. $C=O^T O$

documents \ items	A	B	C	D
C1		1		
C2	1	1		
C3			1	1
C4	1	1		
C5				1

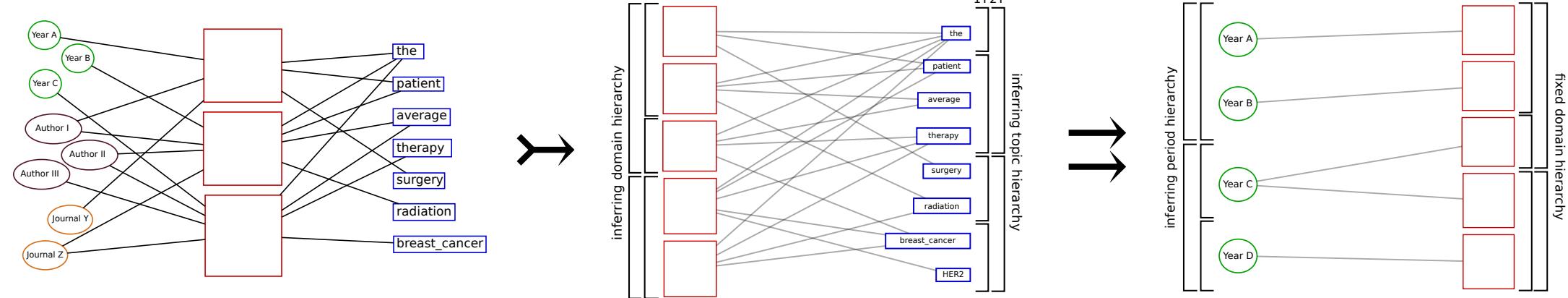
Document hypergraph



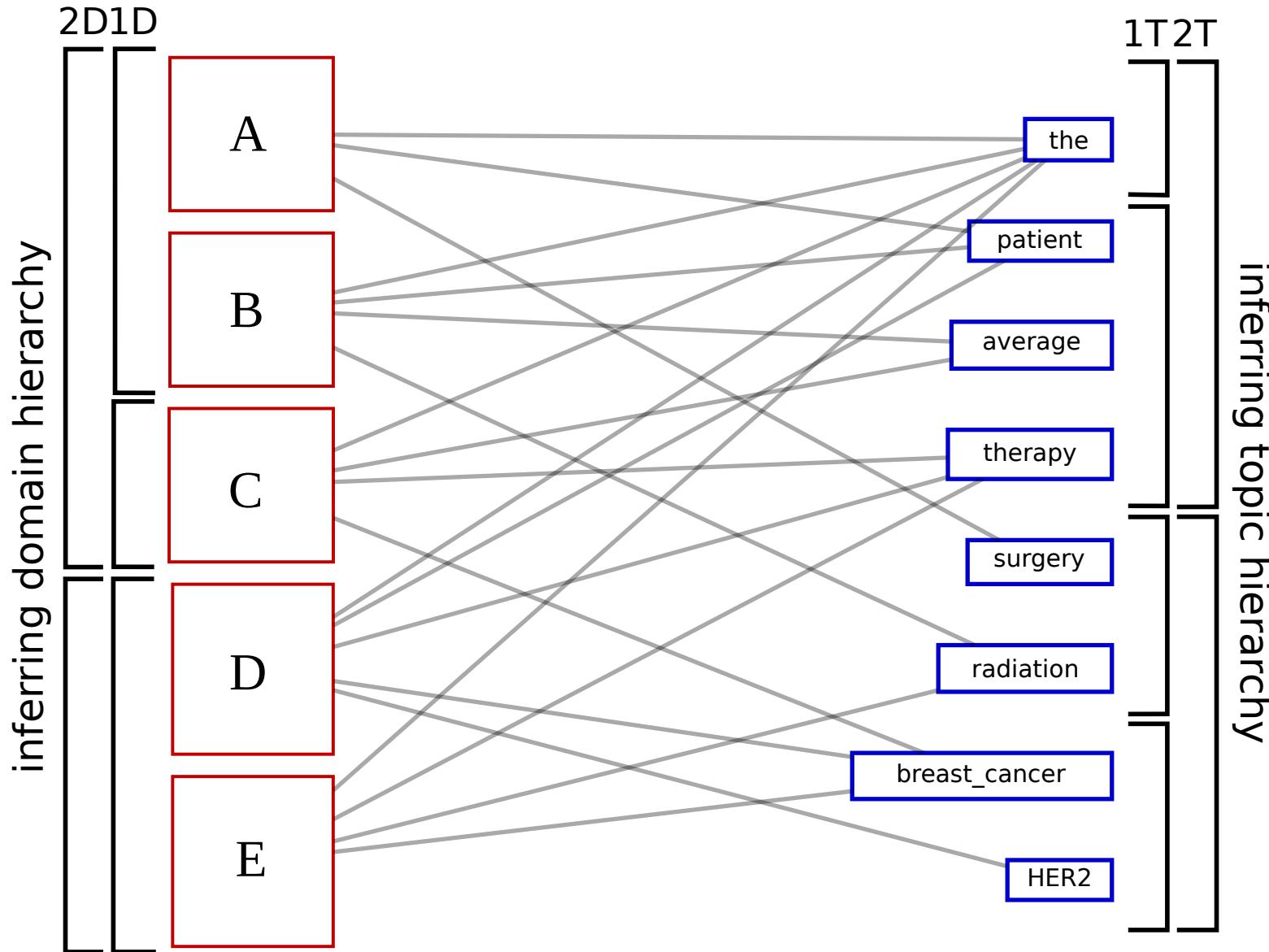
E que tal um pulinho
num capítulo de livro...

<https://arxiv.org/abs/1705.10225>

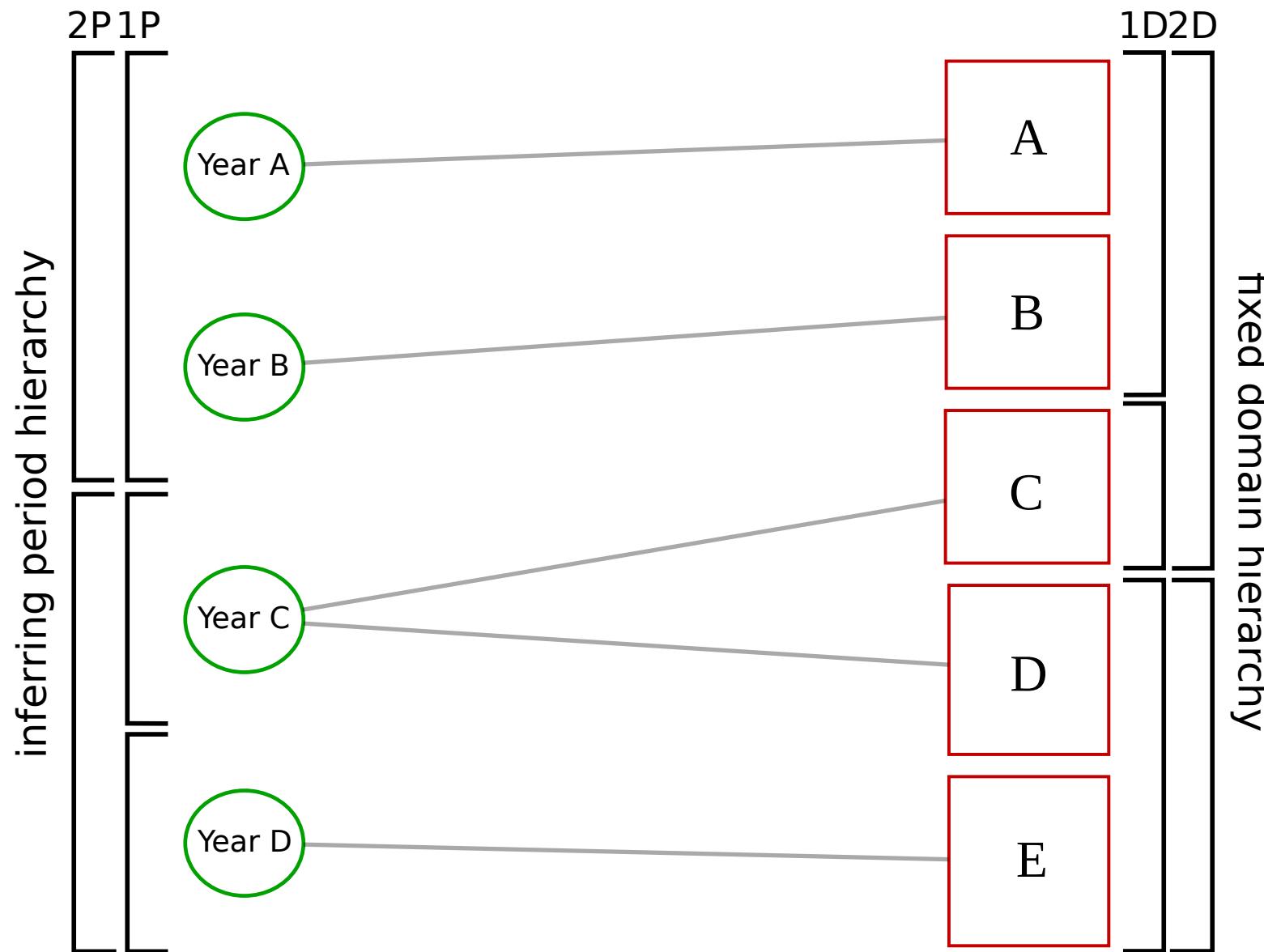
$n\text{SBM} \not\leftrightarrow \text{MDL} \otimes \text{Hypergraphs}$



Domain-topic model



Chained model



E que tal uma última escapada
para um preprint...

<https://arxiv.org/abs/1912.13349>

Obrigado!

Ale Abdo

LISIS – Laboratoire Interdisciplinaire Sciences Innovations Sociétés

CorText – <https://cortext.net/>

E-mail : abdo@member.fsf.org

Ni!

Salvo pelos gráficos das páginas 2, 4, 5 e 6 esta apresentação se distribui sob licença CC-BY.