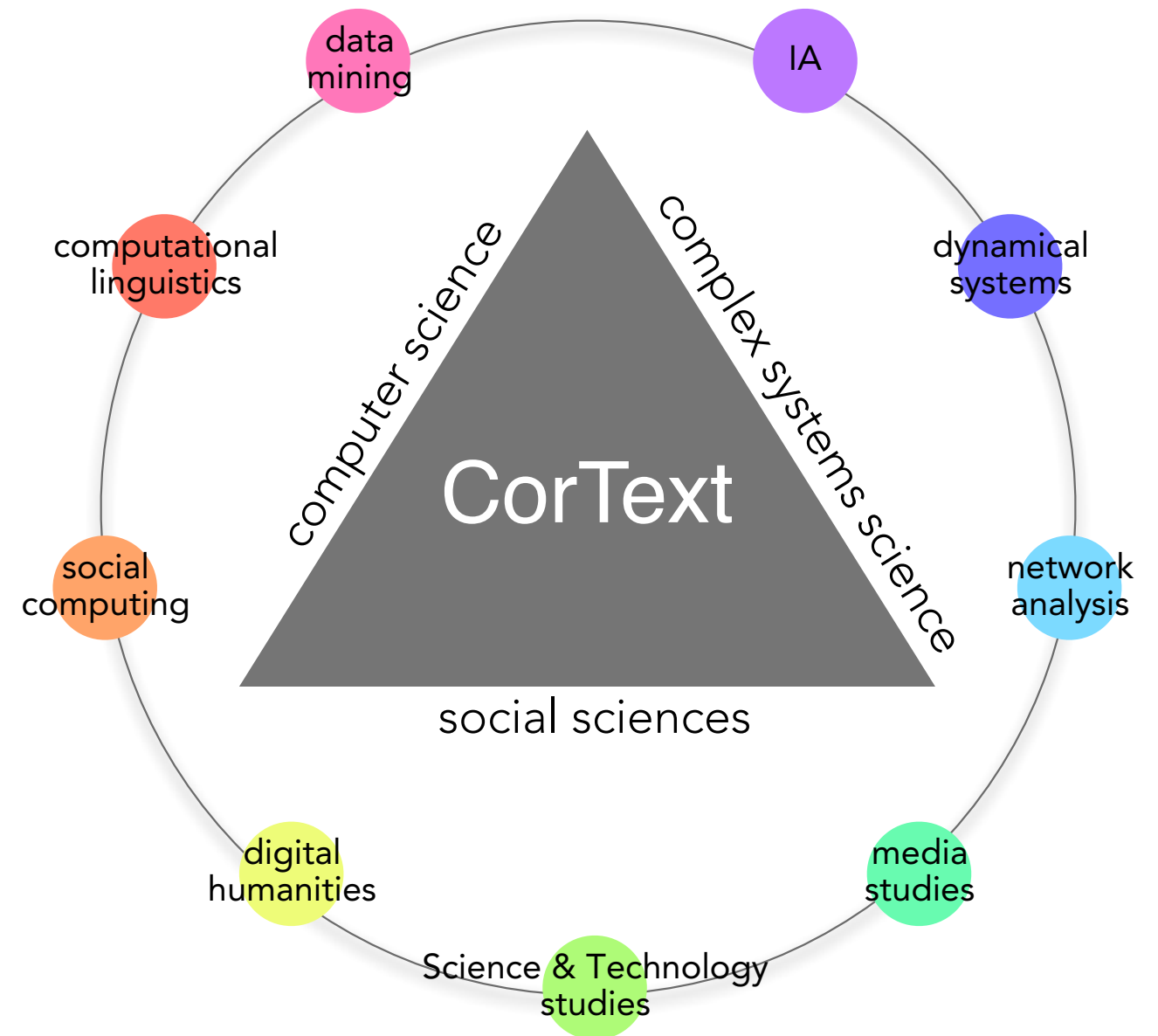


Modeling and mapping knowledge dynamics with CorText platform

Jean-Philippe Cointet
LISIS (INRA) - médialab (Sciences Po)

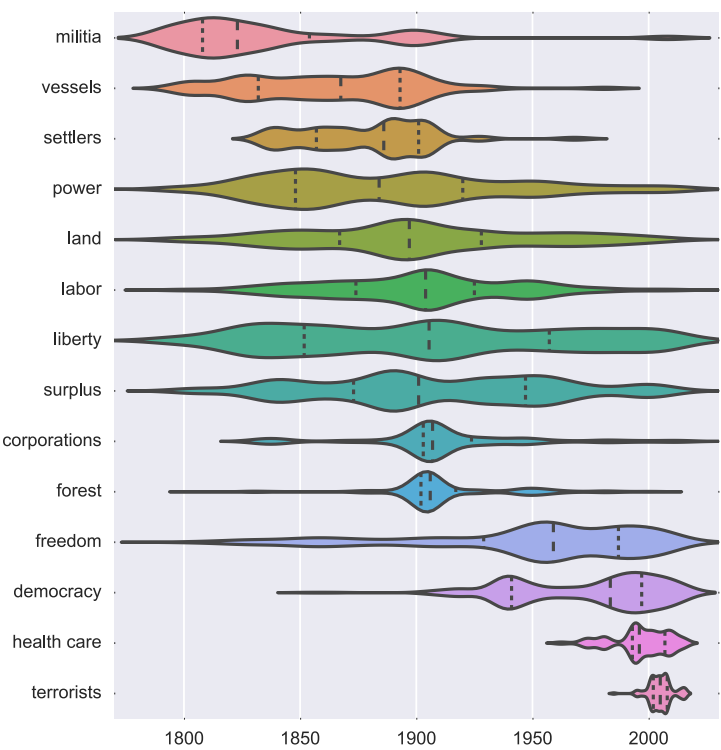
Outline

- 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>



Multi-level perspective

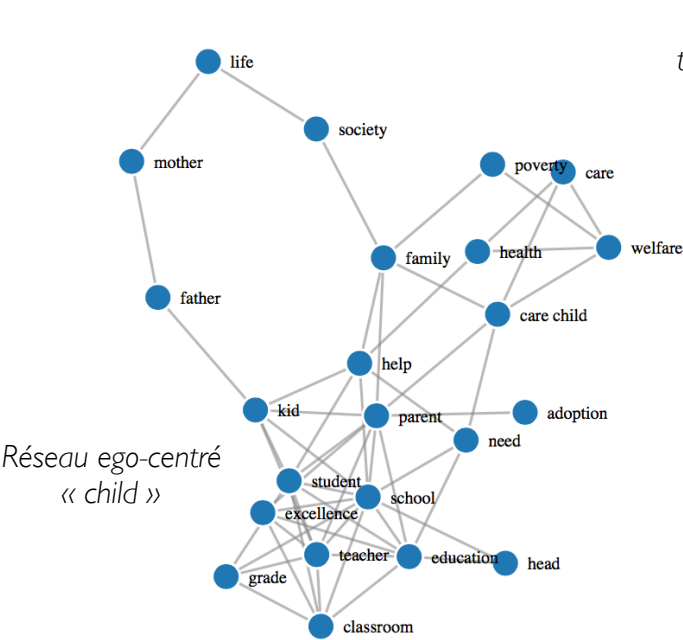
Dynamics



Dynamique des mots

Profils
d'évolution des
quelques termes

Relationships



Réseau ego-centré
« child »

Extraction
terminologique

label	forms
abundance	abundance
acres	acres&lacre
aliens	aliens&lalien
armed forces	armed forces&larmed force&lforce of arms&larms and force
business men	business men&lbusiness man&lmen of business&lman of business&lbusinessman&lbusinessmen
children	children&lchild
commerce and navigation	commerce and navigation&lnavigation and commerce&lnavigation or commerce
crews	crews&lcrew
crime	crime&lcrimes
crisis	crisis&lcrises
crops	crops&lcrop
cruisers	cruisers&lcruiser
democracy	democracy&ldemocracies
diplomatic relations	diplomatic relations
farm products	farm products&lproducts of the farm&lproducts of farm
great importance	great importance&lgreater importance&lreatest importance
health care costs	health care costs&lcost of health care&lhealth care cost
income tax	income tax<he income<he income
peace and freedom	peace and freedom&lfreedom and peace&lpeace with freedom
property rights	property rights&lproperty right&lrights of property&lright of property&lrights and property

Analyse textuelle

Structure locale

ECHELLE MICRO

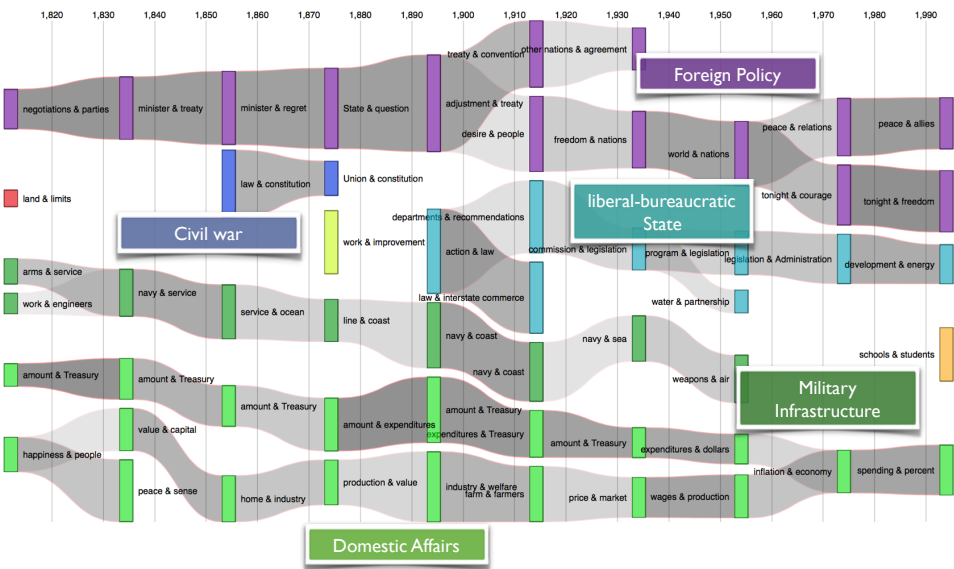
Multi-level perspective

Dynamics

Relationships

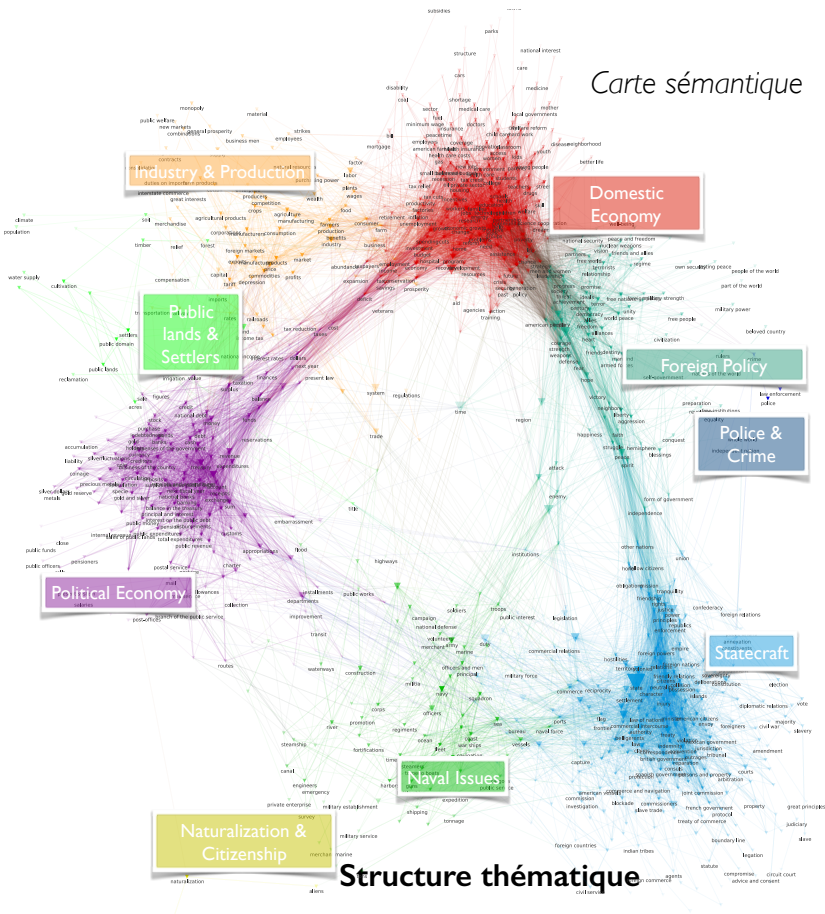
Texts

Diagramme alluvial
de l'évolution des
thématiques



Dynamique des grandes thématiques

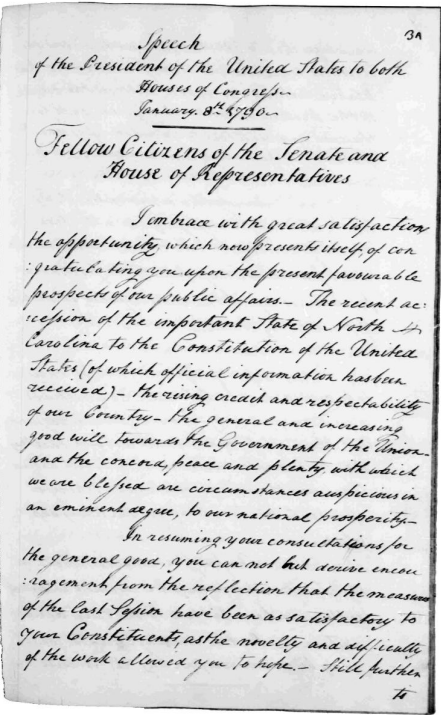
Carte sémantique



Structure thématique

ECHELLE MESO

Discours de
Georges
Washington



Documents

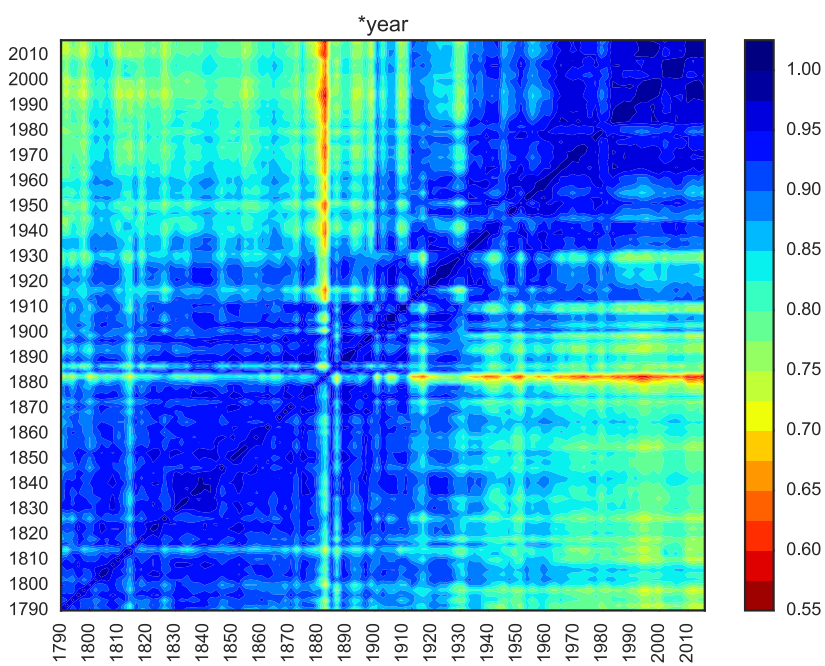
Multi-level perspective

Dynamics

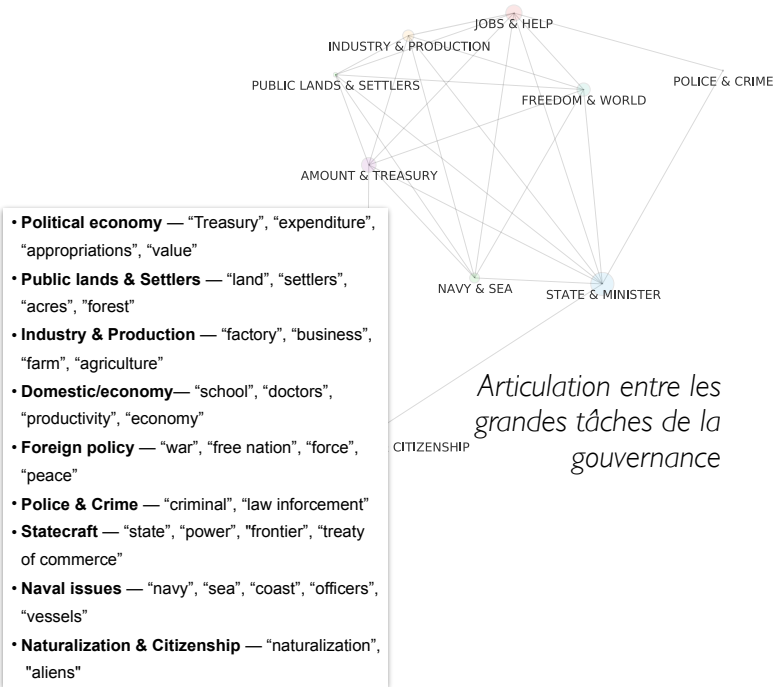
Relationships

Texts

Variations inter-annuelles



Dynamique des discours



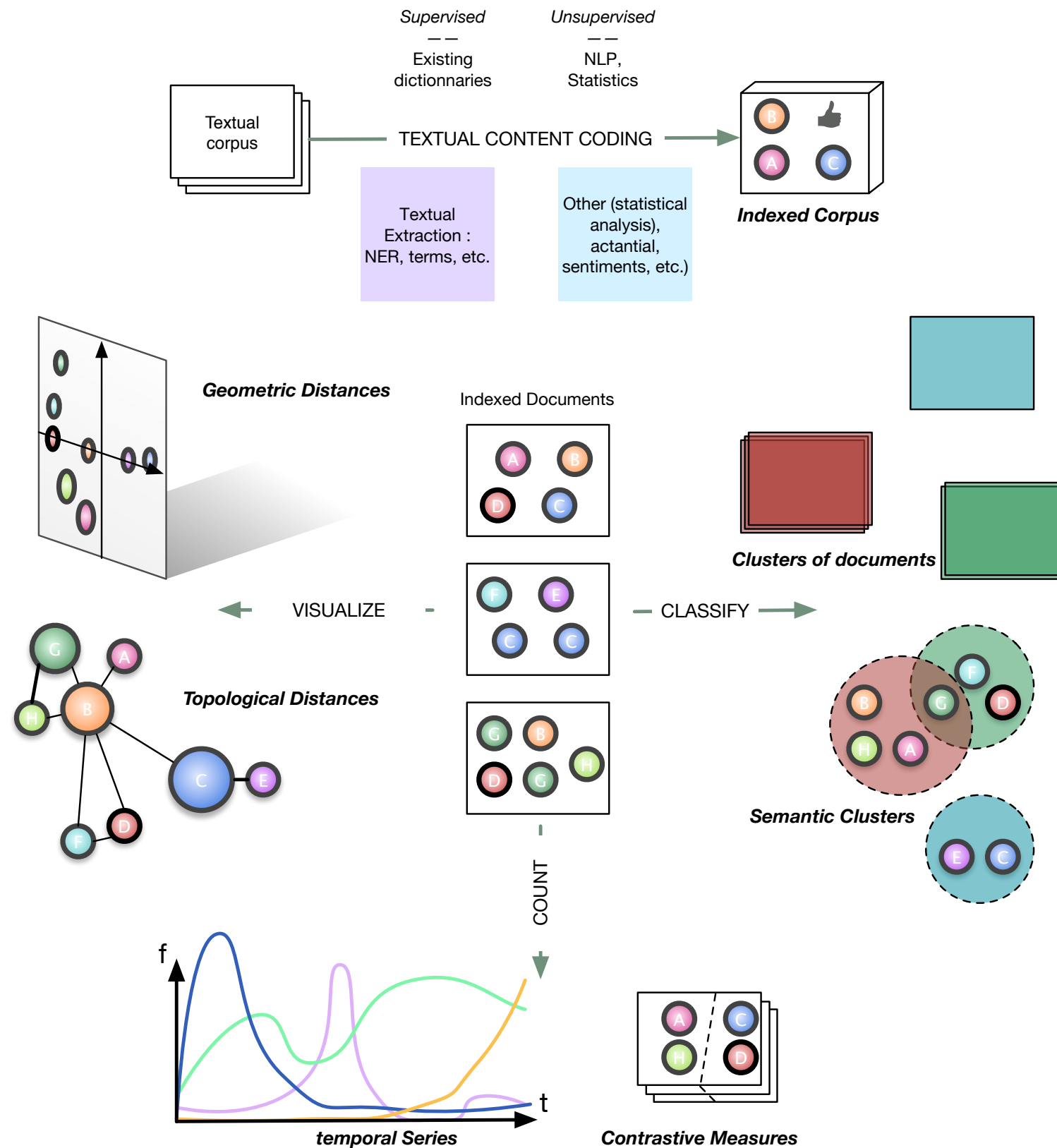
Structure macro des grands thèmes



Corpus

ECHELLE MACRO

Textual Corpus Processing

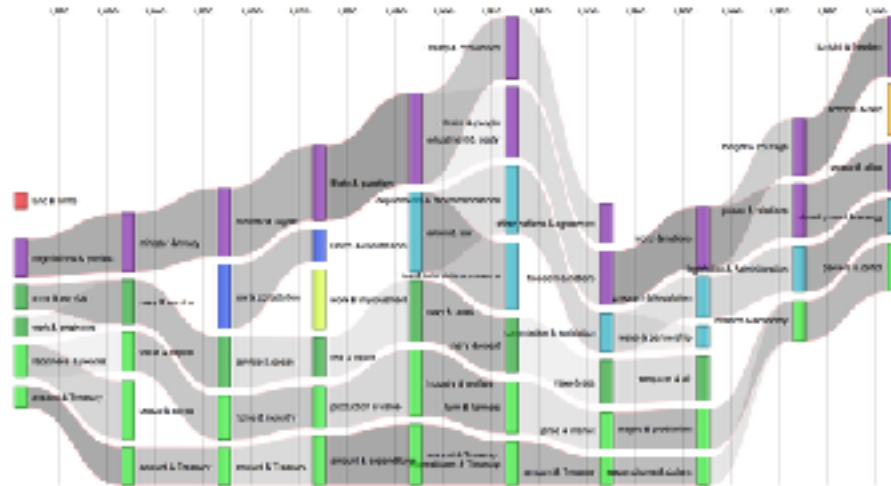


Landscape multi-level dynamics

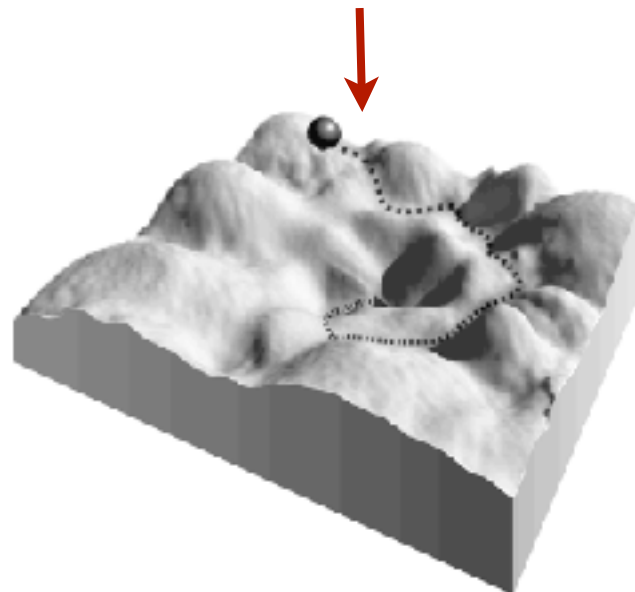
- Network as traces of local heterogeneous dynamics
- Reconstruction of network dynamics
- Landscape modeling of socio-semantic system



Network

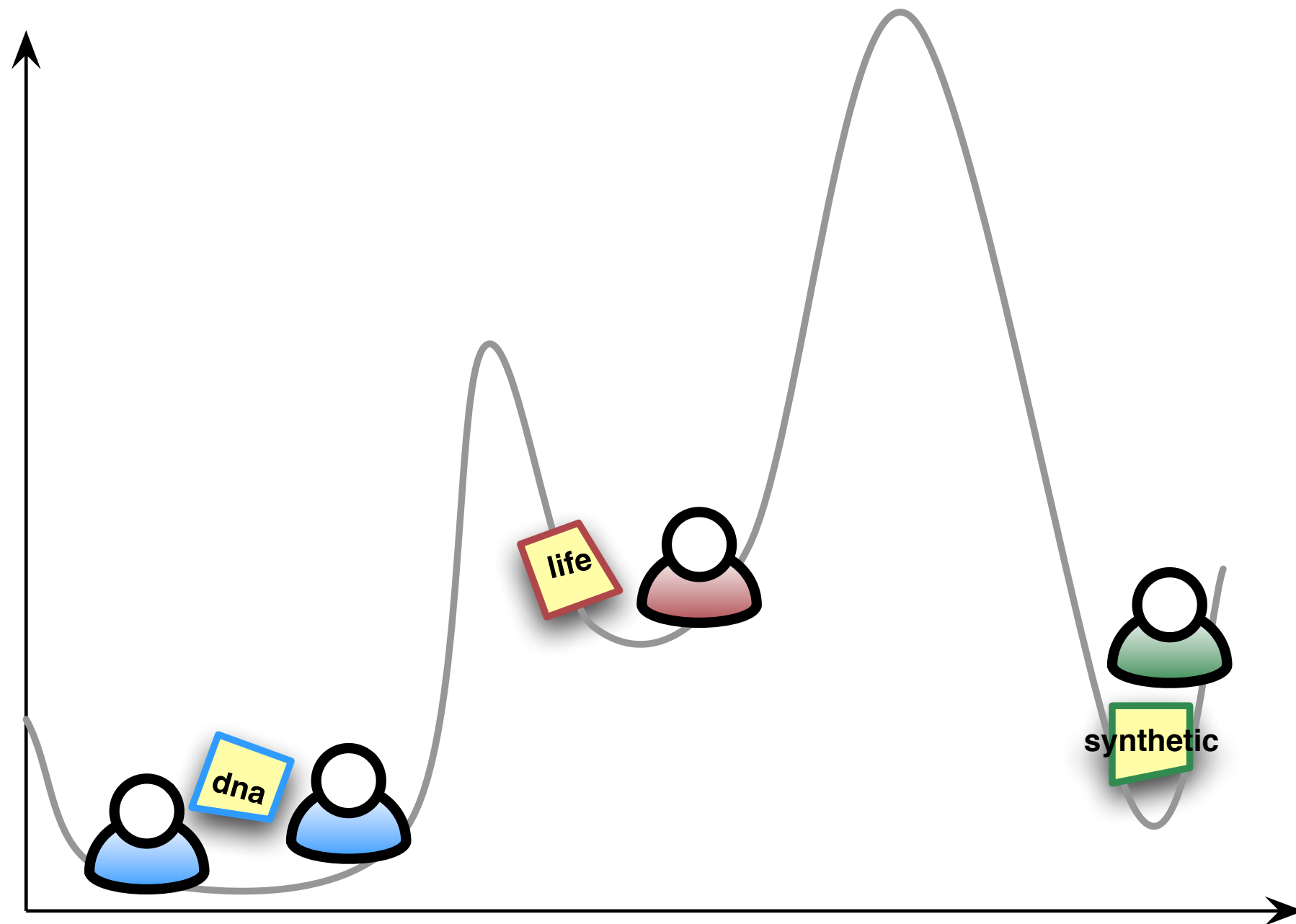


Streams



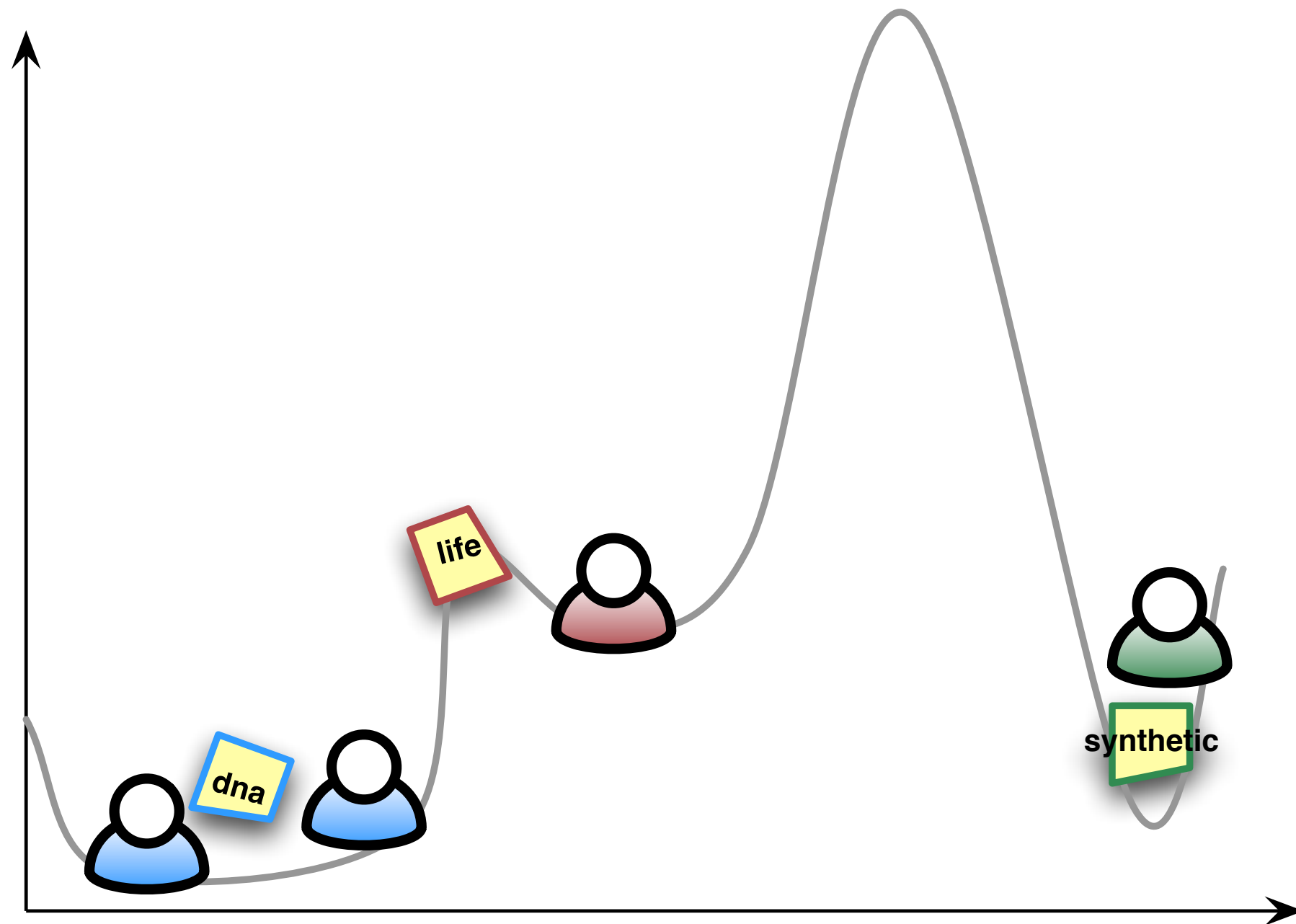
Landscape?

Landscape multi-level dynamics



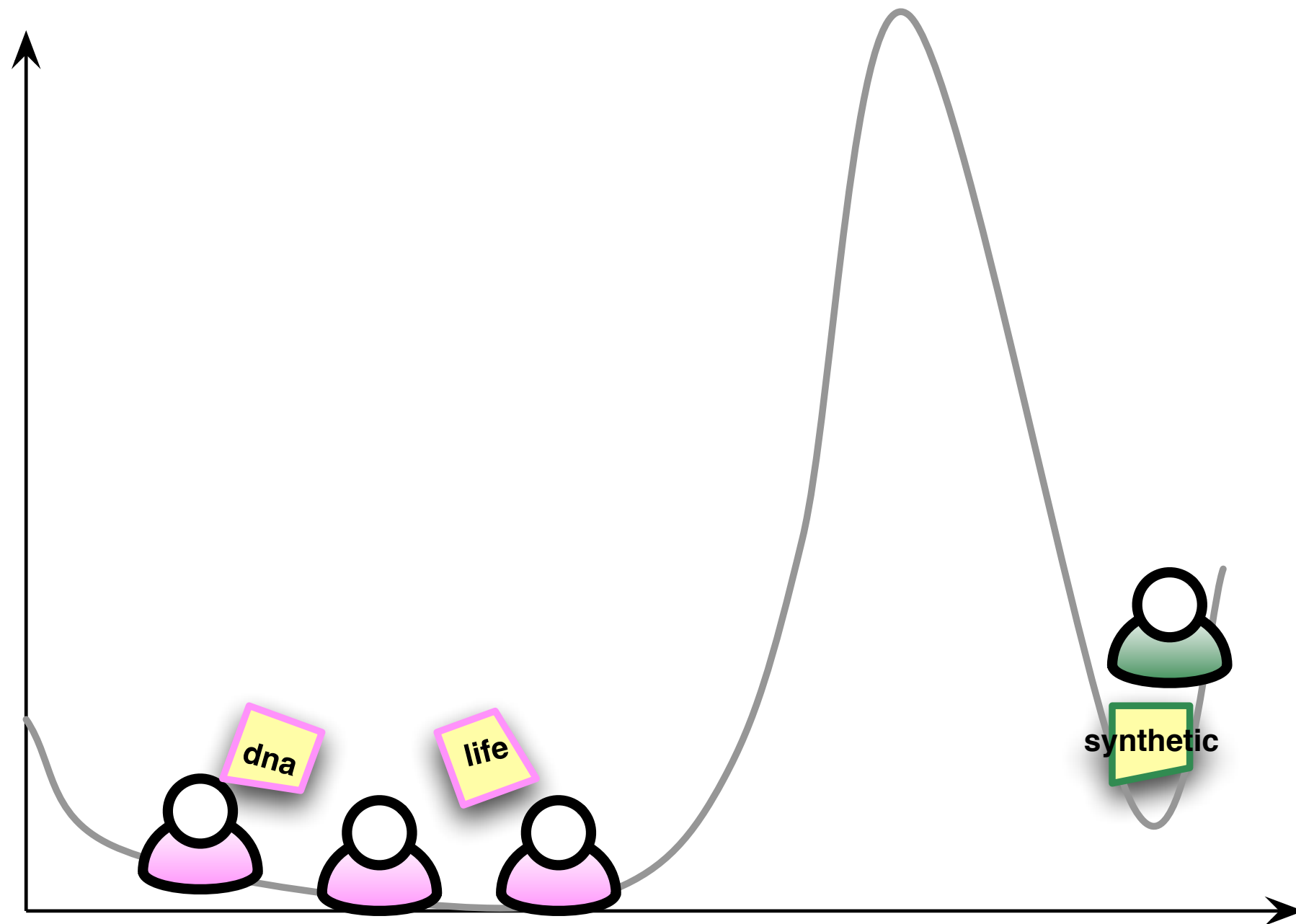
Landscape and its actors ($t=1$)

Landscape multi-level dynamics



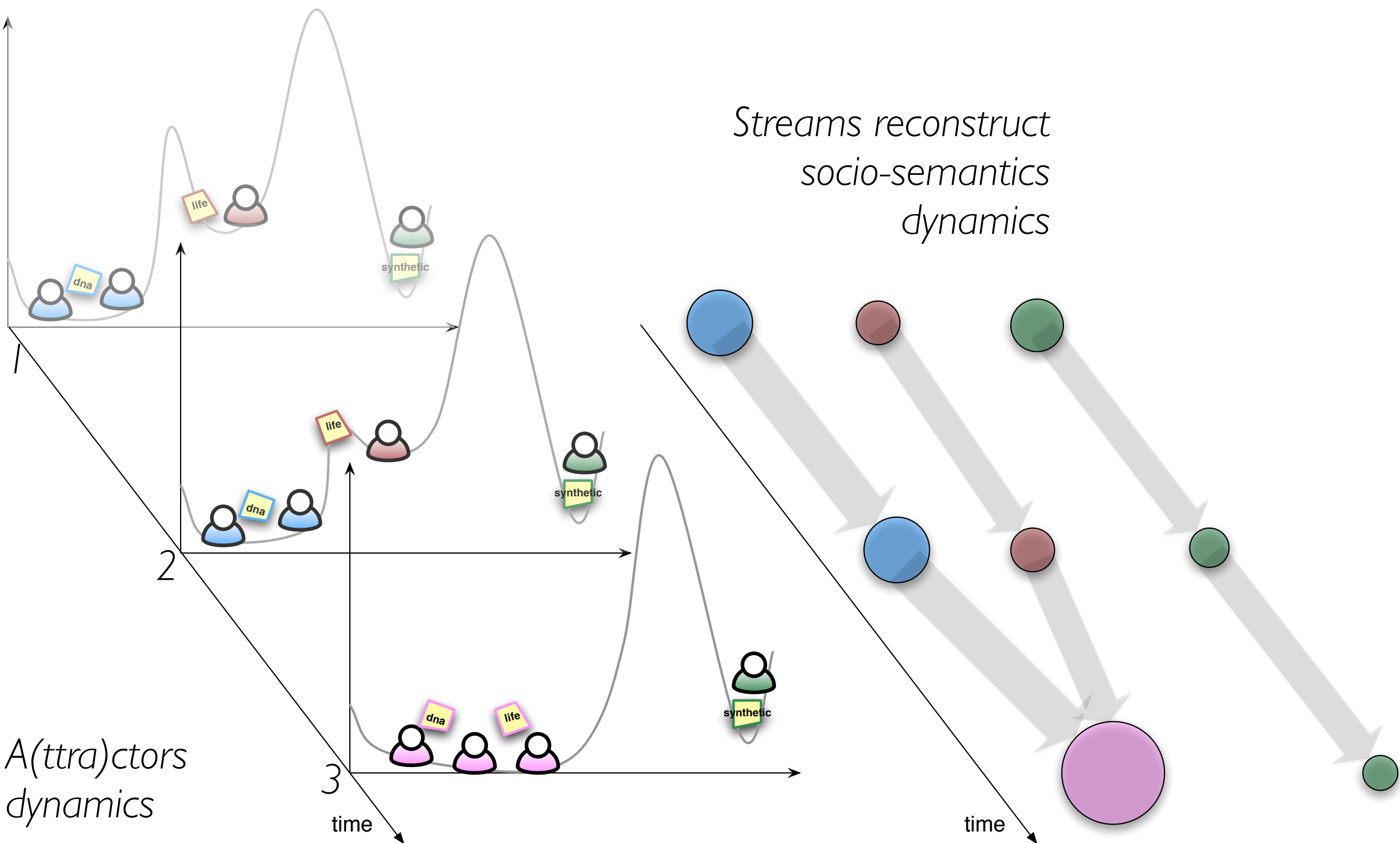
Landscape and its actors ($t=2$)

Landscape multi-level dynamics

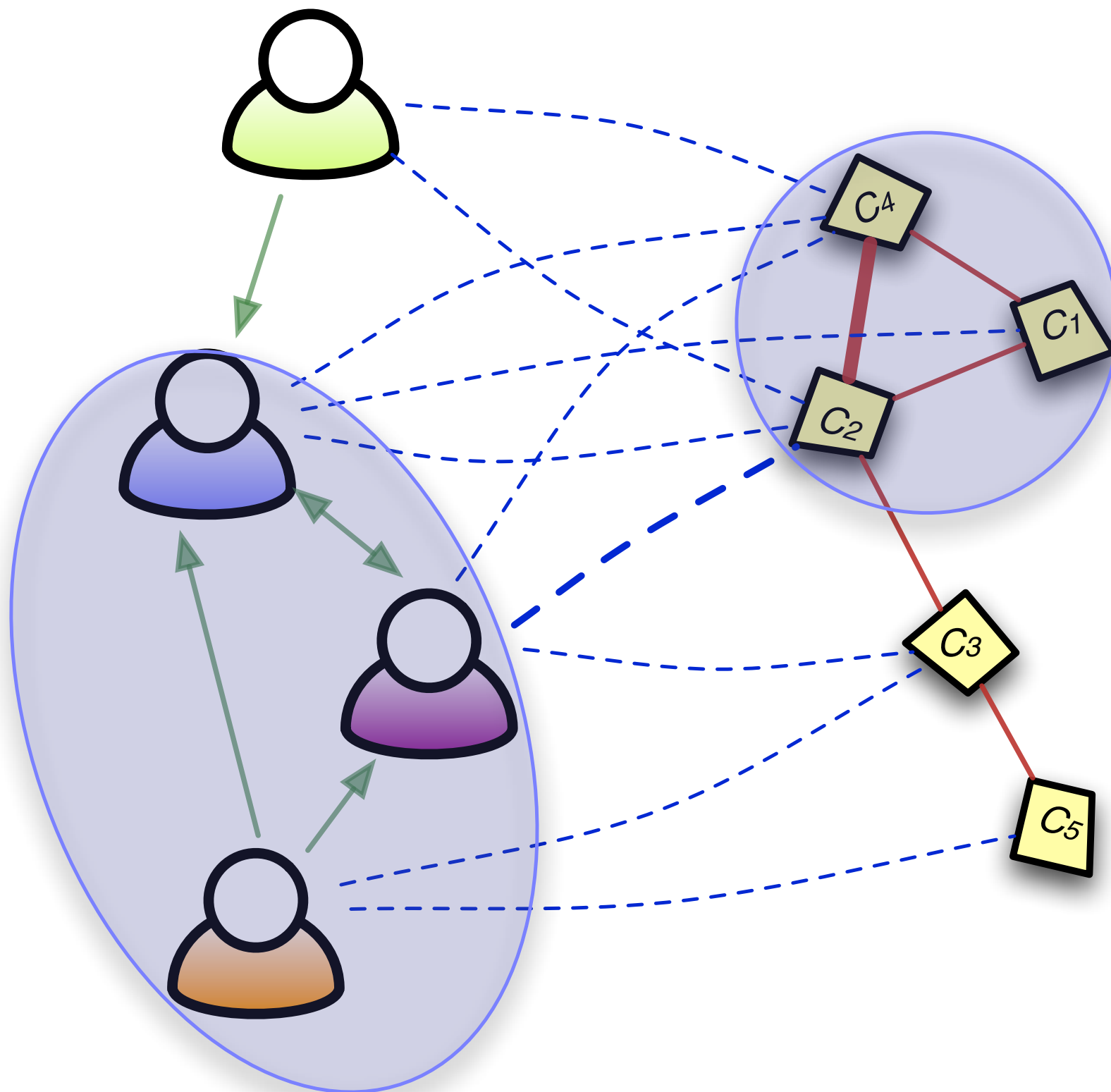


Landscape and its actors ($t=3$)

Landscape multi-level dynamics



Heterogeneous cooccurrences



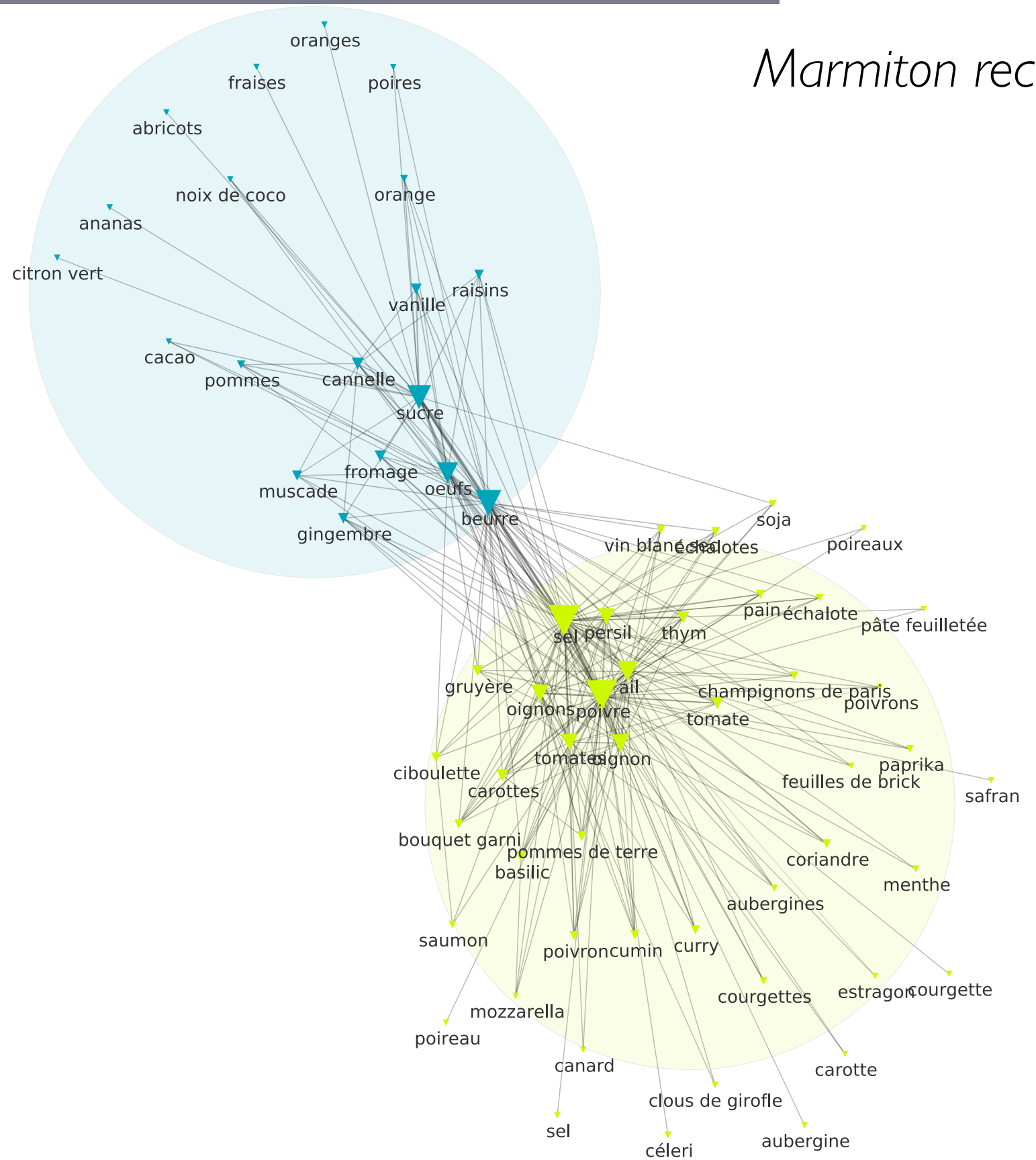
- Generalized co-occurrences analysis framework mixing people, terms, countries, etc...

Understanding proximity metrics

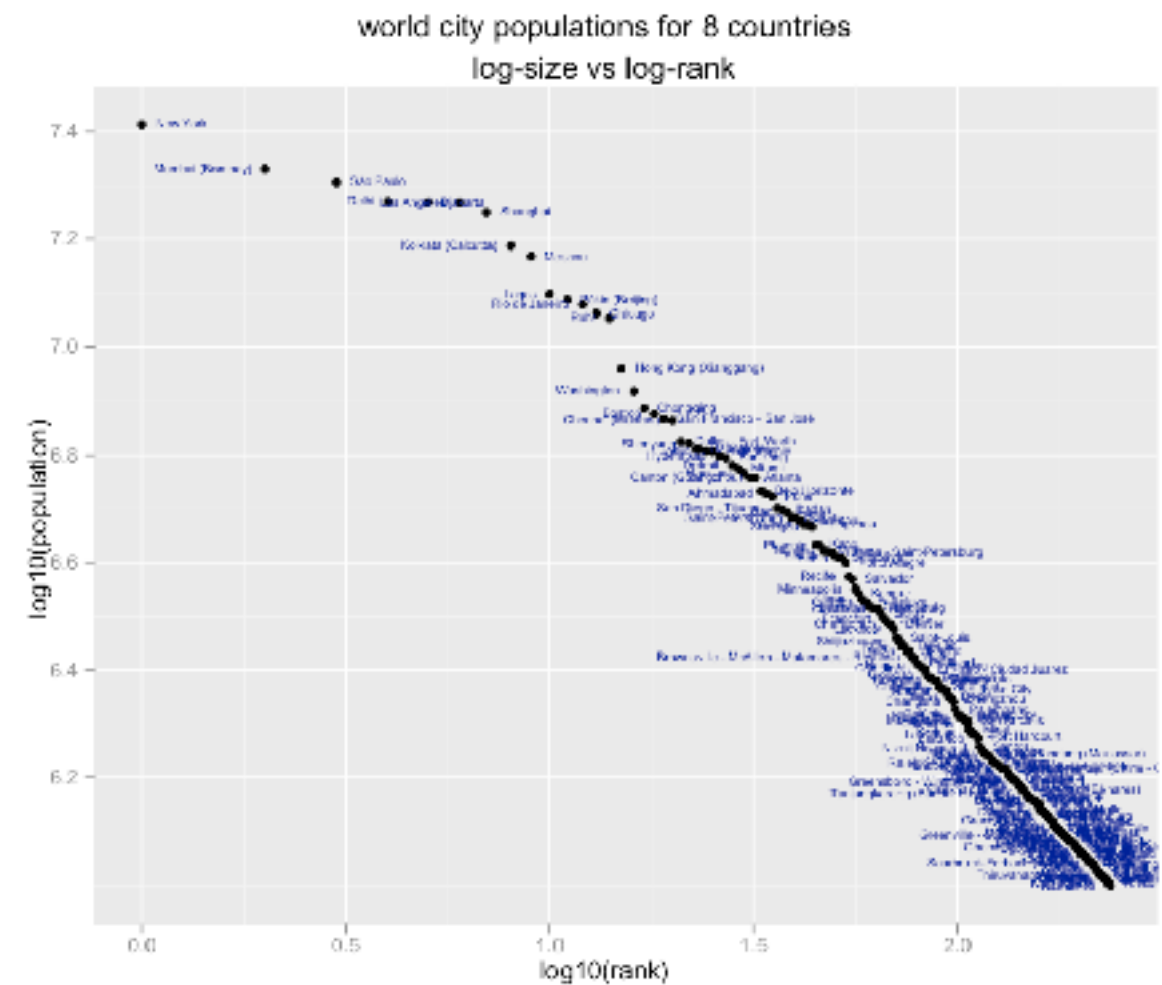
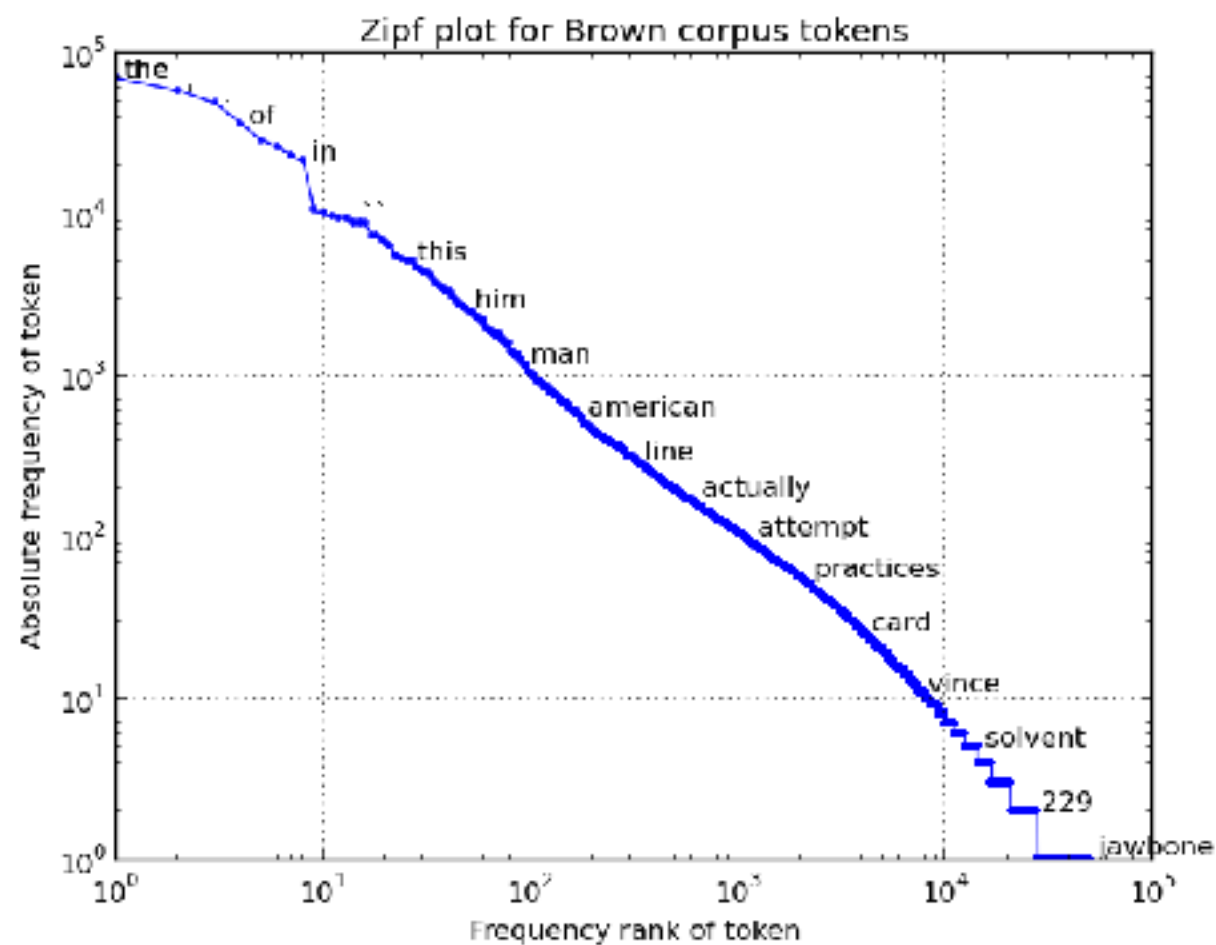
Metrics

Marmiton recipes

- Raw cooccurrence measure



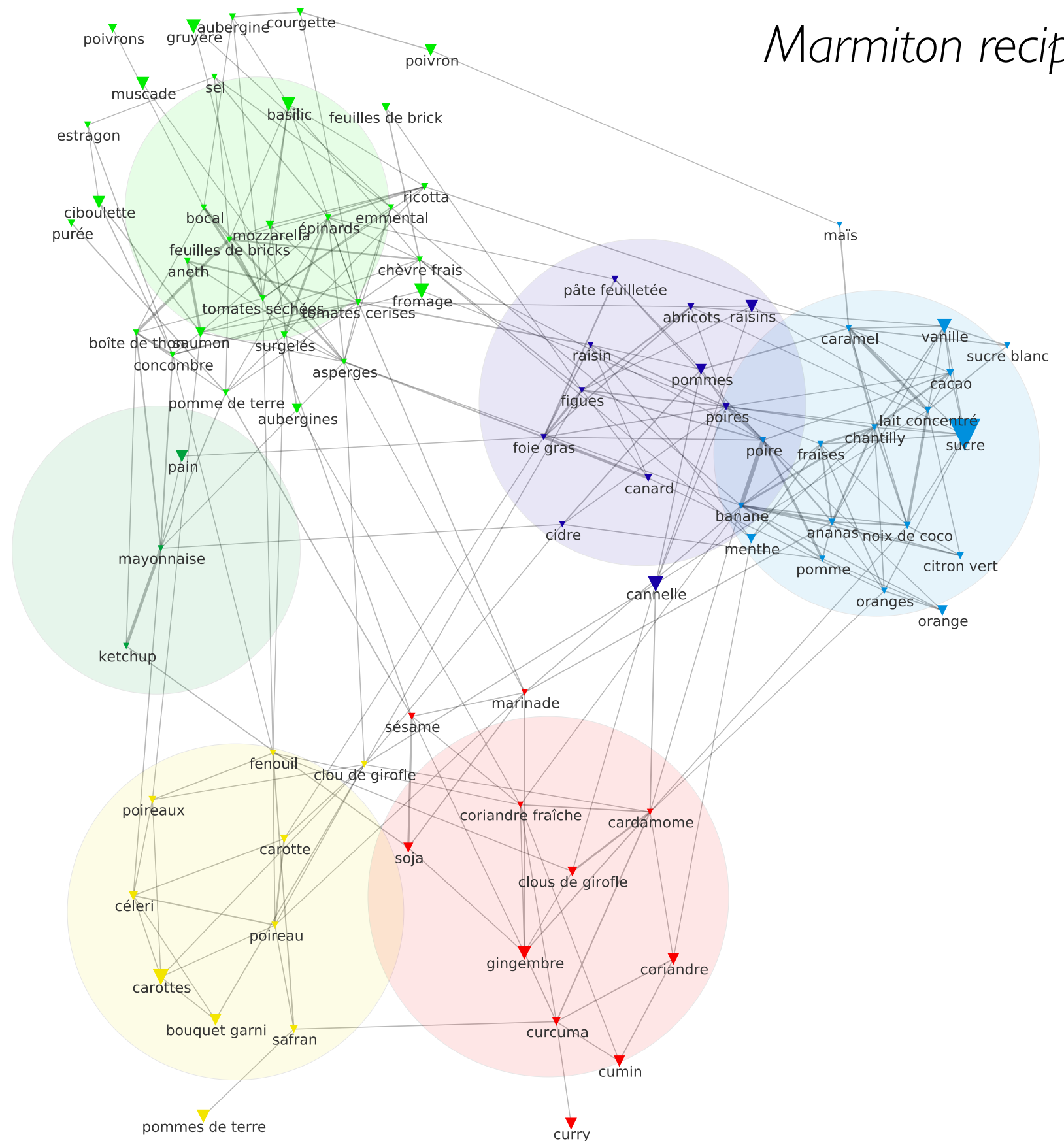
Heterogeneity curse



Metrics

Marmiton recipes

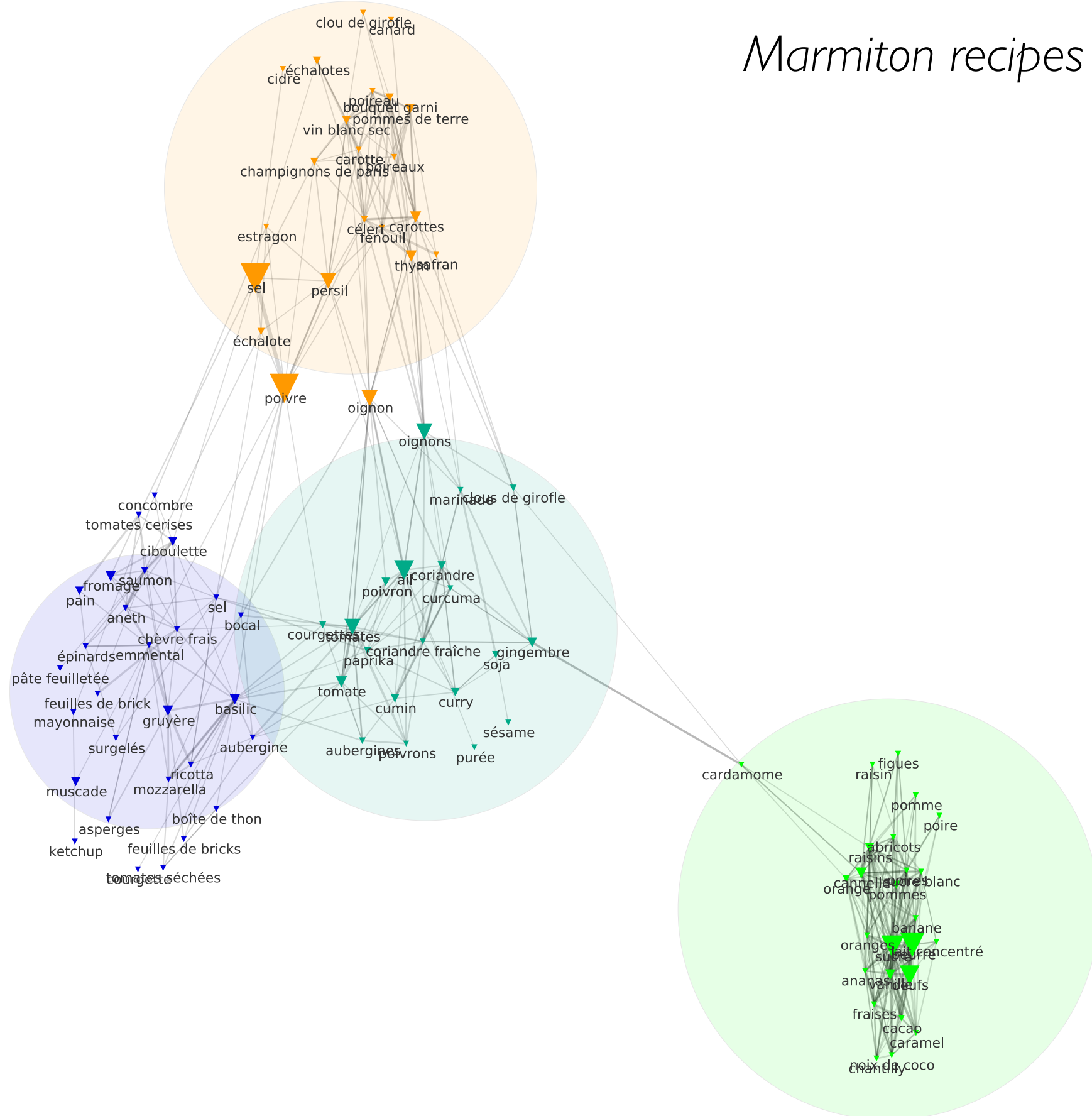
● Cramer



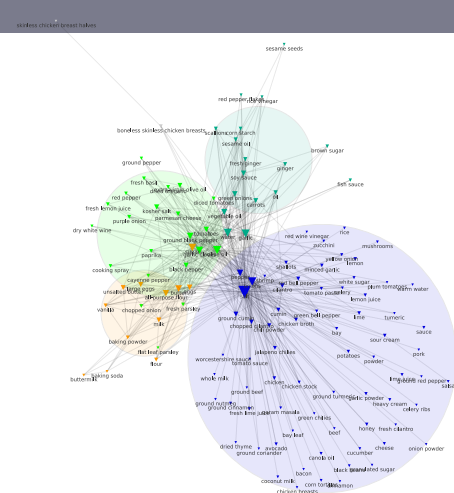
Metrics

- Distributional Measure

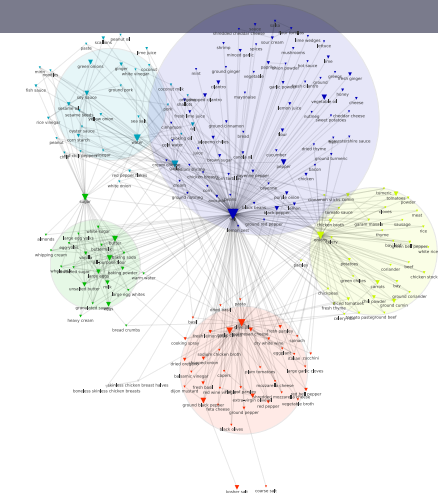
Marmiton recipes



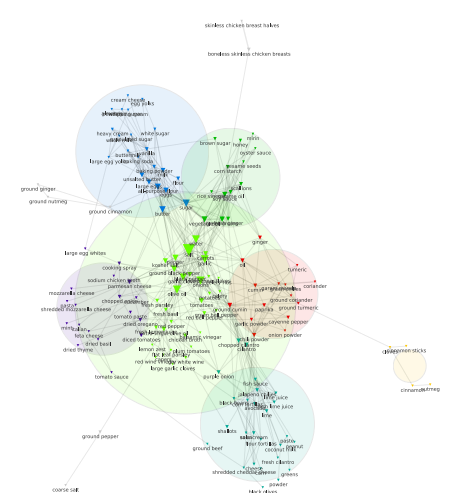
Metrics



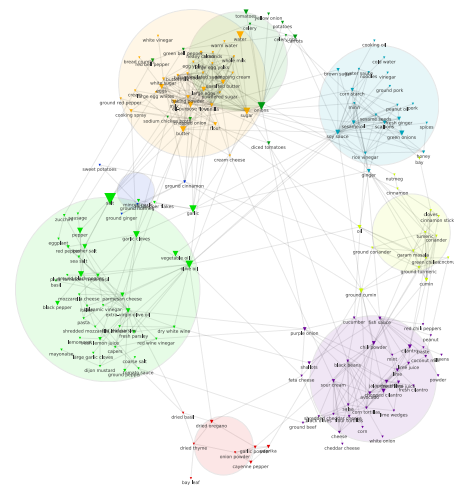
cooccurrences brutes ($Q = 0.2$)



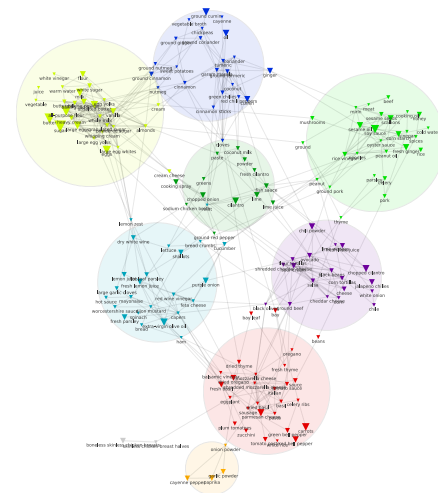
indice d'inclusion ($Q = 0.38$)



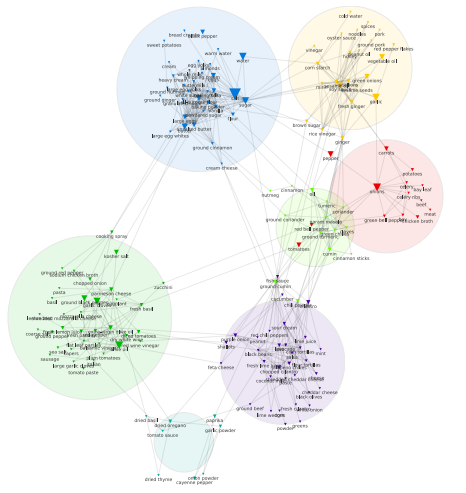
coefficient de Jaccard ($Q = 0.56$)



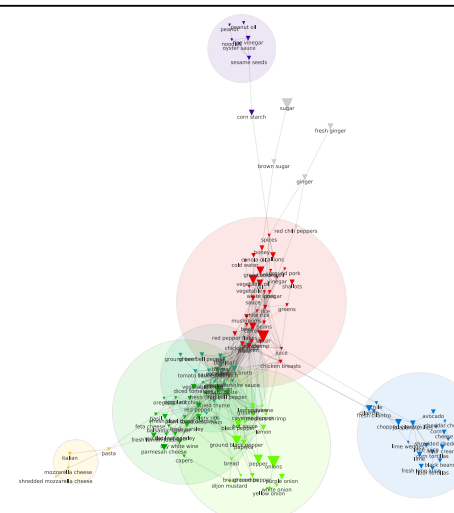
χ^2 ($Q = 0.70$)



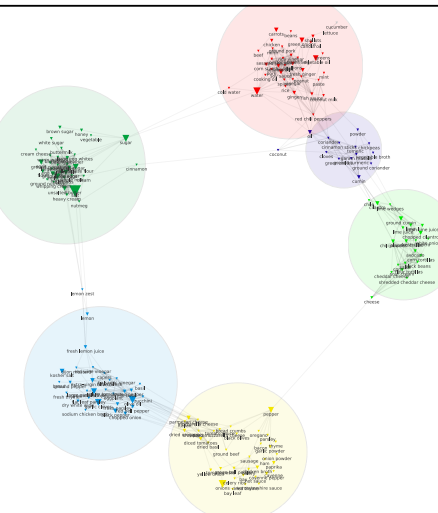
IM ($Q = 0.69$)



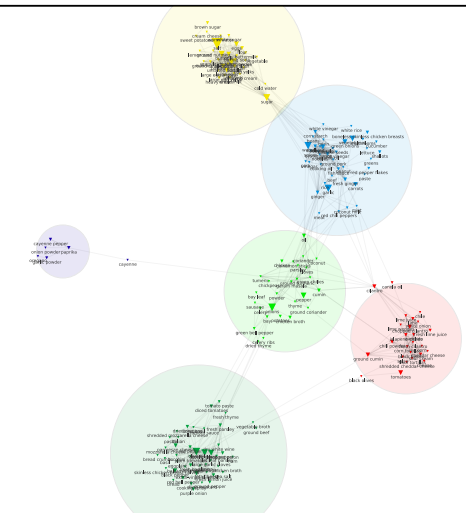
(LLR) ($Q = 0.67$)



cosine ($Q = 0.52$)



s_d IM ($Q = 0.75$)



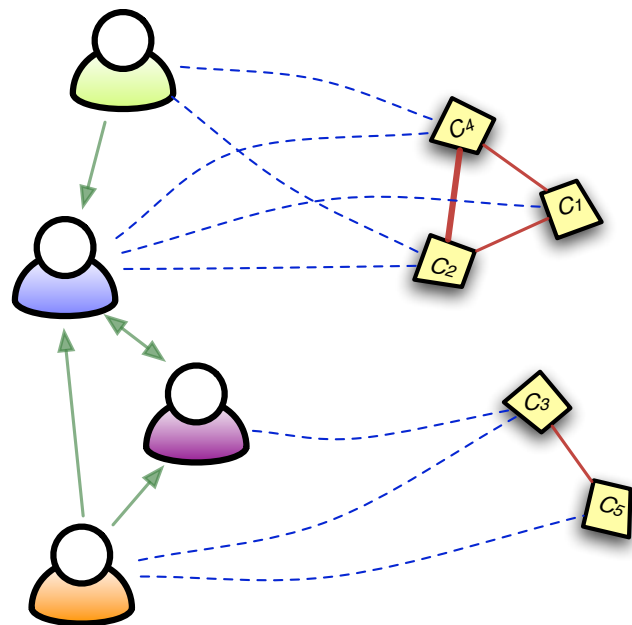
s_d LLR ($Q = 0.73$)

To summarize

Cooccurrences matrices

Cooccurrences matrices construction

- 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. $C = O^t O$

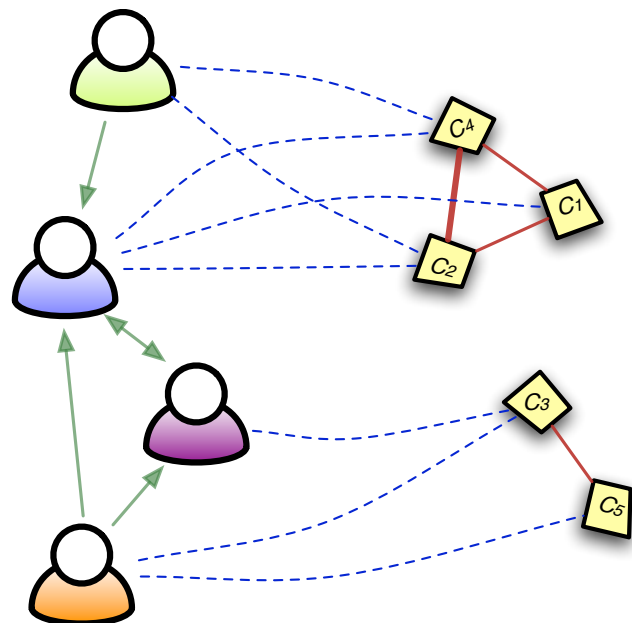


documents items	A	B	C	D
C1		I		
C2	I	I		
C3			I	I
C4	I	I		
C5				I

Cooccurrences matrices

Cooccurrences matrices construction

- 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

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurrences of i : $s_i = \sum_{j, j \neq i} C_{ij}$
- global number of co-occurrences : $N = \sum_i s_i$
- expected number of cooccurrences : $e_{ij} = \frac{s_i s_j}{N}$

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

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurrences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurrences : $N = \sum_i s_i$
- expected number of cooccurrences : $e_{ij} = \frac{s_i s_j}{N}$

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

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurrences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurrences : $N = \sum_i s_i$
- expected number of cooccurrences : $e_{ij} = \frac{s_i s_j}{N}$

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

Proximity Measures

Cooccurrences variables

- Co-Occurrence matrix C :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurrences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurrences : $N = \sum_i s_i$
- expected number of cooccurrences : $e_{ij} = \frac{s_i s_j}{N}$

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

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences : $S_R(i, j) = c_{ij}$*
- *Association strength : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$*
- *Mutual Information : $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$*
- *Cosine : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$*
- *Inclusion index : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$*
- *Jaccard index : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$*
- χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$
- *Cramer index : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$*

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences : $S_R(i, j) = c_{ij}$*
- *Association strength : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$*
- *Mutual Information : $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$*
- *Cosine : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$*
- *Inclusion index : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$*
- *Jaccard index : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$*
- χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$
- *Cramer index : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$*

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences : $S_R(i, j) = c_{ij}$*
- *Association strength : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$*
- *Mutual Information : $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$*
- *Cosine : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$*
- *Inclusion index : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$*
- *Jaccard index : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$*
- *χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$*
- *Cramer index : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$*

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences : $S_R(i, j) = c_{ij}$*
- *Association strength : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$*
- *Mutual Information : $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$*
- *Cosine : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$*
- *Inclusion index : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$*
- *Jaccard index : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$*
- *χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$*
- *Cramer index : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$*

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C* :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences* : $S_R(i, j) = c_{ij}$
- *Association strength* : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$
- *Mutual Information* : $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$
- *Cosine* : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$
- *Inclusion index* : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$
- *Jaccard index* : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$
- χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$
- *Cramer index* : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C* :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences* : $S_R(i, j) = c_{ij}$
- *Association strength* : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$
- *Mutual Information* : $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$
- *Cosine* : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$
- *Inclusion index* : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$
- *Jaccard index* : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$
- χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$
- *Cramer index* : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences :* $S_R(i, j) = c_{ij}$
- *Association strength :* $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$
- *Mutual Information :* $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$
- *Cosine :* $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$
- *Inclusion index :* $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$
- *Jaccard index :* $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$
- χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$
- *Cramer index :* $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C* :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurences* : $S_R(i, j) = c_{ij}$
- *Association strength* : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$
- *Mutual Information* : $S_{MI}(i, j) = \log\left(\frac{c_{ij}}{e_{ij}}\right)$
- *Cosine* : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$
- *Inclusion index* : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$
- *Jaccard index* : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$
- χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$
- *Cramer index* : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurrences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurrences : $N = \sum_i s_i$
- expected number of cooccurrences : $e_{ij} = \frac{s_i s_j}{N}$

Indirect Measures of Similarity :

- Cosine (distributional) :

$$S_{Cd}(i, j) = \frac{\sum_{k \neq i, j} c_{ik} c_{jk}}{\sqrt{\sum_{k \neq i, j} c_{ik}^2 \sum_{k \neq i, j} c_{jk}^2}}$$

- Mutual Information (distributional) :

$$S_{MId} = \frac{\sum_{k \neq i, j; MI_{ik} > 0} \min(MI_{ik}, MI_{jk})}{\sum_{k \neq i, j; MI_{ik} > 0} MI_{ik}}$$

Proximity Measures

Cooccurrences variables

- *Co-Occurrence matrix C :*
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurrences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurrences : $N = \sum_i s_i$
- expected number of cooccurrences : $e_{ij} = \frac{s_i s_j}{N}$

Indirect Measures of Similarity :

- Cosine (distributional) :

$$S_{Cd}(i, j) = \frac{\sum_{k \neq i, j} c_{ik} c_{jk}}{\sqrt{\sum_{k \neq i, j} c_{ik}^2 \sum_{k \neq i, j} c_{jk}^2}}$$

- Mutual Information (distributional) :

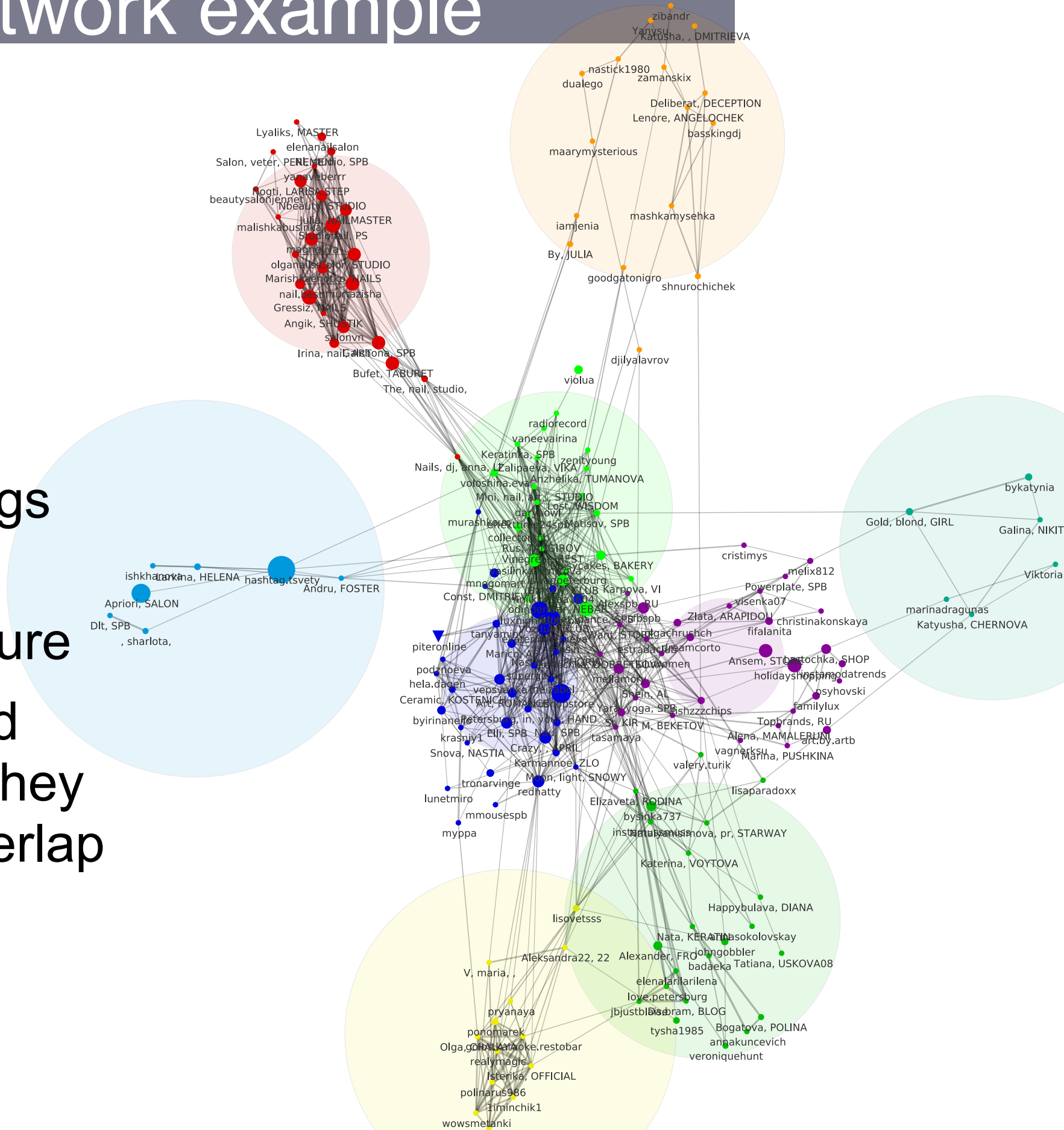
$$S_{MId} = \frac{\sum_{k \neq i, j; MI_{ik} > 0} \min(MI_{ik}, MI_{jk})}{\sum_{k \neq i, j; MI_{ik} > 0} MI_{ik}}$$

Metrics Summary

proximity measures	type of network	normalisation	special properties
raw	interaction network (<i>e.g.</i> social network)	no	-
χ^2	all	yes	normalization tend to create links toward higher degree nodes
MI	all	yes	Inspired from information theory
Cramer	all	yes	-
cosine	homogeneous network (eg. semantic)	yes	Classical measure (originating from scientometrics)
distributional	homogeneous network (eg. semantic)	yes	very robust measure (coming from computational linguistics)
cosine_het	affiliation network (eg. users sharing the same hashtags)	yes	two fields are required but the final network is homogeneous
dot_product_het	affiliation network (eg. users sharing the same hashtags)	no	two fields are required but the final network is homogeneous

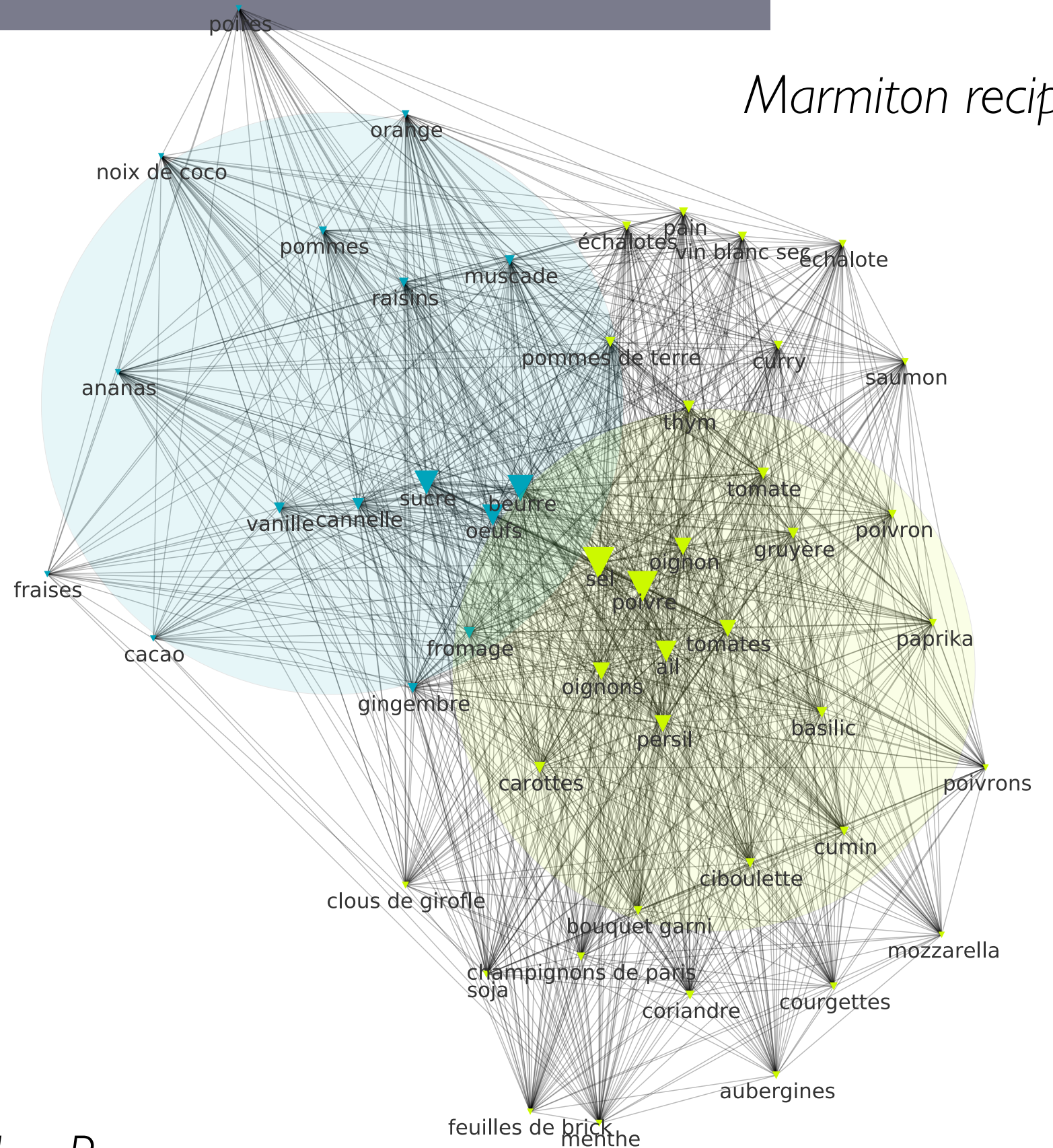
Affiliation network example

- parameters:
 - Field1: Users
 - Field 2: Hashtags
 - cosine_het proximity measure
- Two Users are linked when the hashtags they use have a large overlap



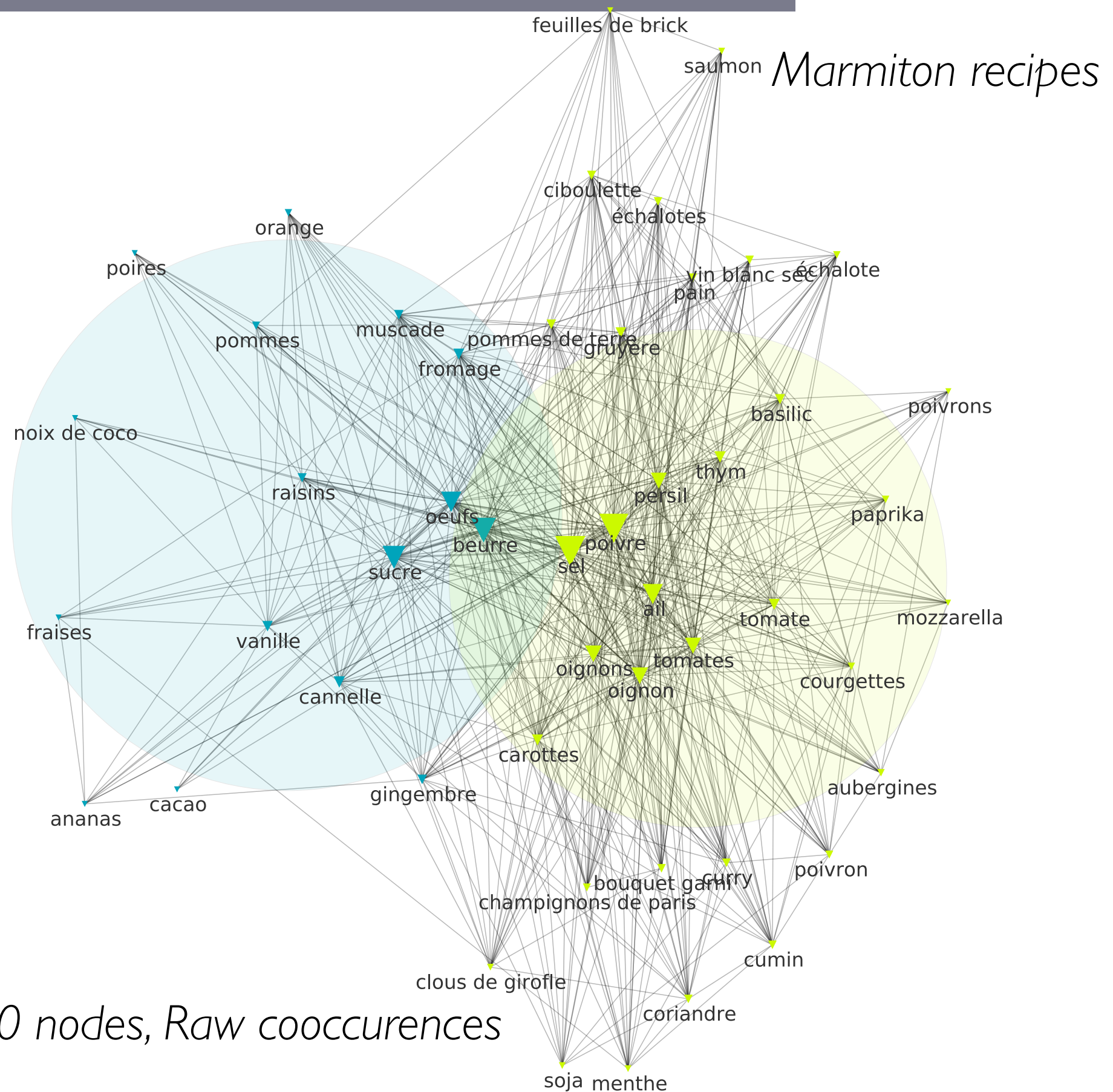
Filtering

- No filtering



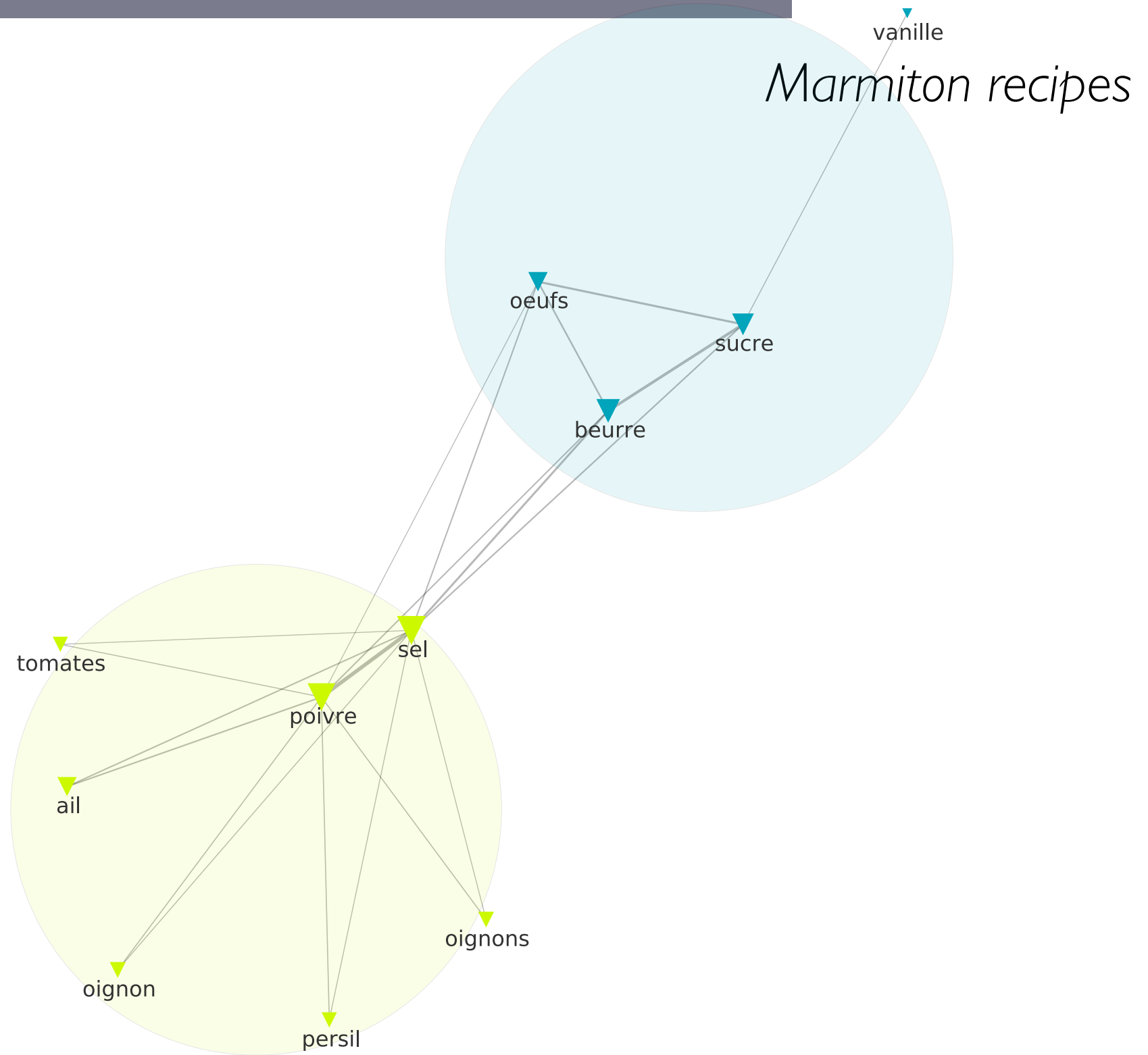
Filtering

- filtering on edge strength values: 10 cooc at least



Filtering

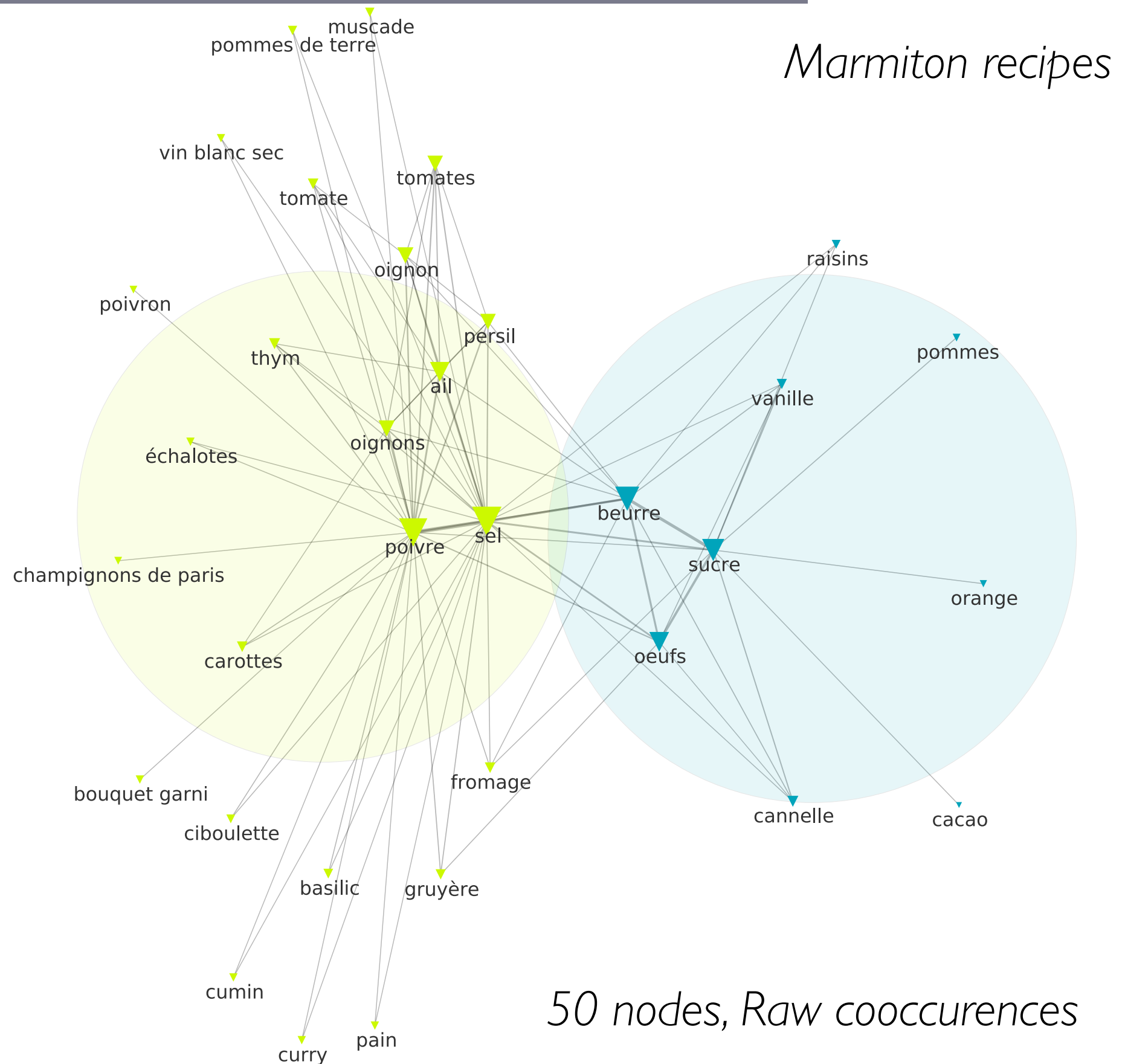
- filtering on the number of edges: 20 Top edges



50 nodes, Raw cooccurrences

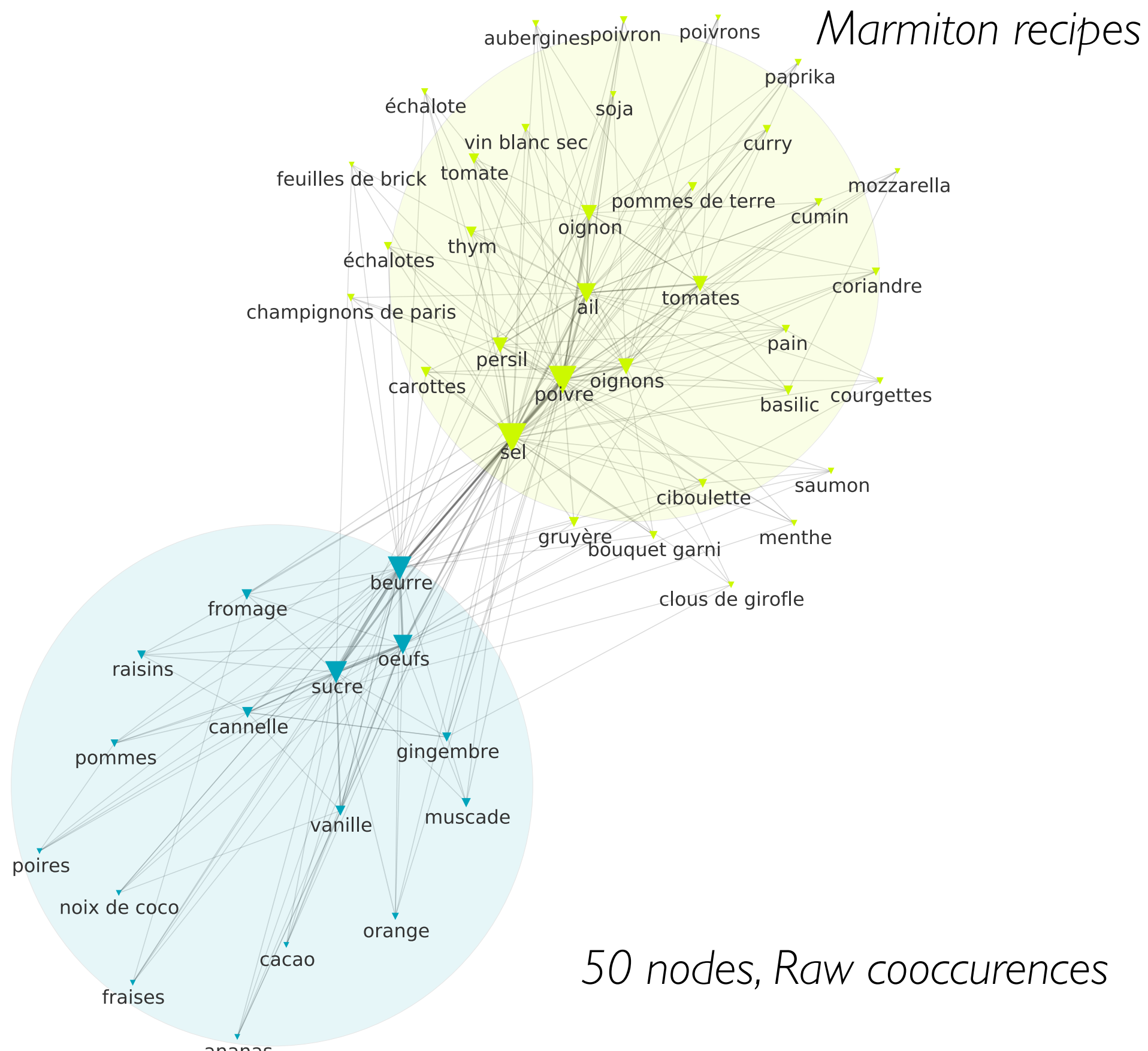
Filtering

- Auto-threshold (the edge strength threshold is automatically inferred from graph topology)

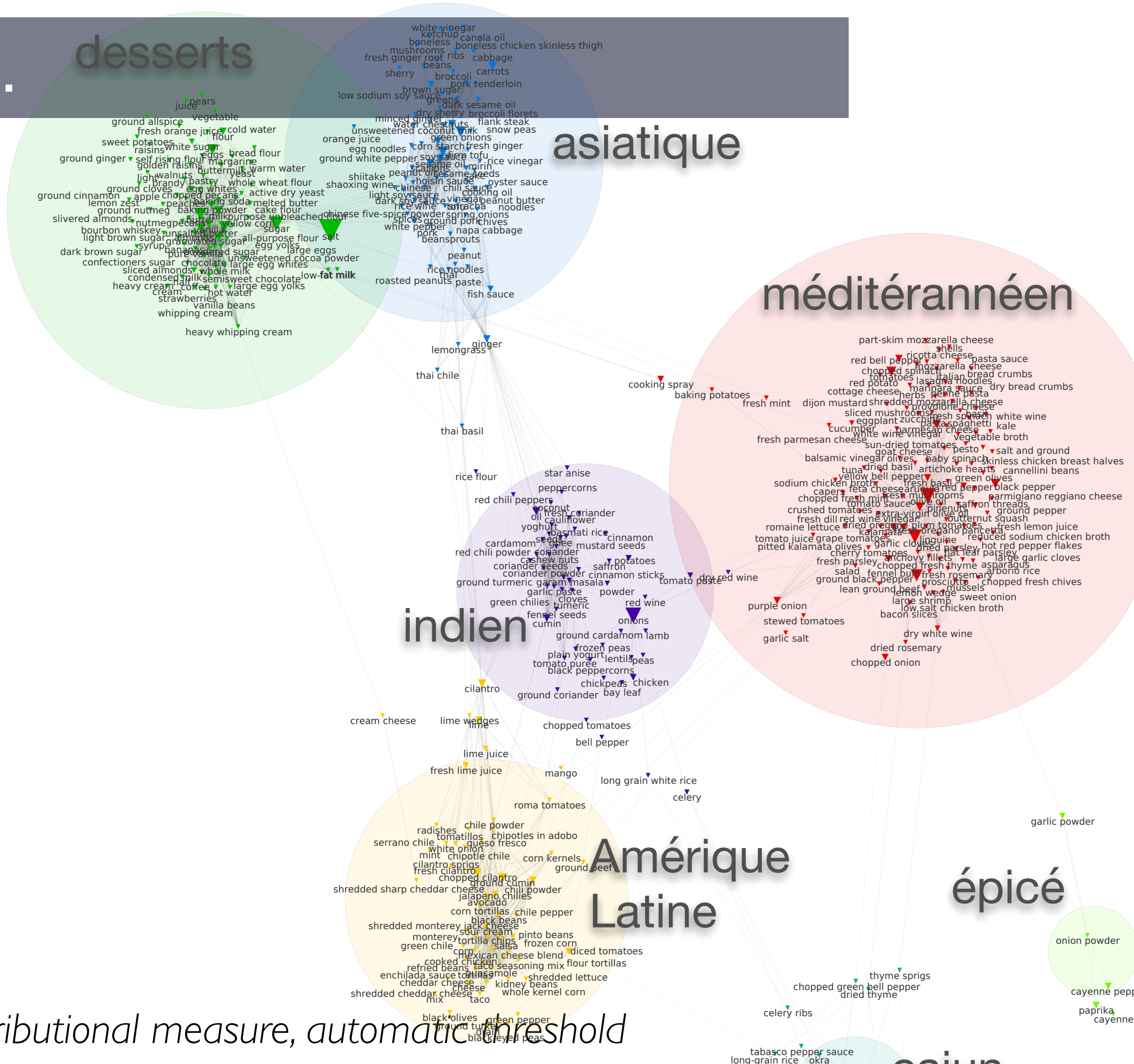


Filtering

- Local Filtering: 5 top neighbours



Recipes...

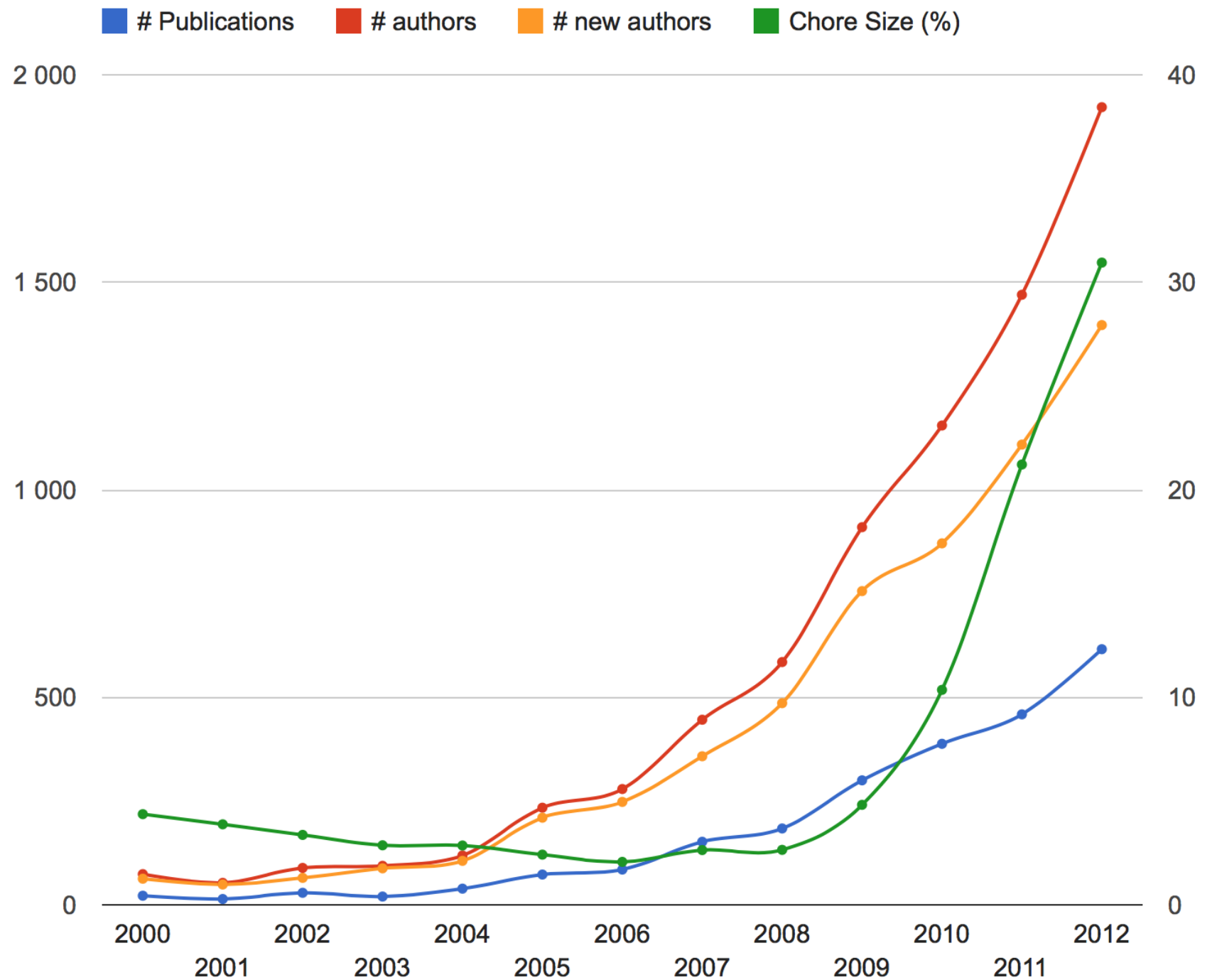
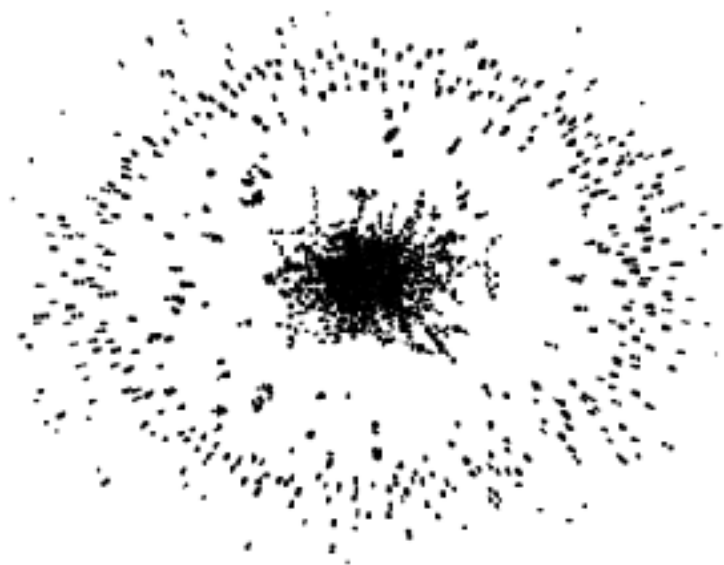


500 nodes, Distributional measure, automatic threshold

Synthetic Biology Emergence

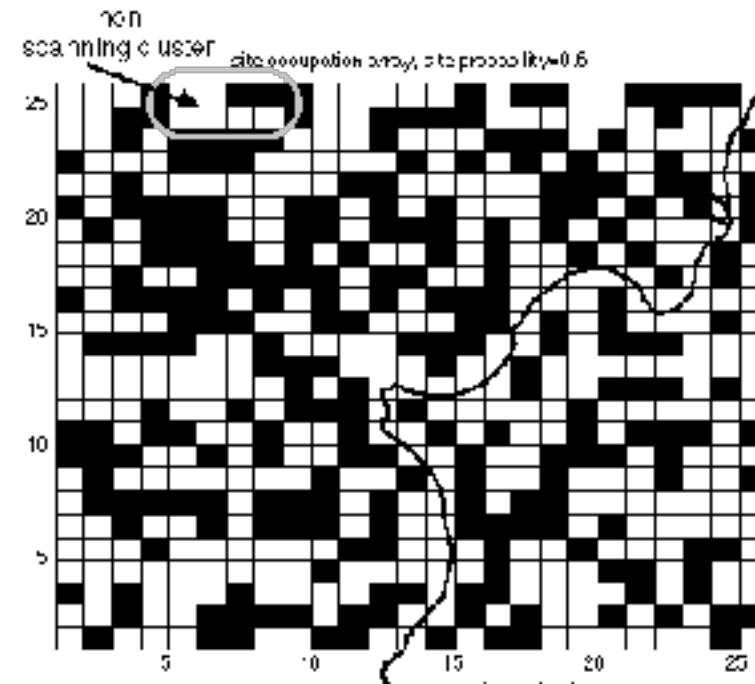
Collaboration Network Dynamics

- Unstabilized field at the crossroads between different disciplinary origins
- Exponential growth of publications with a very high rate of newcomers
- Although a central community is emerging

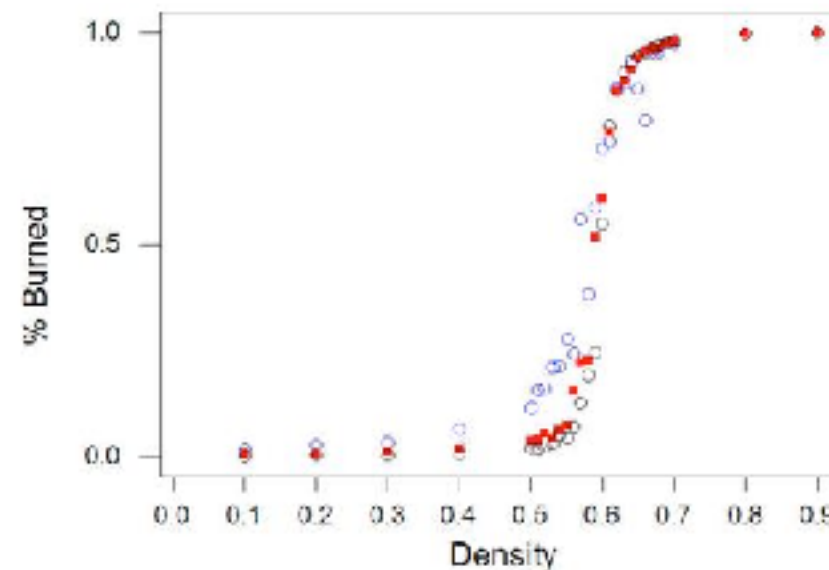


Collaboration Network Dynamics

- Unstabilized field at the crossroads between different disciplinary origins
- Exponential growth of publications with a very high rate of newcomers
- Although a central community is emerging



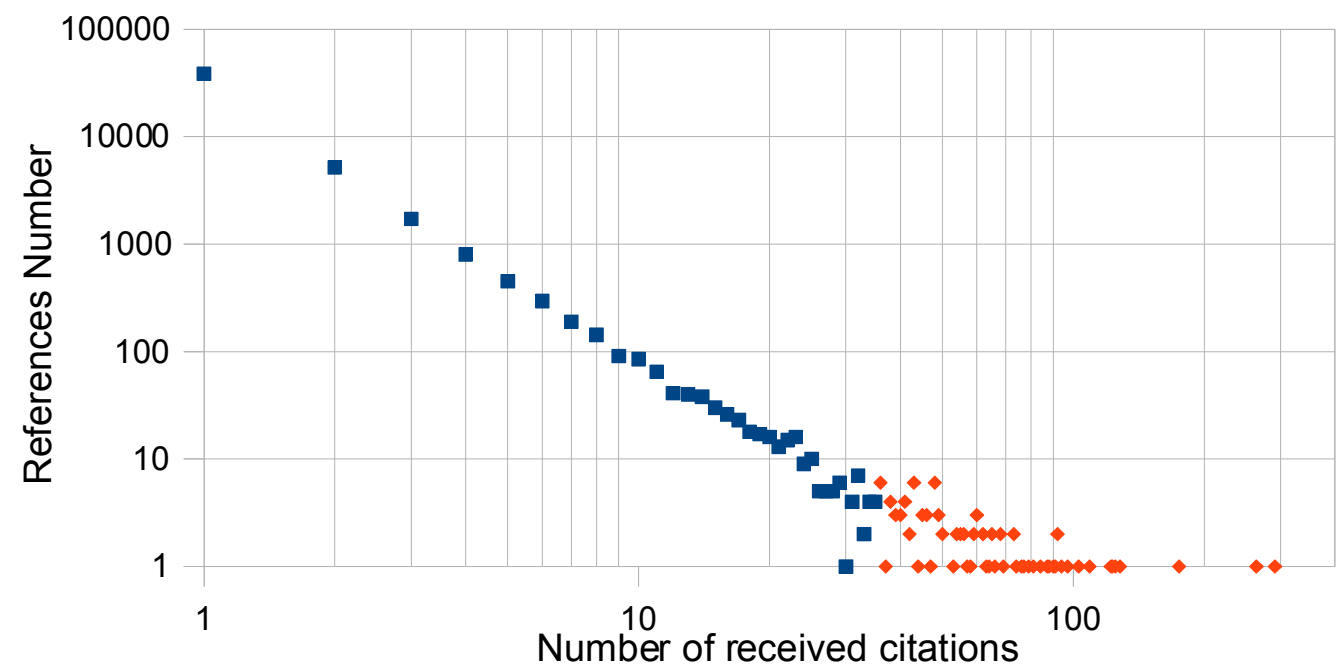
Percent of Forest Lost



«percolation threshold» above which the giant connected component collapse in disconnected subgraphs

Scientific references co-citation map

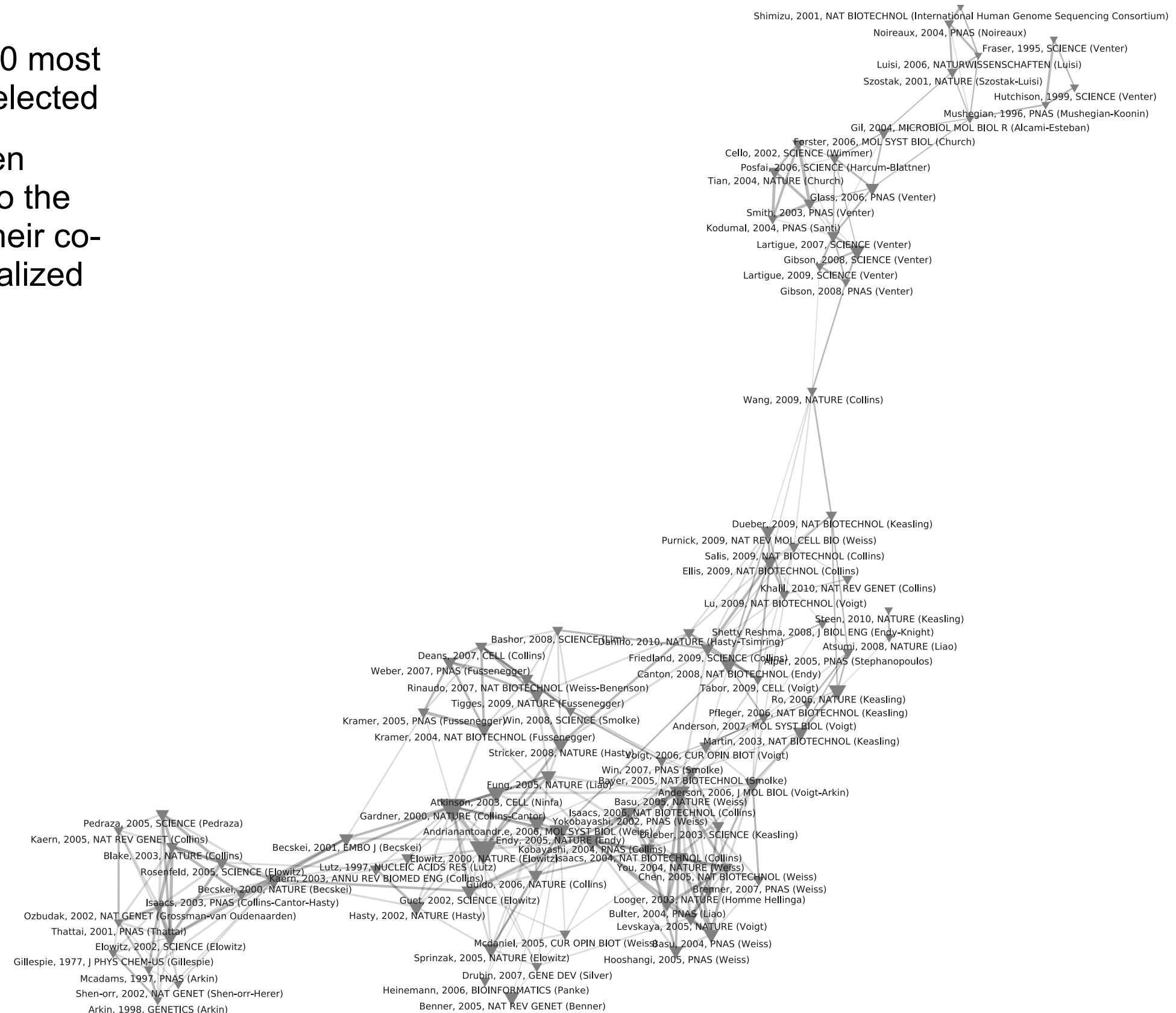
- **Selection:** only the 100 most cited references are selected



Received citations distributions

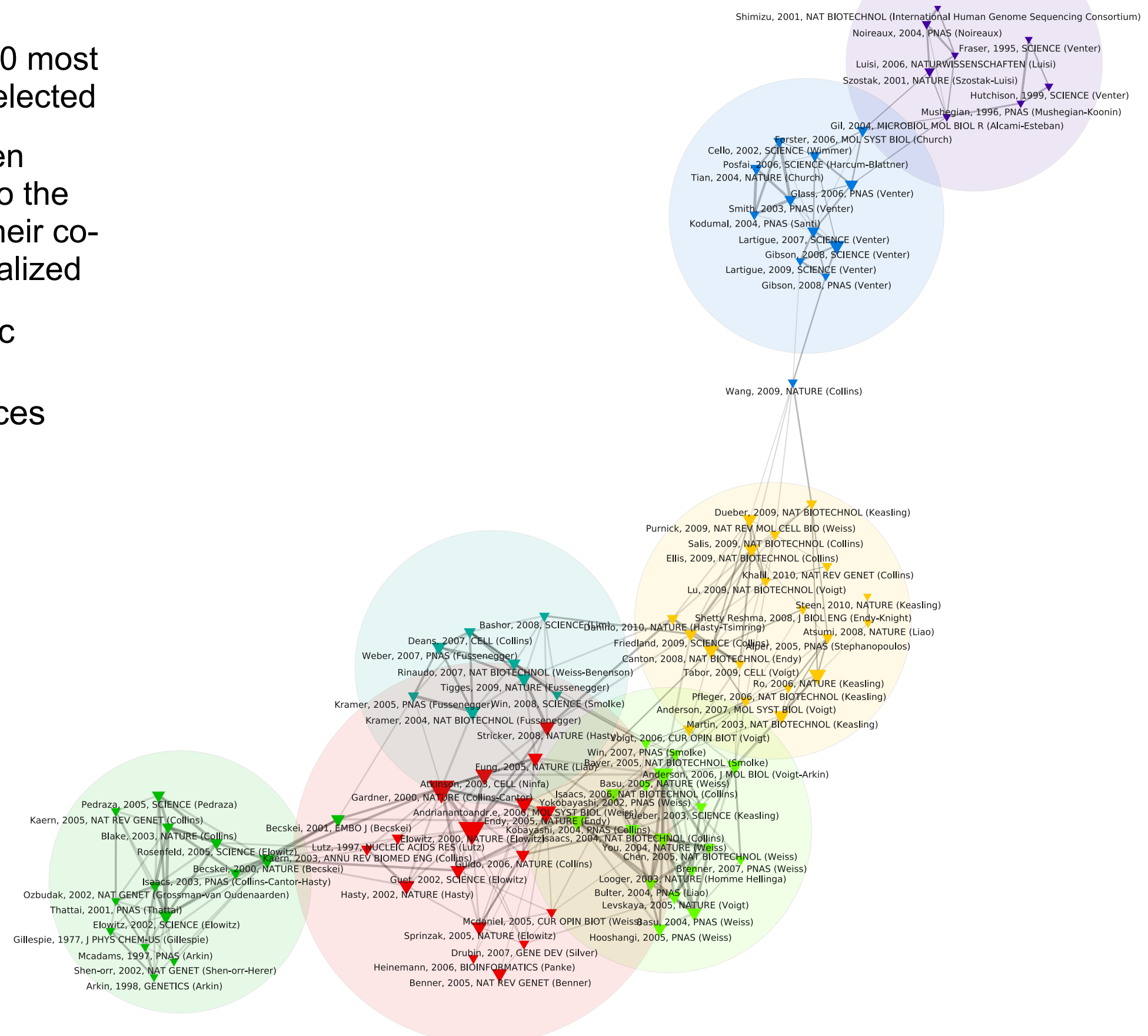
Scientific references co-citation map

- **Selection:** only the 100 most cited references are selected
- **Network:** They are then connected according to the relative frequency of their co-occurrences and spatialized



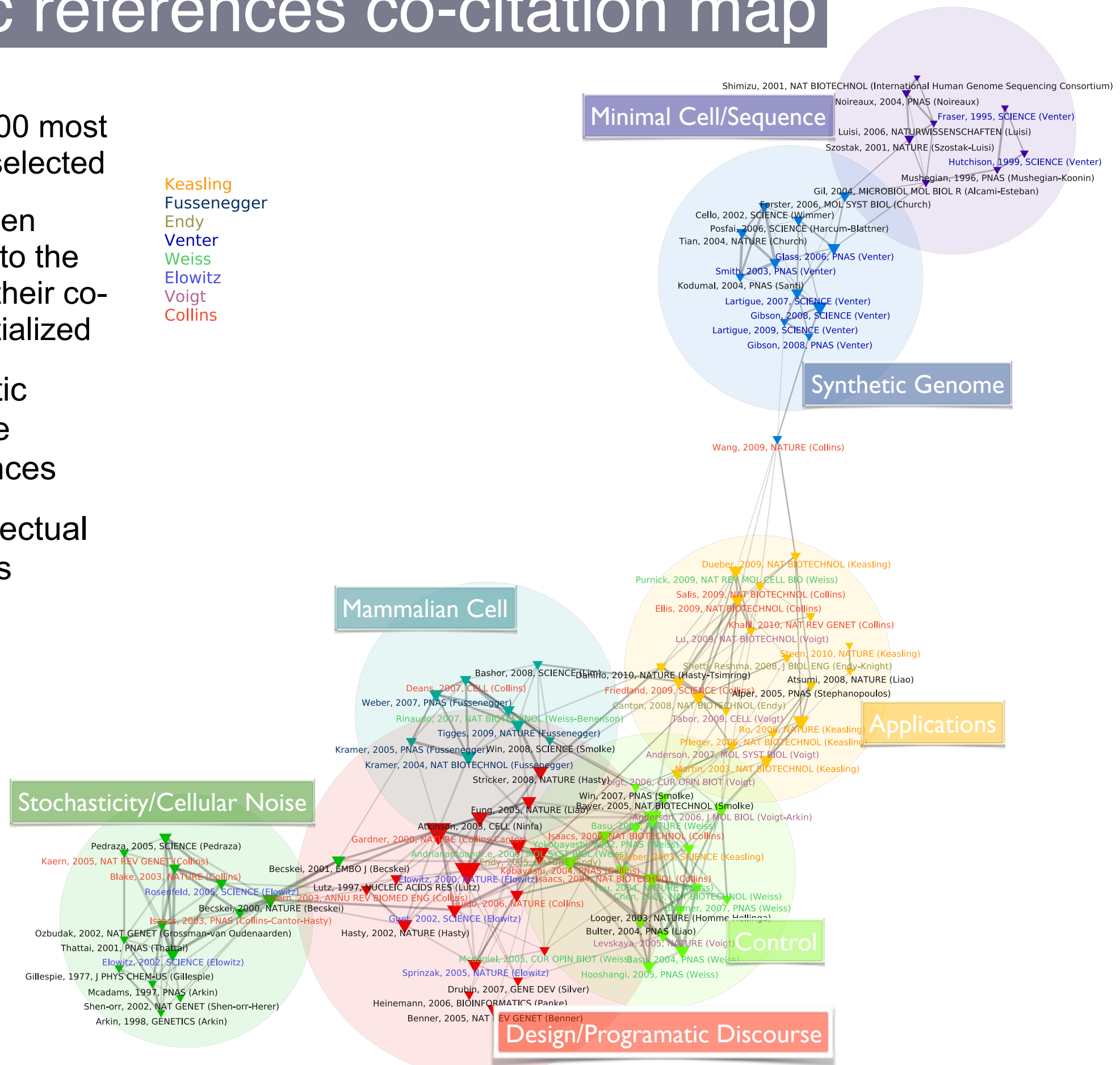
Scientific references co-citation map

- **Selection:** only the 100 most cited references are selected
- **Network:** They are then connected according to the relative frequency of their co-occurrences and spatialized
- **Clustering:** Automatic detection of cohesive subgroups of references



Scientific references co-citation map

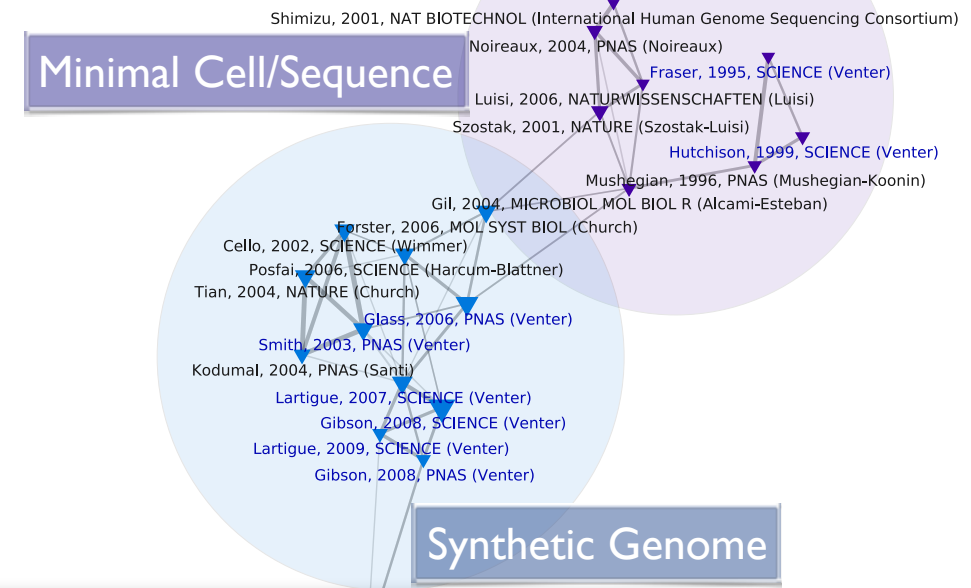
- **Selection:** only the 100 most cited references are selected
- **Network:** They are then connected according to the relative frequency of their co-occurrences and spatialized
- **Clustering:** Automatic detection of cohesive subgroups of references
- **Interpretation:** intellectual schools, major actors



Scientific references co-citation map

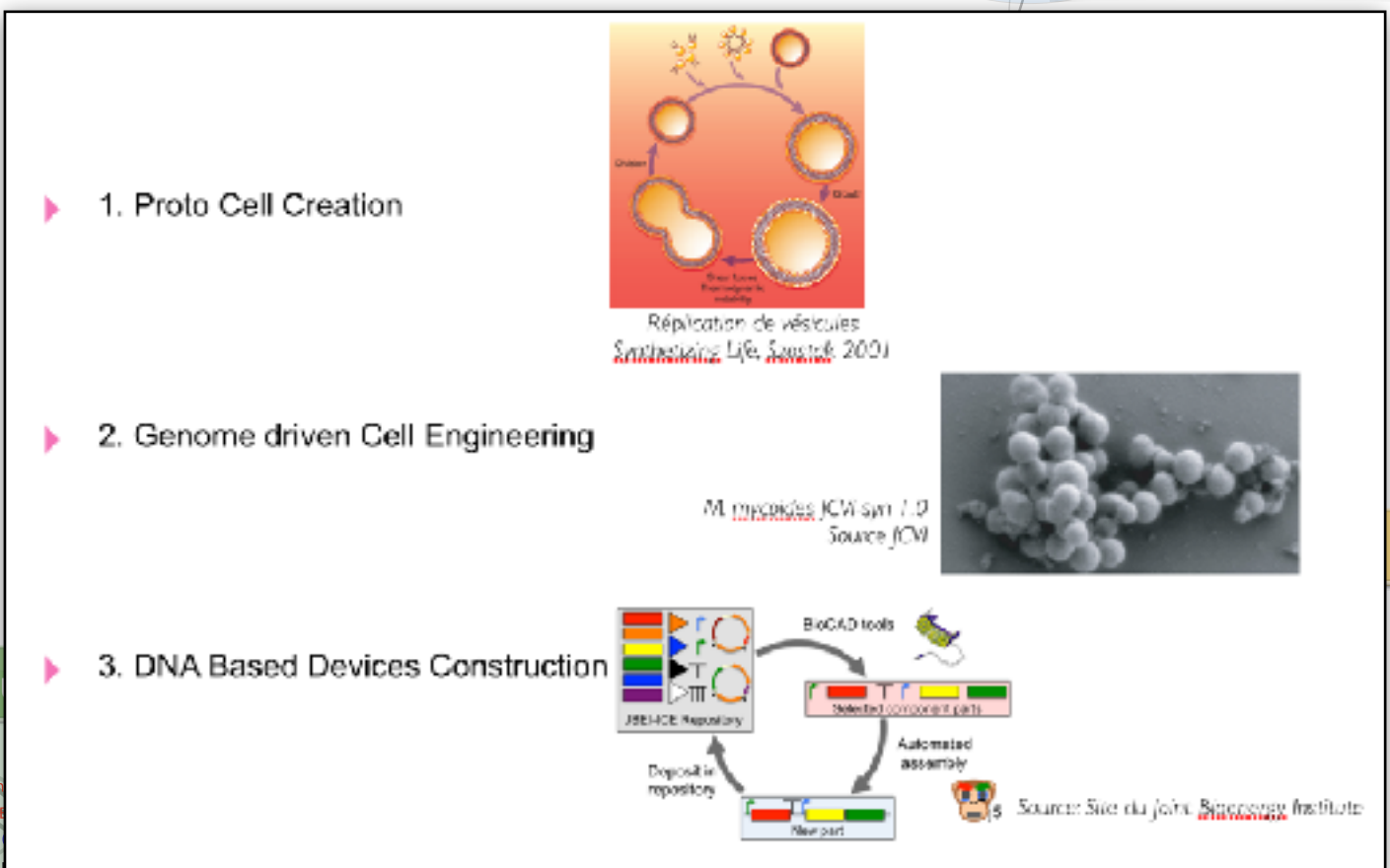
- **Selection:** only the 100 most cited references are selected
- **Network:** They are then connected according to the relative frequency of their co-occurrences and spatialized
- **Clustering:** Automatic detection of cohesive subgroups of references
- **Interpretation:** intellectual schools, major actors

Keasling
Fussenegger
Endy
Venter
Weiss
Elowitz
Voigt
Collins



Stochasticity/Cell

Pedraza, 2005, SCIENCE
Kaern, 2005, NAT REV GENET (Collins)
Blake, 2003, NATURE
Rosenfeld, 2002, SCIENCE
Isaacs, 2003, PNAS (Collins-Cantor-Hasty)
Ozbudak, 2002, NAT GENET (Grossman-van Oudenaarden)
Thattai, 2001, PNAS (Thattai)
Elowitz, 2002, SCIENCE (Elowitz)
Gillespie, 1977, J PHYS CHEM-US (Gillespie)
McAdams, 1997, PNAS (Arkin)
Shen-orr, 2002, NAT GENET (Shen-orr-Herer)
Arkin, 1998, GENETICS (Arkin)



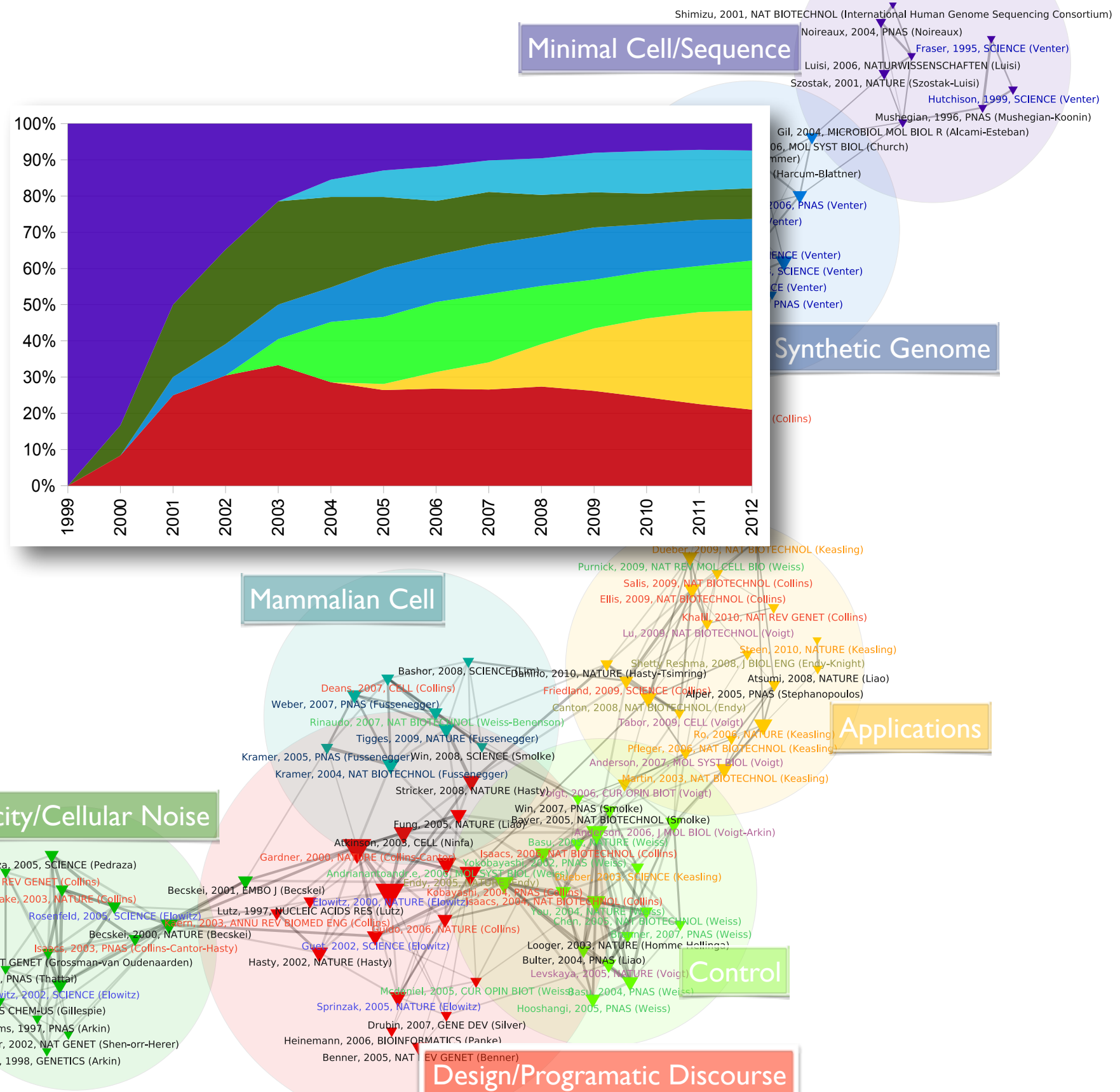
Control

Design/Programatic Discourse

Guell, 2002, SCIENCE (Elowitz)
Hasty, 2002, NATURE (Hasty)
Looger, 2003, NATURE (Homme Hallings)
Bulter, 2004, PNAS (Liao)
Levskaya, 2005, NATURE (Voigt)
McDaniel, 2005, CUR OPIN BIOT (Weiss)
Sprinzak, 2005, NATURE (Elowitz)
Hooshangi, 2005, PNAS (Weiss)
Drubin, 2007, GENE DEV (Silver)
Heinemann, 2006, BIOINFORMATICS (Panke)
Benner, 2005, NAT REV GENET (Benner)

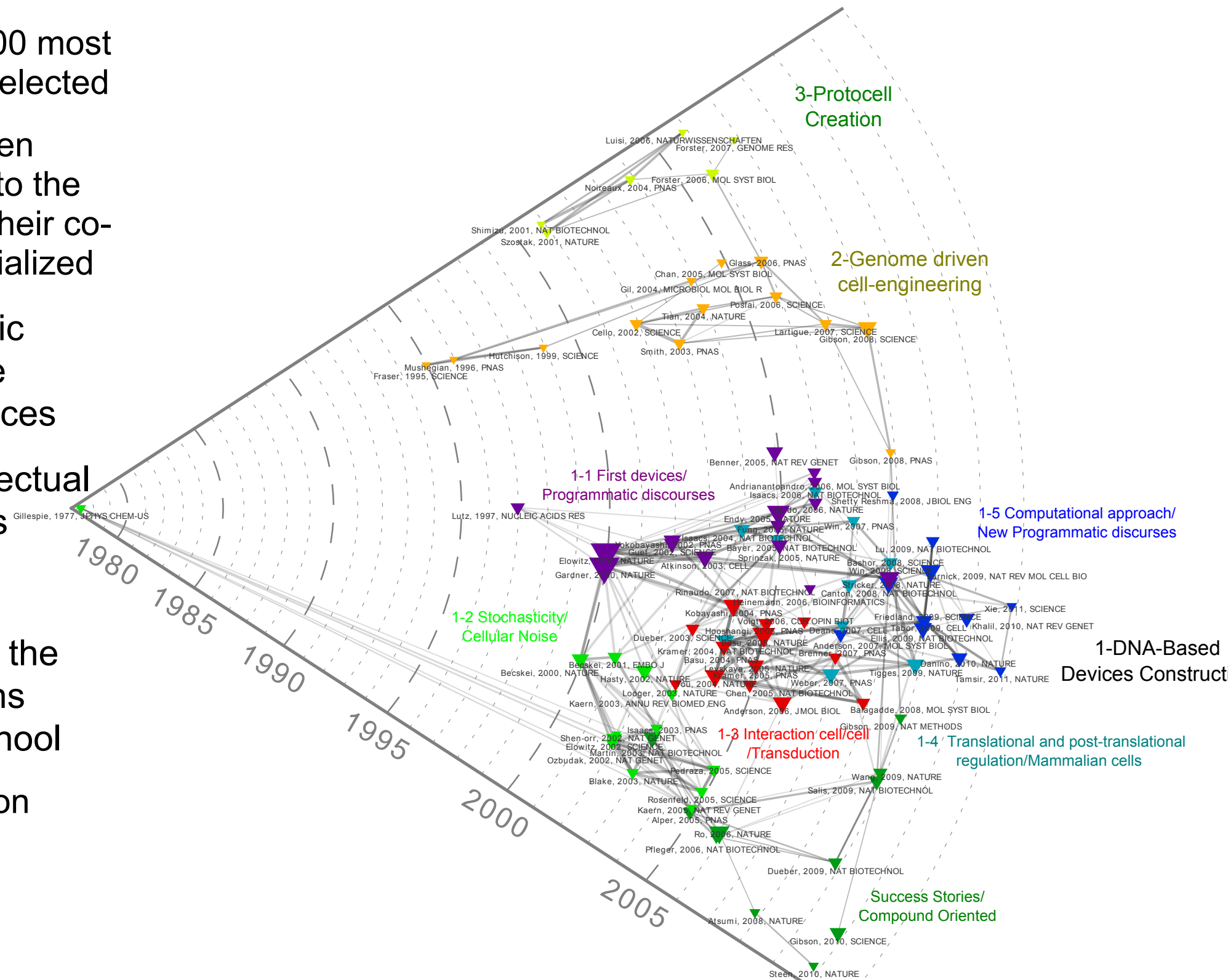
Scientific references co-citation map

- **Selection:** only the 100 most cited references are selected
- **Network:** They are then connected according to the relative frequency of their co-occurrences and spatialized
- **Clustering:** Automatic detection of cohesive subgroups of references
- **Interpretation:** intellectual schools, major actors
- **Dynamics:**
 - > Time Evolution of the number of publications belonging to each school



Scientific references co-citation map

- **Selection:** only the 100 most cited references are selected
- **Network:** They are then connected according to the relative frequency of their co-occurrences and spatialized
- **Clustering:** Automatic detection of cohesive subgroups of references
- **Interpretation:** intellectual schools, major actors
- **Dynamics:**
 - > Time Evolution of the number of publications belonging to each school
 - > radar representation

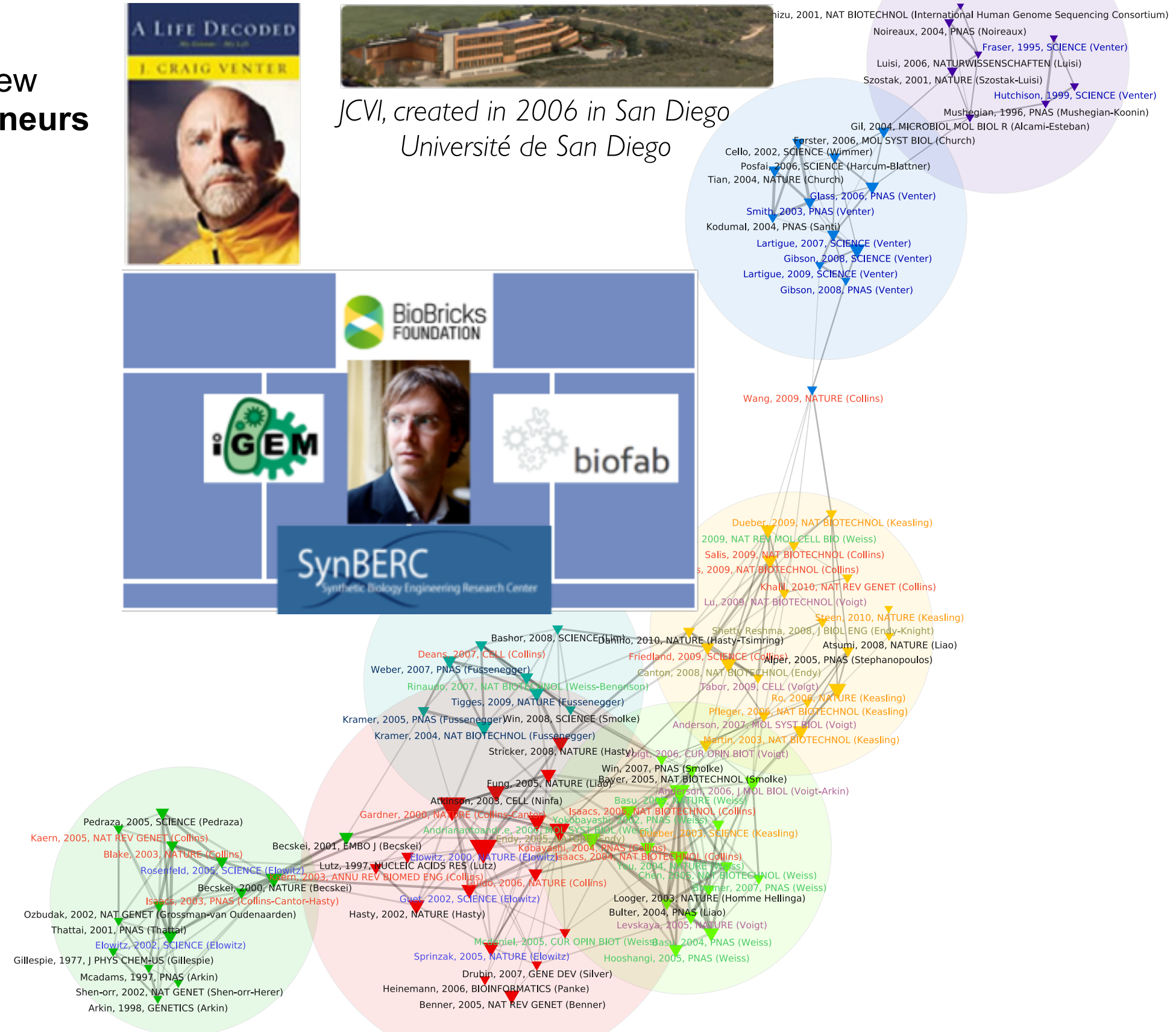
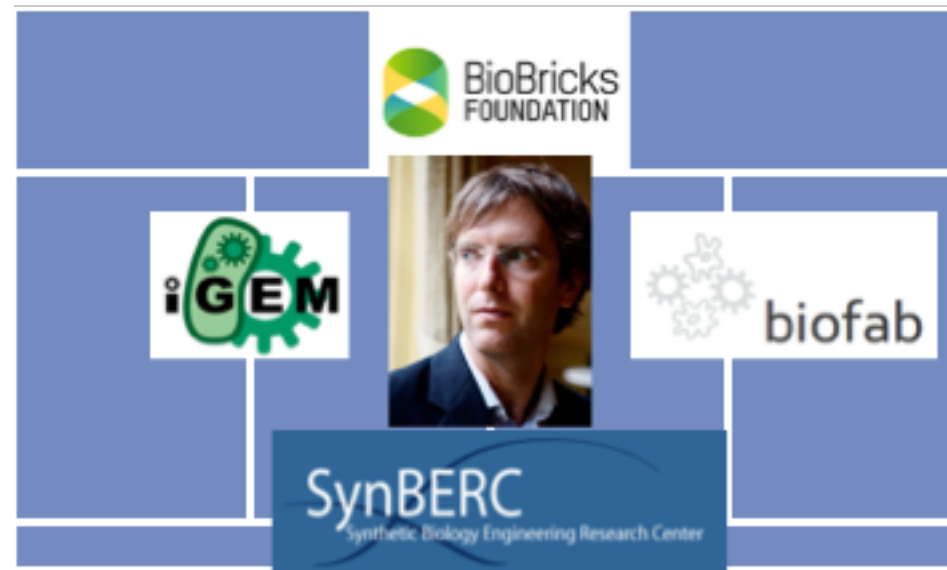


Actors and discourses

- Determining role of a few institutional entrepreneurs



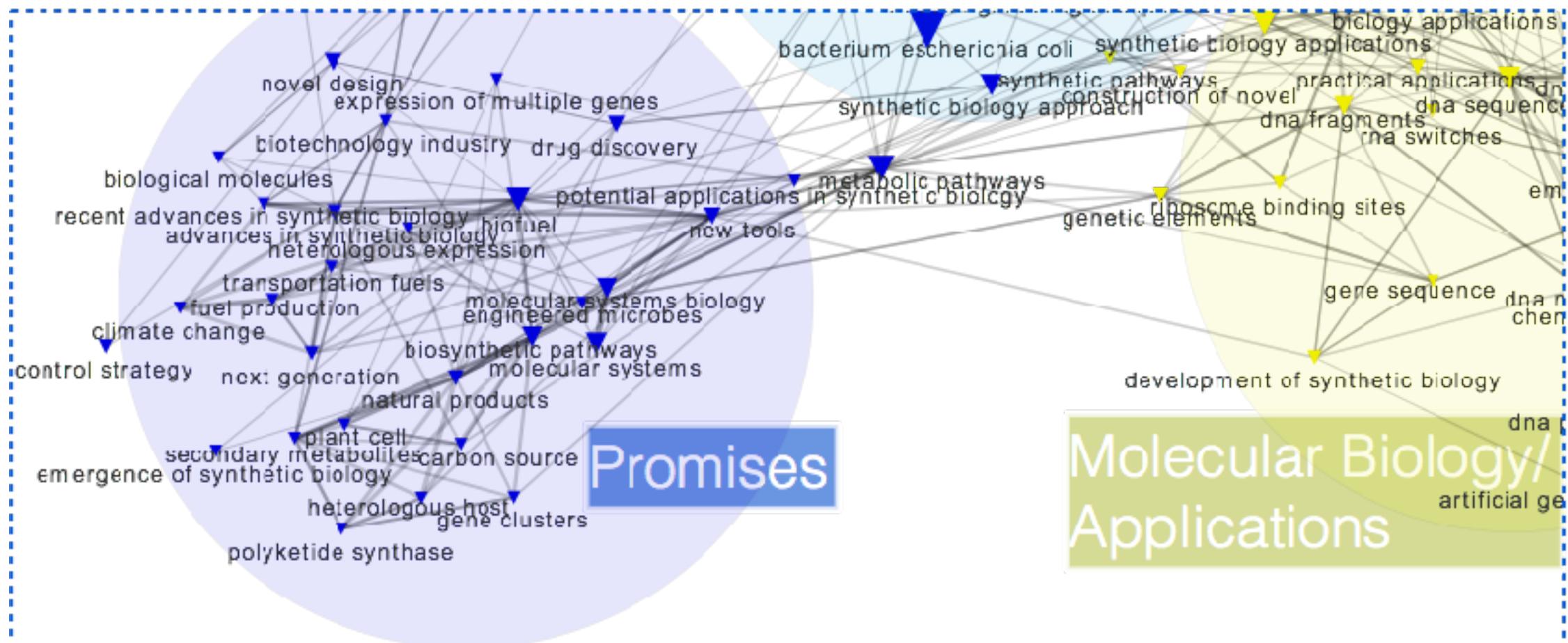
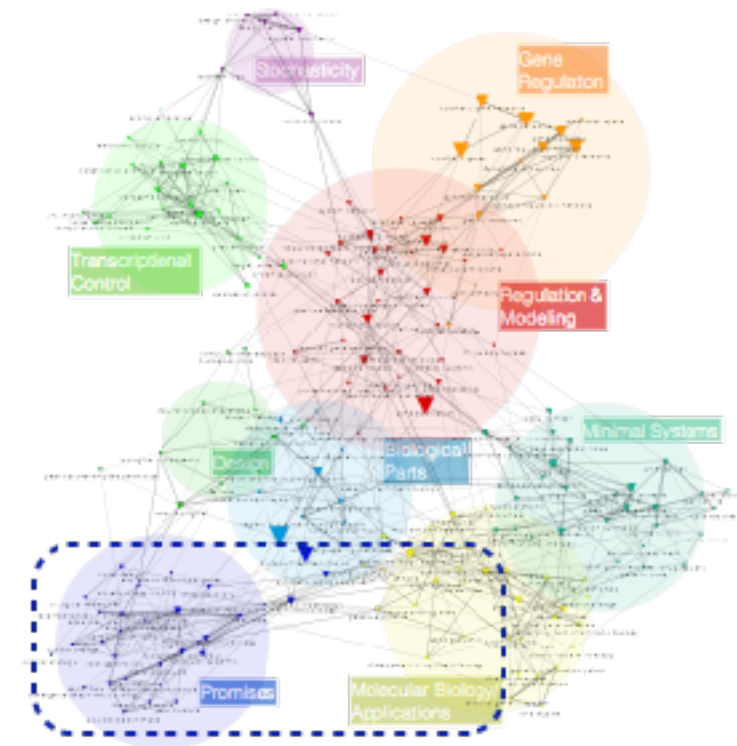
JCVI, created in 2006 in San Diego
Université de San Diego



Actors and discourses

- Determining role of a few **institutional entrepreneurs**
- Central role of **promises** discourses in the article

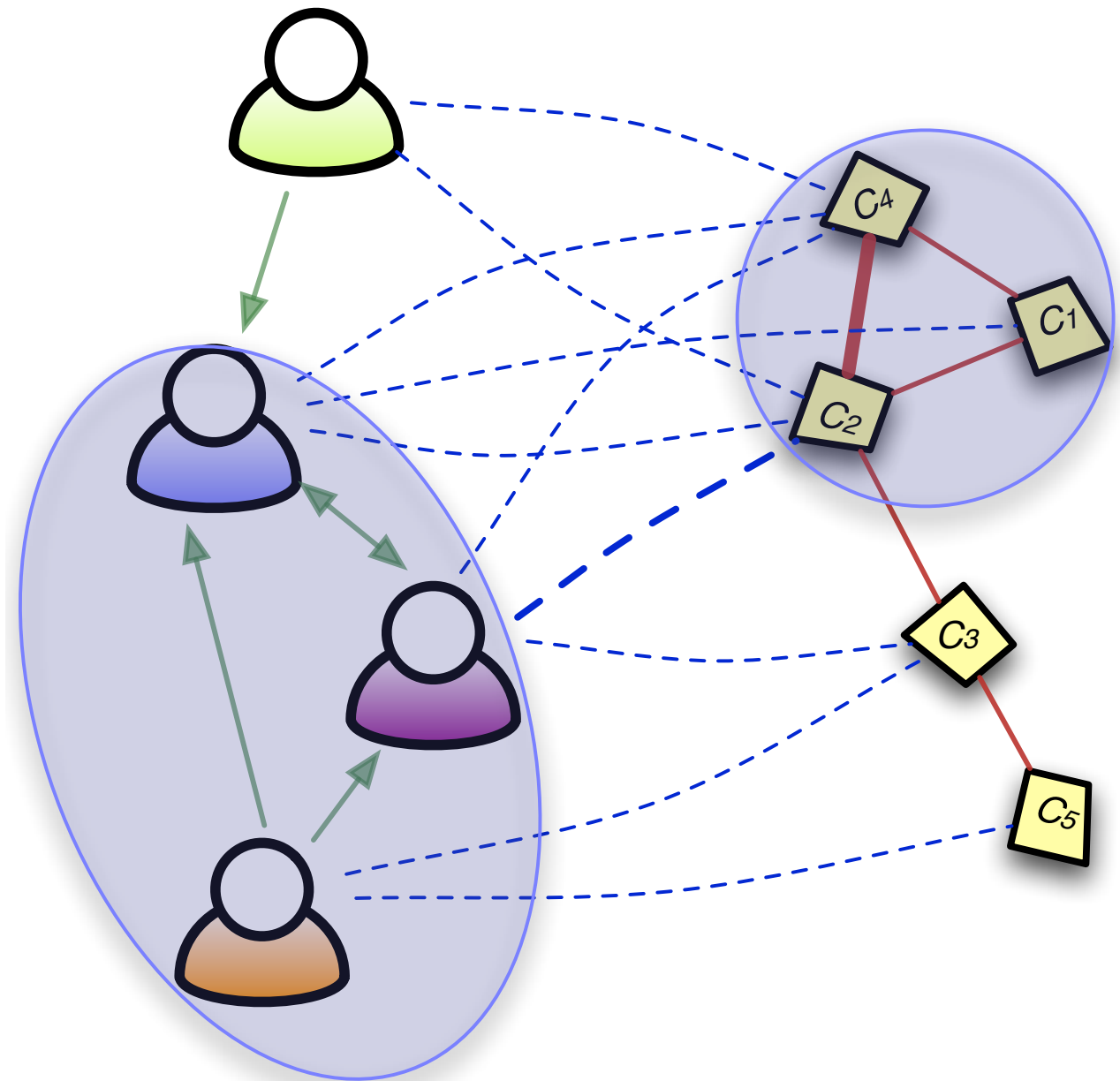
lexical map based on the analysis of titles and abstracts



Unveiling Networks Community Structure

Community Detection Algorithms

- Generalized co-occurrences analysis framework mixing people, terms, countries, etc...
- Clustering techniques are being used to circulate from micro to macro levels - clusters are made of possibly heterogeneous nodes associated in a singular manner.



Socio-semantic network and related clusters

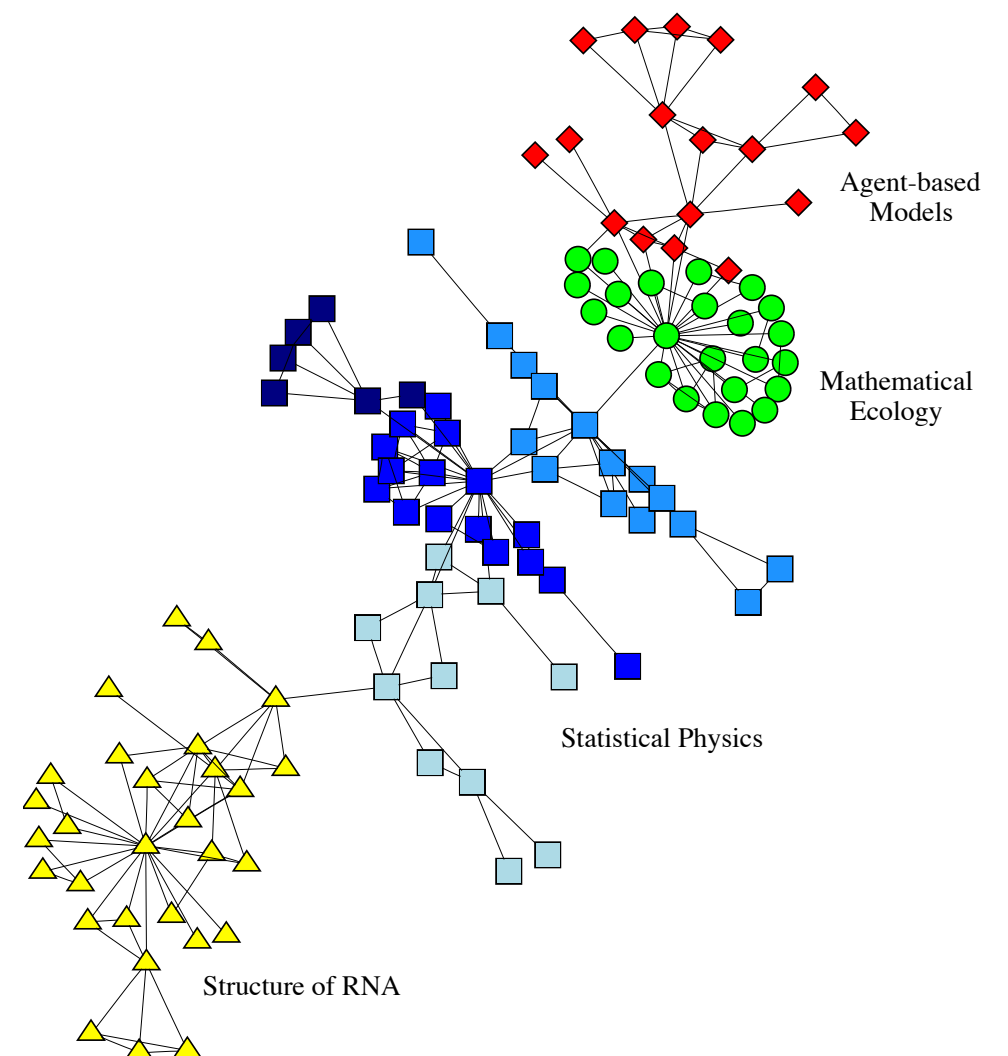
Community Detection Algorithms

- Modularity optimization (algorithmic definition)

$$Q \equiv \frac{1}{2|E|} \sum_c \sum_{i,j \in c} (A_{ij} - P_{ij})$$

$$P_{ij} \equiv \frac{d_i d_j}{2|E|}$$

Network Modularity Measure

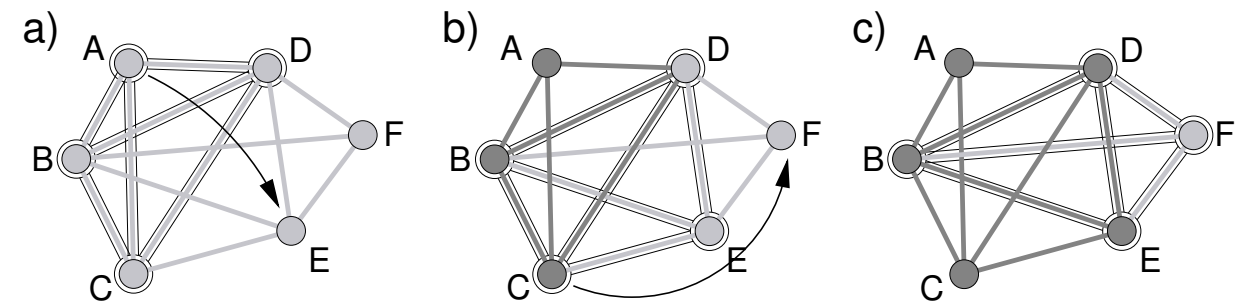


Santa Fe Collaboration Network

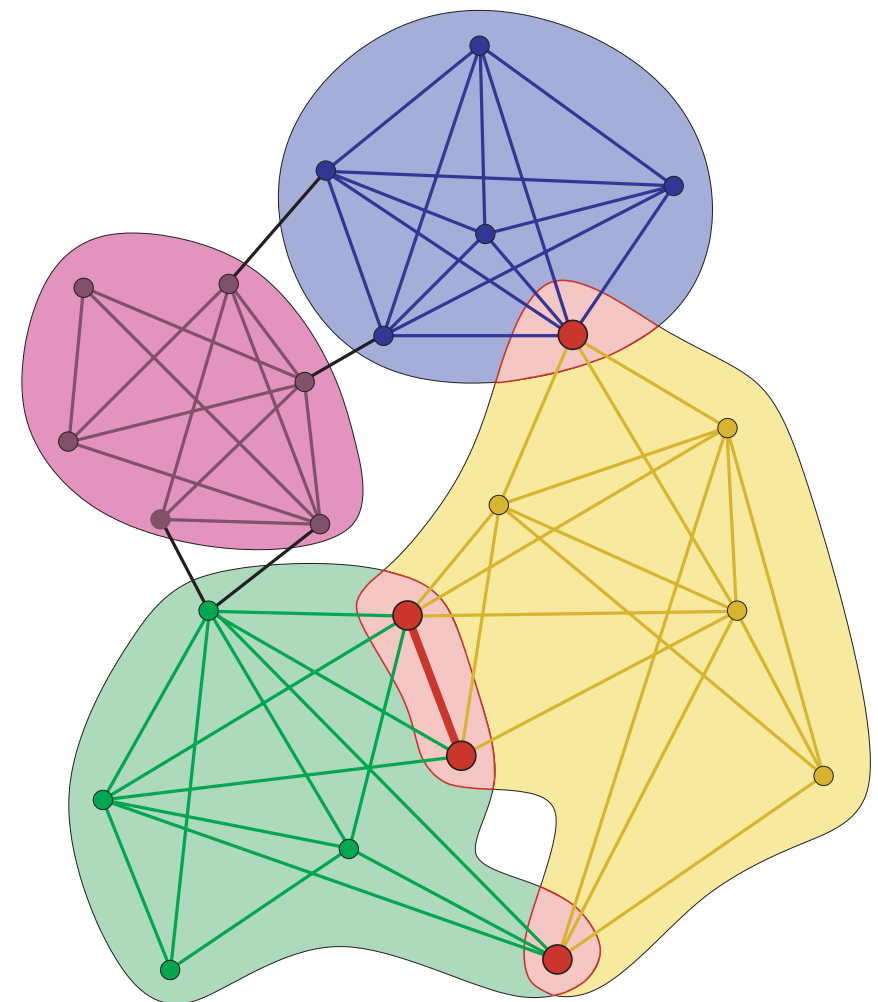
Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 7821–7826.

Community Detection Algorithms

- Modularity optimization (algorithmic definition)
- Clique Percolation (algebraic definition)



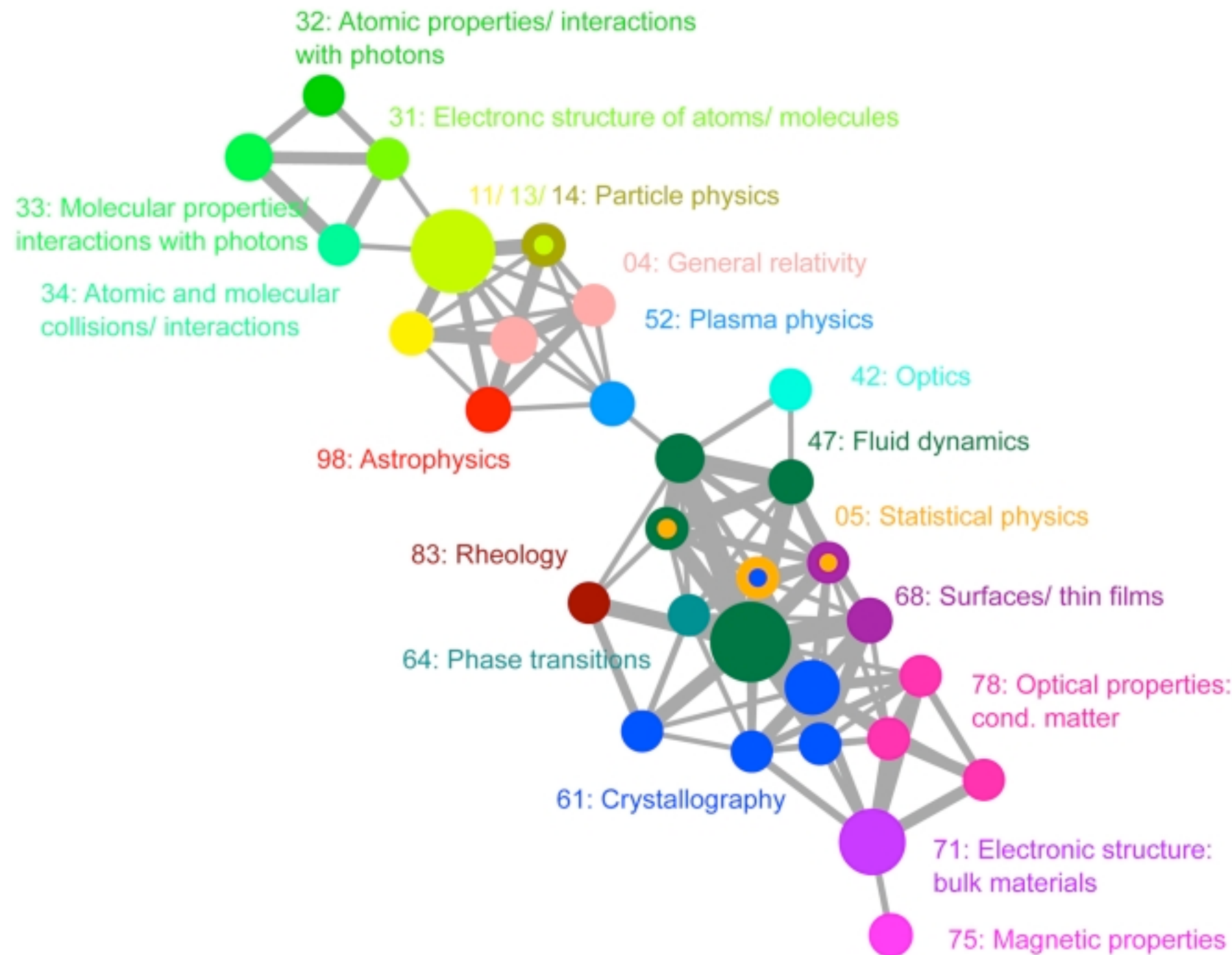
Rolling 4-clique



Clique percolation communities

Community Detection Algorithms

- Modularity optimization (algorithmic definition)
- Clique Percolation (algebraic definition)

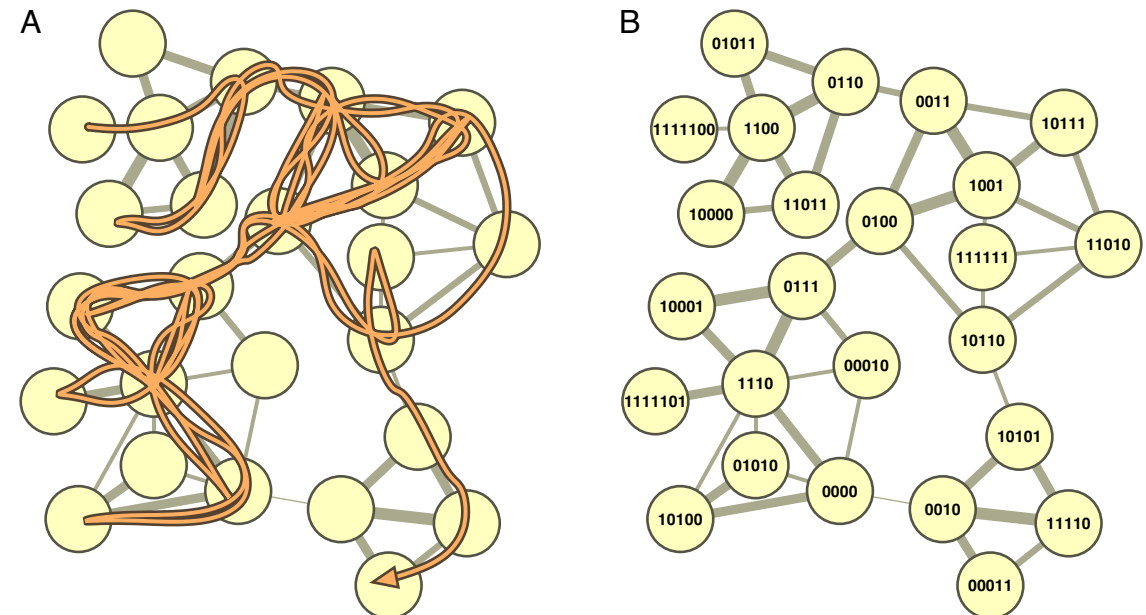


ex: Herrera, Mark, David C. Roberts, and Natali Gulbahce. Mapping the evolution of scientific fields. *PloS one* 5.5 (2010): e10355.

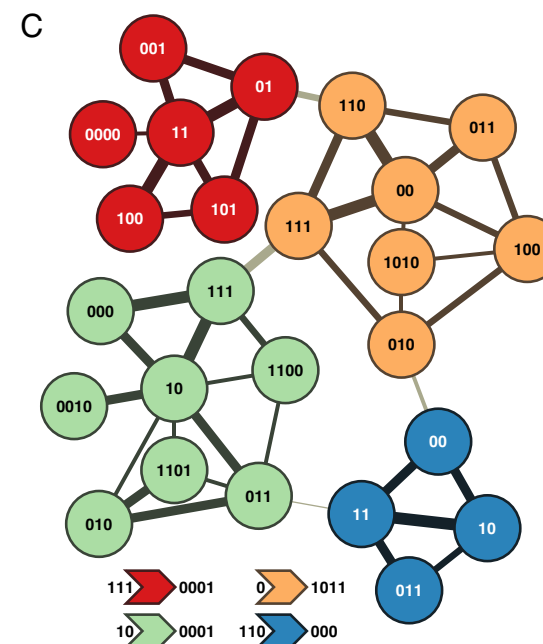
Map based on PACS code of APS publications

Community Detection Algorithms

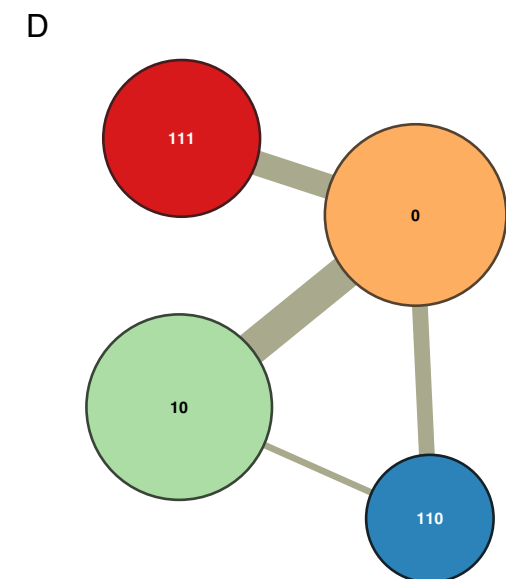
- Modularity optimization (algorithmic definition)
- Clique Percolation (algebraic definition)
- Maps of Random walks (information theoretic approach)



1111100 1100 0110 11011 10000 11011 0110 0011 10111 1001 0011
1001 0100 0111 10001 1110 0111 10001 0111 1110 0000 1110 10001 0111
0111 1110 0111 1110 1111101 1110 0000 10100 0000 1110 10001 0111
0100 10110 11010 10111 1001 0100 1001 10111 1001 0100 1001 0100
0011 0100 0011 0110 11011 0110 0011 0100 1001 10111 0011 0100
0111 10001 1110 10001 0111 0100 10110 111111 10110 10101 11110
00011



111 0000 11 01 101 100 101 01 0001 0 110 011 00 110 00 111 1011 10
111 000 10 111 000 111 10 011 10 000 111 10 111 10 0010 10 011 010
011 10 000 111 0001 0 111 010 100 011 00 111 00 011 00 111 00 111
110 111 110 1011 111 01 101 01 0001 0 110 111 00 011 110 111 1011
10 111 000 10 000 111 0001 0 111 010 1010 010 1011 110 00 10 011



111 0000 11 01 101 100 101 01 0001 0 110 011 00 110 00 111 1011 10
111 000 10 111 000 111 10 011 10 000 111 10 111 10 0010 10 011 010
011 10 000 111 0001 0 111 010 100 011 00 111 00 011 00 111 00 111
110 111 110 1011 111 01 101 01 0001 0 110 111 00 011 110 111 1011
10 111 000 10 000 111 0001 0 111 010 1010 010 1011 110 00 10 011

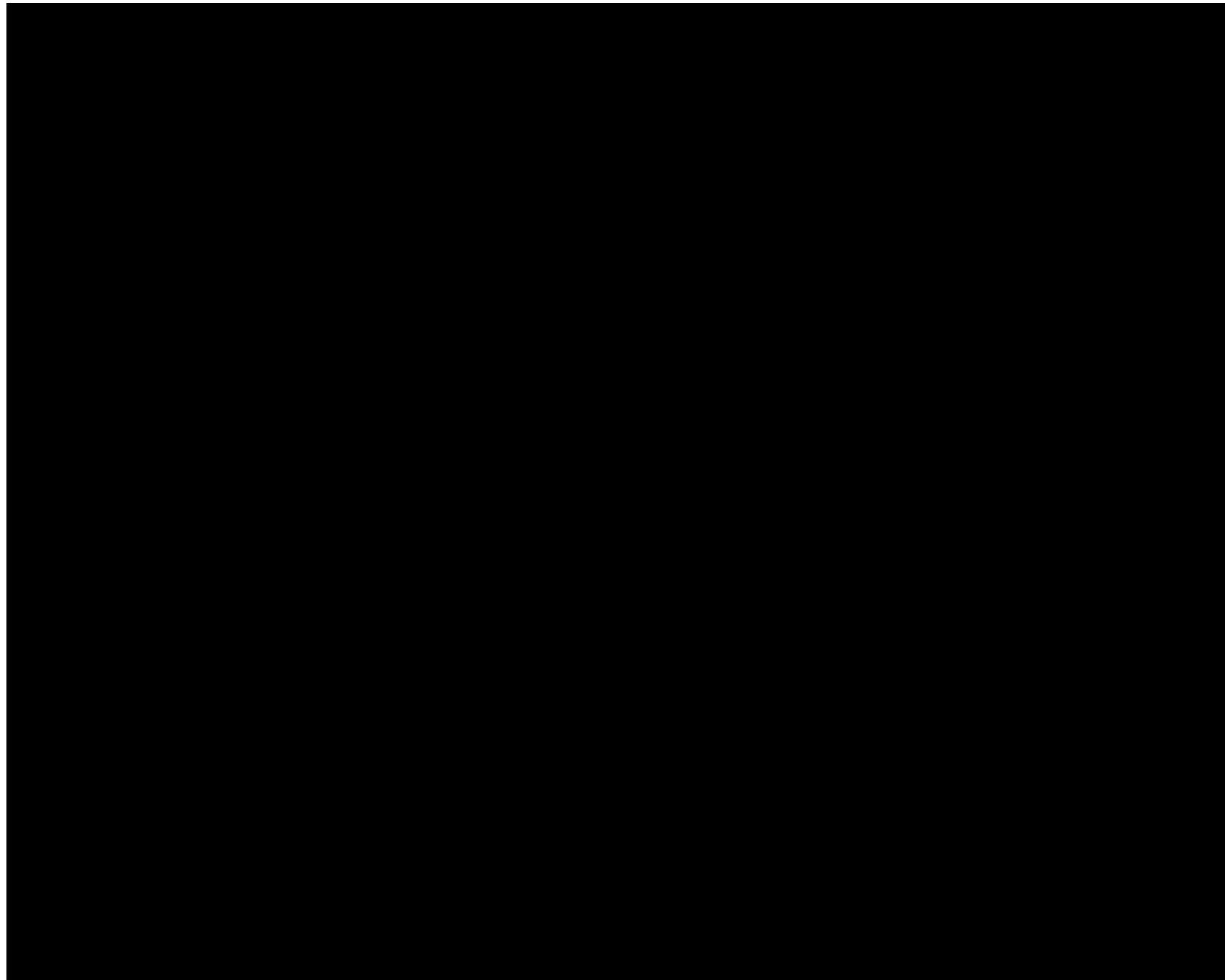
Rosvall, M., & Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(4), 1118.

compressing probability flow description

Spatializing Networks

Mapping Networks

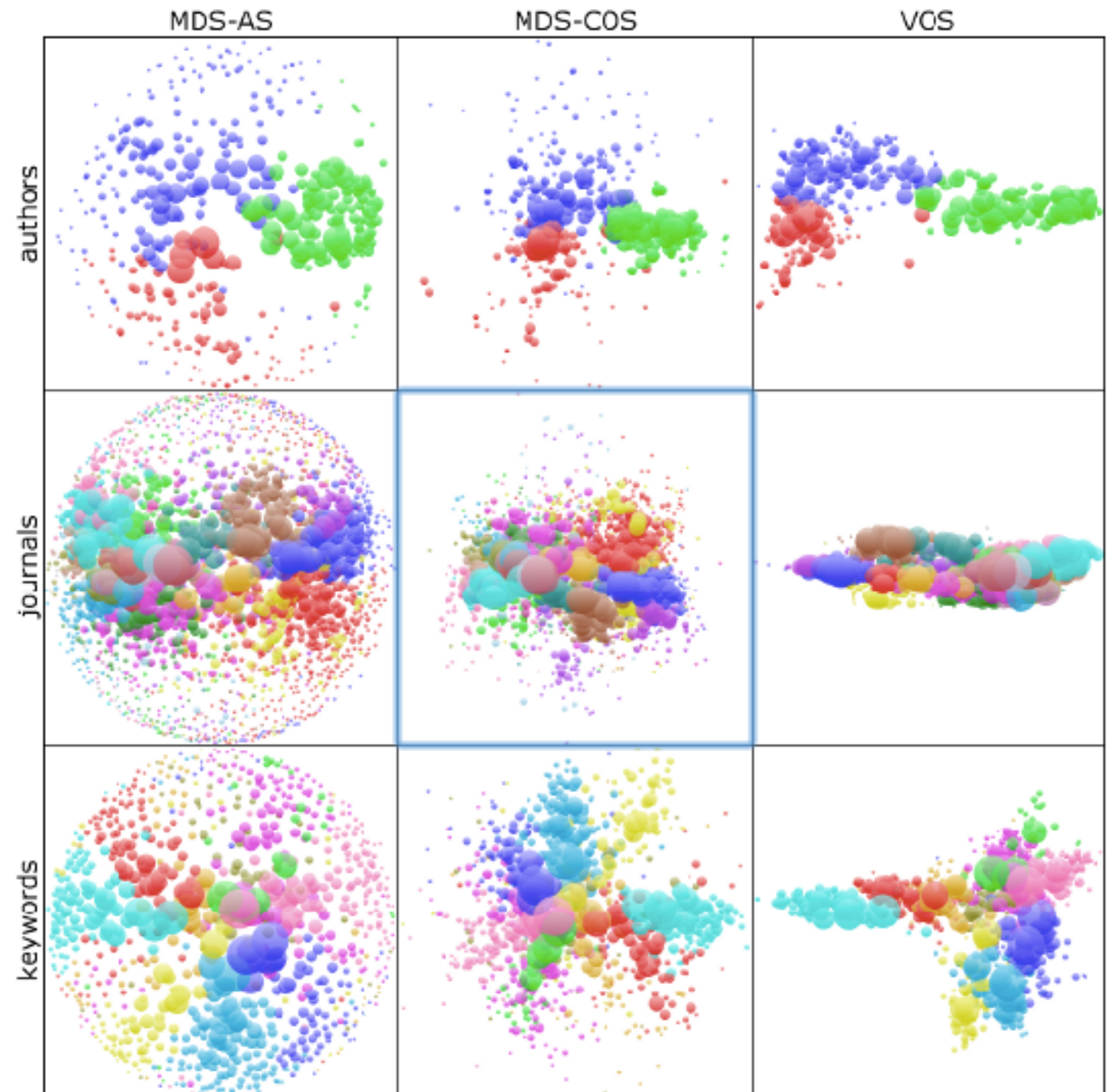
- Force-directed graph drawing



Gephi demo

Mapping Networks

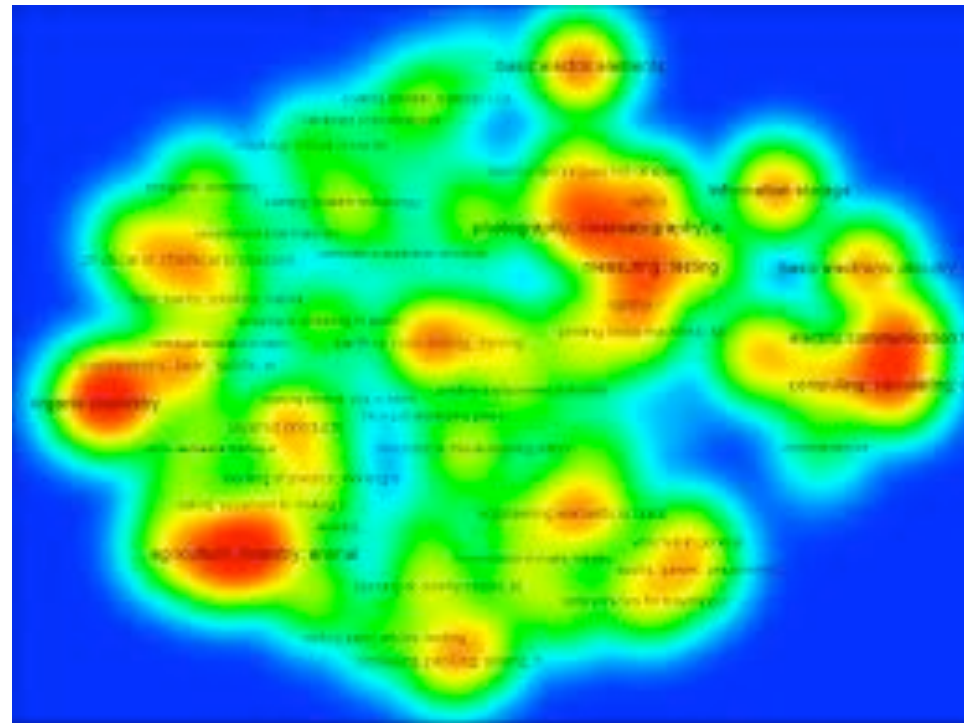
- Force-directed graph drawing
- Multidimensional Scaling and related techniques



Van Eck, N.J., & Waltman, L. (2010) "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Vol 84, No. 2, pp. 523–538.

Mapping Networks

- Force-directed graph drawing
- Multidimensional Scaling and related techniques



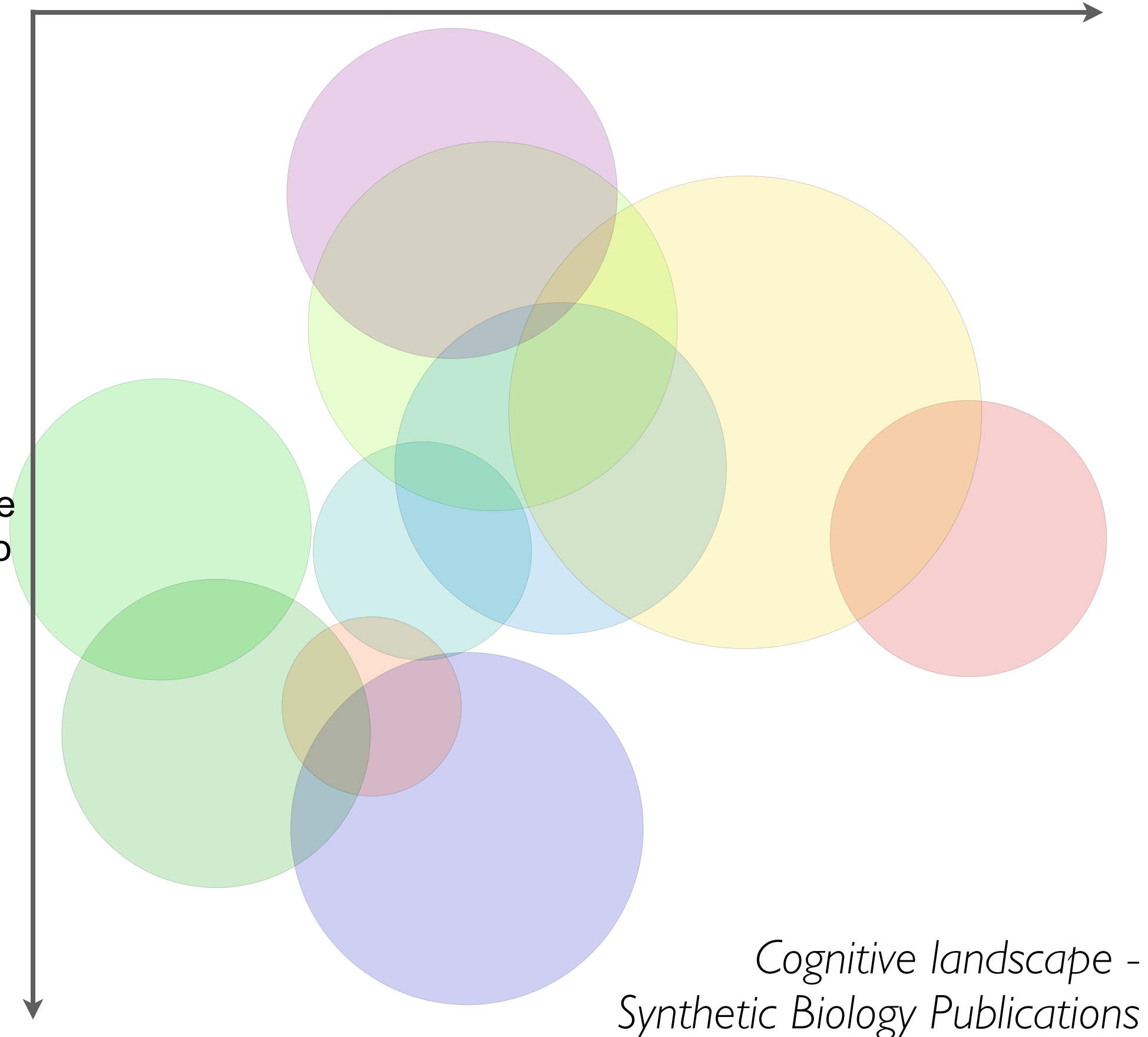
Van Eck, N.J., & Waltman, L. (2010)
“Software survey: VOSviewer, a computer
program for bibliometric mapping”,
Scientometrics, Vol 84, No. 2, pp. 523–538.



Skupin, A., Biberstine, J. R., & Börner, K.
(2013). Visualizing the Topical Structure of
the Medical Sciences: A Self-Organizing
Map Approach. *PLoS ONE*, 8(3), e58779.
doi:10.1371/journal.pone.0058779.s001

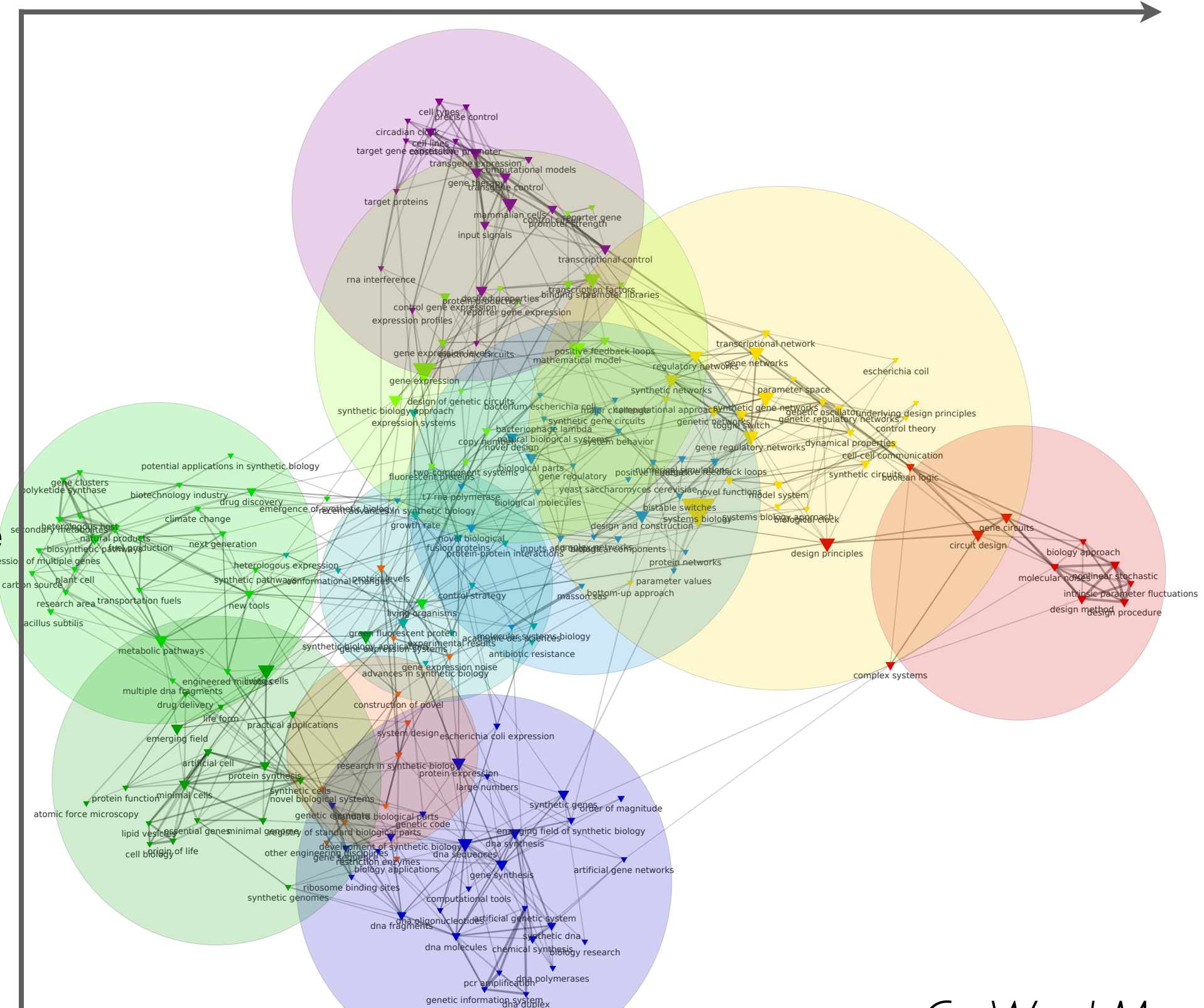
From Communities to Network Mapping

- Force-directed graph drawing
- Multidimensional Scaling and related techniques
- Actors, (here terms) are spatialized according to the position of their communities



From Communities to Network Mapping

- Force-directed graph drawing
- Multidimensional Scaling and related techniques
- Actors, (here terms) are spatialized according to the position of their communities



Co-Word Map
synthetic biology publications

Textual Analysis

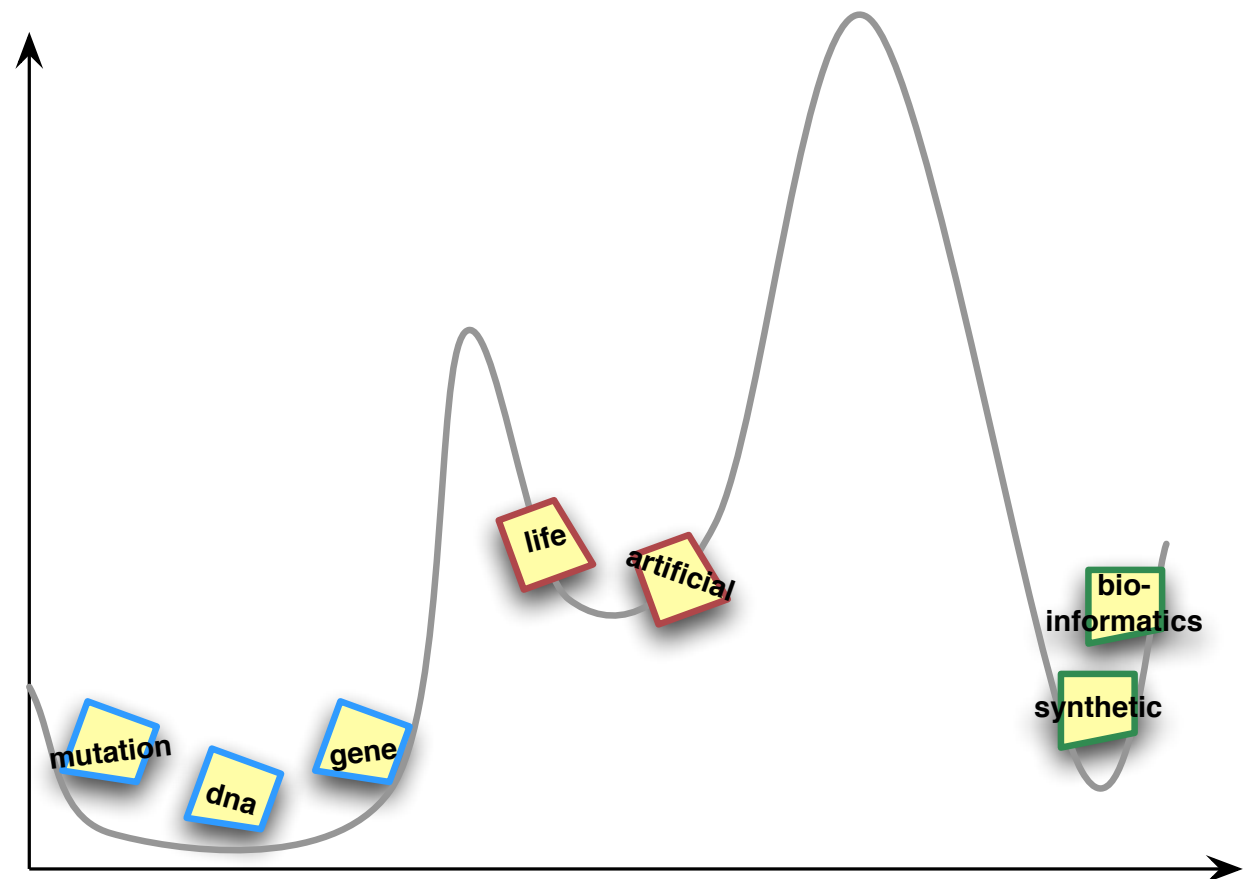
Textual Analysis

«Indexing is an intervention between the text and the co-word analysis, and the validity of the map will depend, to a certain extent, on the nature of the indexing. Yet since indexers try to capture what it is about a text that is interesting, they partially reproduce the readings that the texts are given within the field itself'. Thus, despite the fact that indexing is not entirely reliable, validity is never totally absent.»

Callon, M.; Law, J.; & Rip, A. (Eds.). (1986a). *Mapping the dynamics of science and technology: Sociology of Science in the real world*. London: The Macmillan Press Ltd.

what it is about a text that is interesting?

- **grammatical** criterion, candidate terms are usually limited *noun phrases*,
- **unithood**, phrases should represent a proper semantic unit,
- **termhood**, terms should be domain specific to carry substantial information



Textual Analysis

We believed we could reduce our dependence on foreign oil and protect our planet. And today, America is number one in oil and gas.

Textual Analysis

Part-Of-Speech Tagging

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil

CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in

NN CC NN
oil and gas.

Textual Analysis

Chunking

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil
CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in
NN CC NN
oil and gas.

Extracted noun phrases:

- *dependence*
- *planet*
- *oil*
- *gas*

Textual Analysis

Chunking

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil

CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in

NN CC NN
oil and gas.

Extracted noun phrases:

- *dependence*
- *planet*
- *oil*
- *gas*
- *foreign oil*

Textual Analysis

Chunking

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil

CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in

NN CC NN
oil and gas.

Extracted noun phrases:

- *dependence*
- *planet*
- *oil*
- *gas*
- *foreign oil*
- *dependence on foreign oil*
- *oil and gas*

Textual Analysis

Stemming, Filtering and Standardizing

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil

CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in

NN CC NN
oil and gas.

Extracted classes:

- *dependence on foreign oil: {dependence on foreign oil ; foreign oil dependence}*
- *oil and gas: {oil and gas; gas and oil}*
- *planet: {planet, planets}*
- *etc.*

Textual Analysis

Unithood: extracting semantic units with C-value

- Simple frequency-based approach : «Real» Terms tend to appear more frequently than non-terms
- C-value approach (Frantzi K. & Ananiadou S., 2000):
 - Longer phrases are more likely to be relevant,
 - Nested terms may induce false positive, ex: self organizing maps.

Textual Analysis

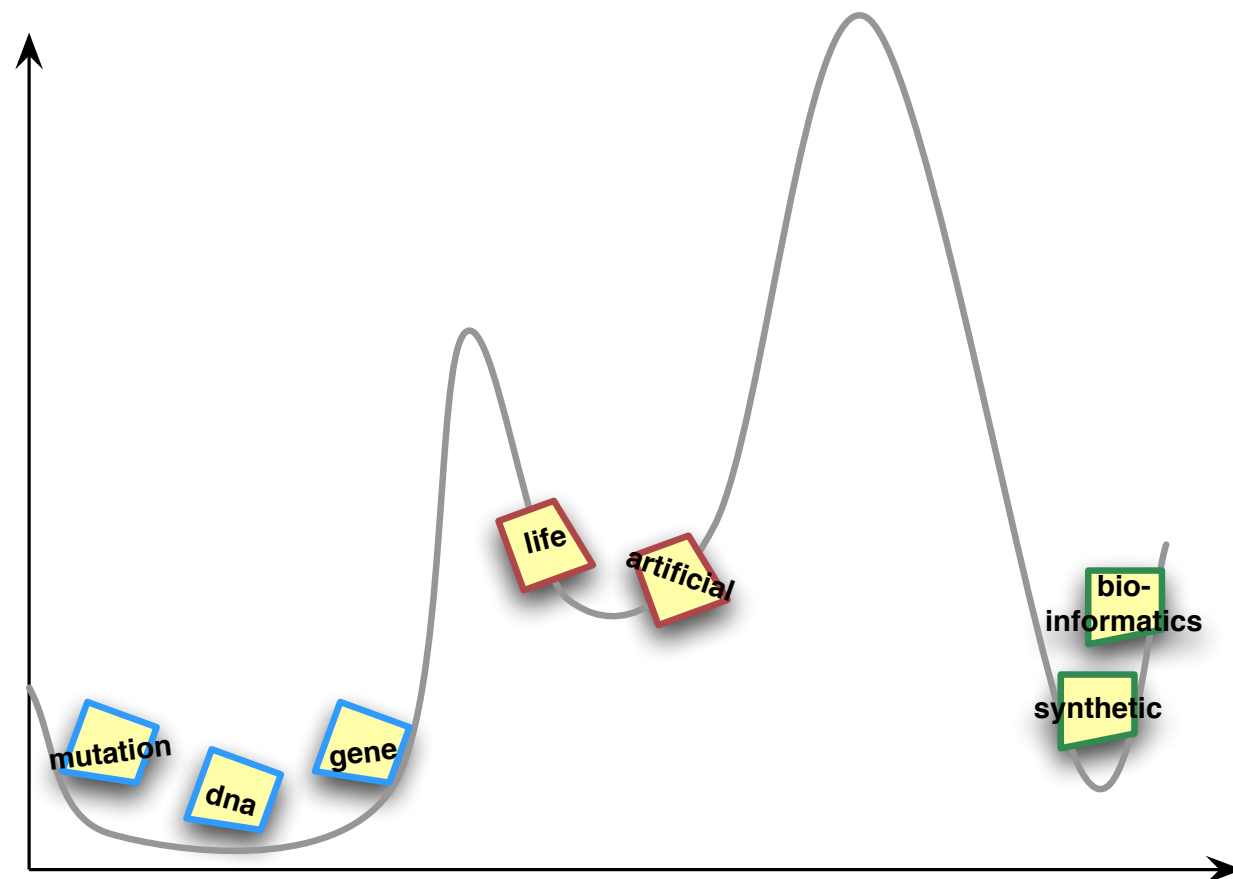
Termhood

- Candidate terms should be thematically specific ; terms not specific to a specific thematic subfield have neutral meaning given the whole domain and should be excluded
- On the contrary, terms which distribution is biased toward certain topics are more likely to have interesting meaning.
- Co-occurrences between existing candidate terms are extracted to compute the **Xhi2 score** of specificity of each term compared to other terms (Matsuo Y. & Ishizuka M., 2004).

$$\chi^2(w) = \sum_{g \in G} \frac{(\text{freq}(w, g) - n_w p_g)^2}{n_w p_g}$$

Textual Analysis

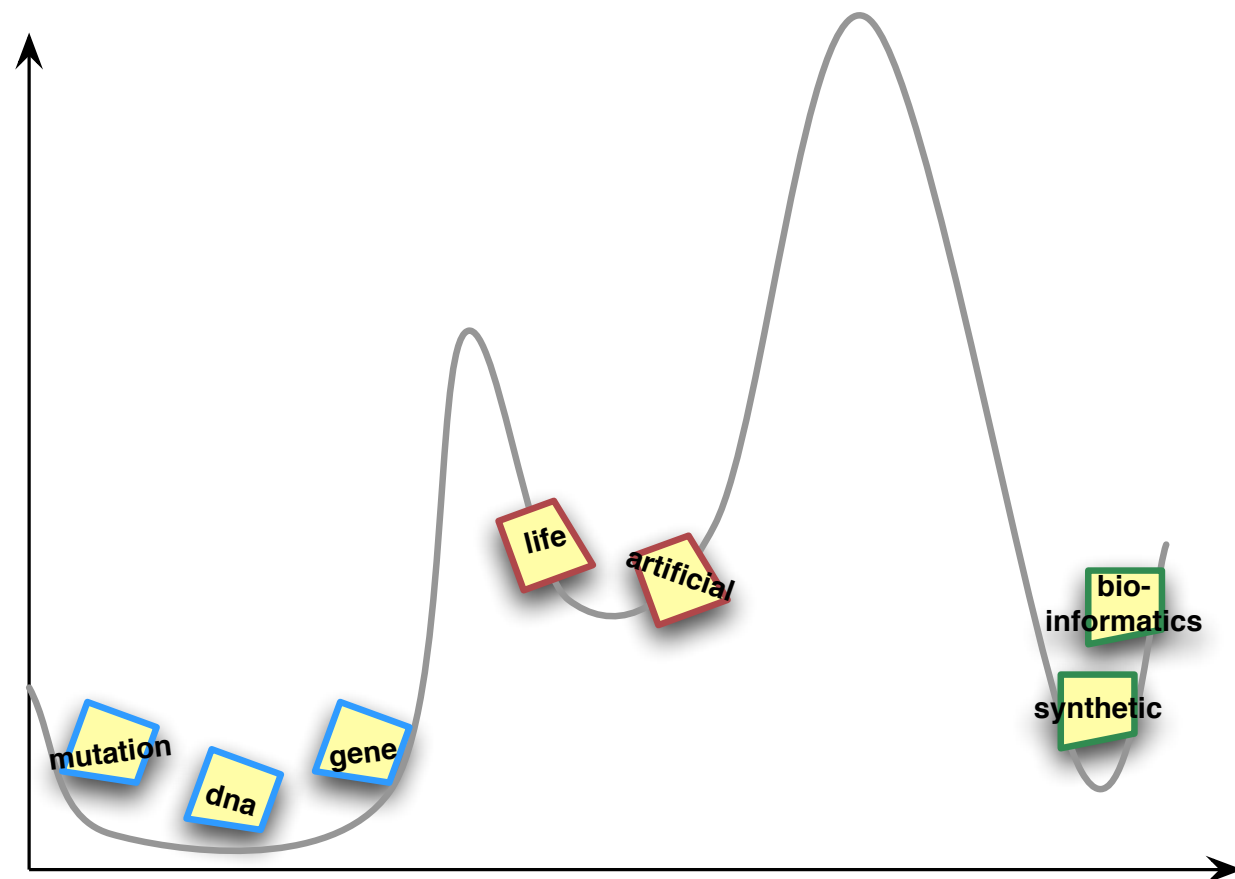
NLP analysis outcome



label	forms
heterologous expression	heterologous expression
complex networks	complex networks; complex network
design principles	design principle; design principles
transgene control	transgene control
gene expression noise	gene expression noise; noise in gene expression; gene-expression noise; noise in gene-expression
biosynthetic pathways	biosynthetic pathways; biosynthetic pathway
mathematical model	mathematical model; mathematical models; Mathematical models
metabolic pathways	metabolic pathway; metabolic pathways
control theory	control theory
reporter gene expression	reporter gene expression; reporter gene-expression
gene expression systems	gene expression systems; gene-expression systems
biological molecules	biological molecules; biological molecule
control circuit	control circuit; control circuits; circuit from a control; circuits control
toggle switch	Toggle switch; toggle switch; toggle switches
protein networks	protein networks; networks of proteins; proteins and networks
restriction enzymes	restriction enzymes; restriction enzyme
transcriptional network	transcriptional networks; transcriptional network; transcription networks
system design	system design; design these systems; systems design
Bacillus subtilis	Bacillus subtilis; bacillus subtilis
multiple DNA fragments	multiple DNA fragments; multiple dna fragments
drug delivery	drug delivery; delivery of drugs
biological clock	biological clock; biological clocks; Biological Clocks
mammalian cells	mammalian cells; Mammalian Cells
fuel production	production of fuels; Production of fuels; fuel production
artificial gene networks	artificial gene networks
control strategy	control strategy; control strategies; strategies that control
gene expression levels	gene expression levels; level of gene expression; gene-expression levels; level of gene-expression
computational approach	computational approach; Computational approaches; computational approaches; computational approach-; computational Approach; Computational approach-; Computational approach; Computational Approach
DNA synthesis	DNA synthesis; dna synthesis
gene networks	gene networks; gene network; Gene networks; networks of genes; Gene Network; genes and networks
biological parts	Biological Parts; biological parts
genetic information system	genetic information system; genetic -information system; genetic Information system
engineered microbes	engineered microbes; Engineering microbes; engineering microbes
Registry of Standard Biological Parts	registry of standard biological parts; Registry of Standard Biological Parts; Registry of Standard-Biological Parts; Registry of STANDARDS Biological Parts; Registry of standards Biological Parts; Registry of Standards Biological Parts

Textual Analysis

NLP analysis outcome



label	forms
abundance	abundance
acres	acres&lacre
aliens	aliens&lalien
alliances	alliances&llalliance
allies	allies&lally
American life	American lifel&lAmerican lives
armed forces	armed forces&larmed force&lforce of arms&larms and force
boundary line	boundary linel&lline of boundary
business men	business menl&lbusiness manl&lmen of businessl&lman of business
children	childrenl&lchild
commerce and navigation	commerce and navigationl&lavigation and commercel&lavigation or commerce
construction work	construction workl&lconstruction of worksl&lwork of constructionl&lworks constructionl&lconstruction of these worksl&lwork in construction
crews	crewsl&lcrew
crime	crimel&lcrimes
crisis	crisisl&lcrises
crops	cropsl&lcrop
cruisers	cruisersl&lcruiser
democracy	democracyl&ldemocracies
diplomatic relations	diplomatic relations
expenses of the Government	expenses of the Governmentl&lGovernment expensesl&lGovernment expendel&lexpenses of Governmentl&lexpense of the Government
farm products	farm productsl&lproducts of the farml&lproducts of farm
great importance	great importancel&lgreater importancel&lreatest importance
health care costs	health care costsl&lcost of health carel&lhealth care cost
income tax	income taxl<he income<he income tax
peace and freedom	peace and freedoml&lpeace and freedom&lpeace with freedom
property rights	property rightsl&lproperty rightl&lrights of propertyl&lright of propertyl&lrights and property

Semantic Map

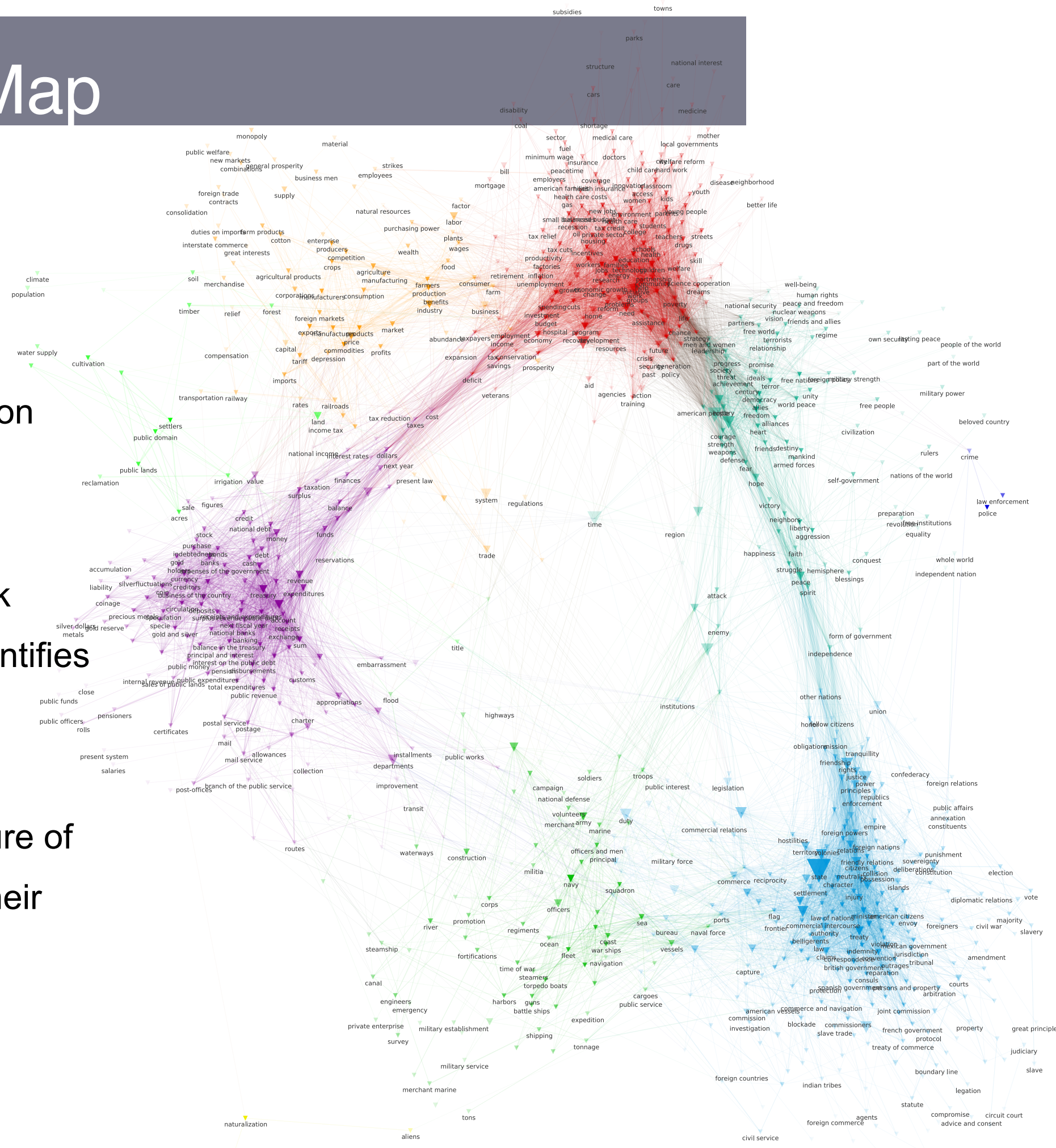
Four steps process:

1. **Similarity Network** based on terms cooccurrences

2. **Filter Edges** to obtain the sparsest connected network

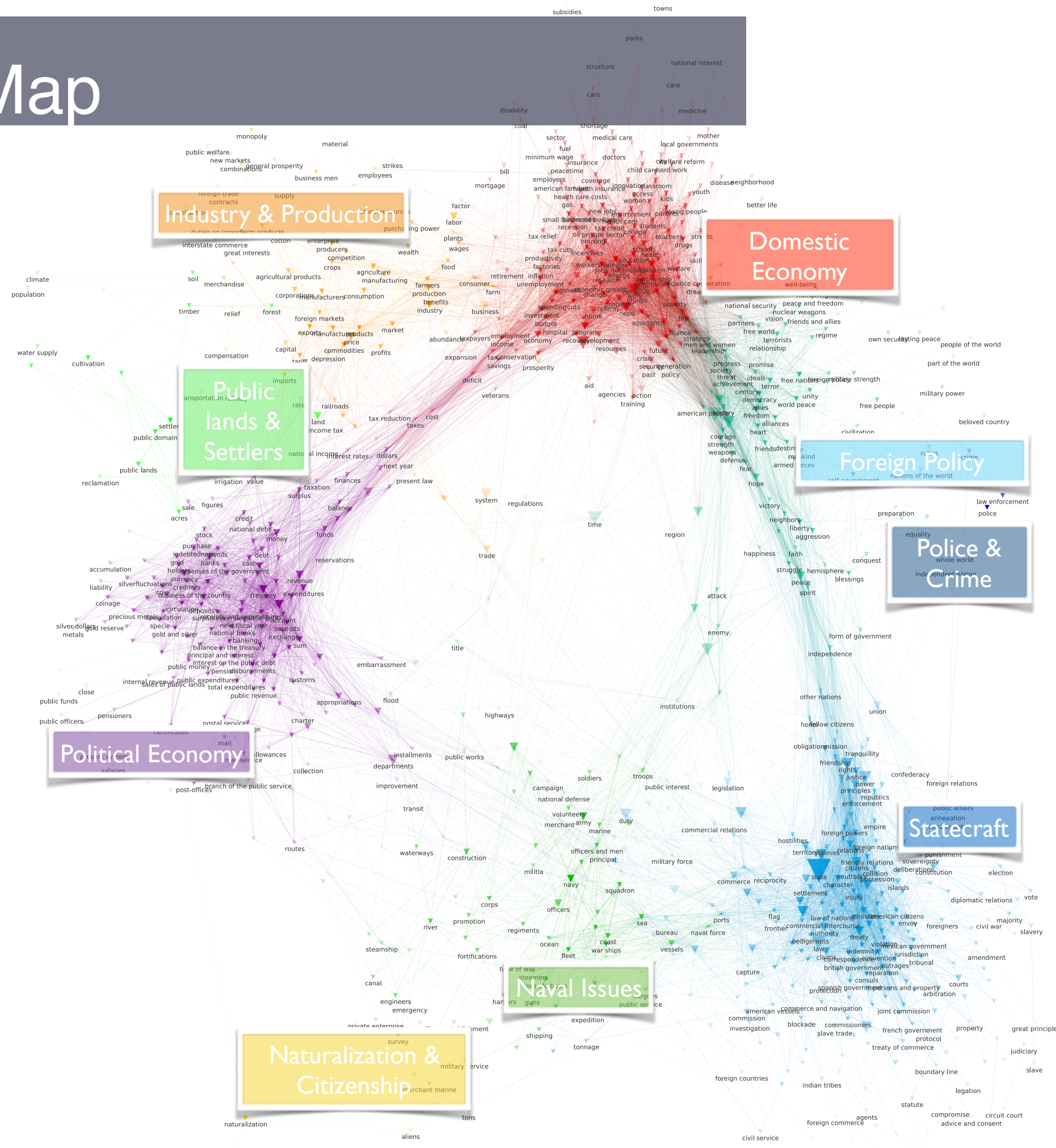
3. **Community detection** identifies cohesive thematic areas

4. **Mapping** allows to visually investigate the inner structure of those clusters, as well as their relationships



9 Topics:

- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy**— “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Public lands & Settlers** — “land”, “settlers”, “acres”
- **Naturalization & Citizenship** — “naturalization”, “aliens”
- **Police & Crime** — “criminal”, “law enforcement”



[illegible]

- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Public lands & Settlers** — “land”, “settlers”, “acres”
- **Naturalization & Citizenship** — “naturalization”, “aliens”
- **Police & Crime** — “criminal”, “law enforcement”

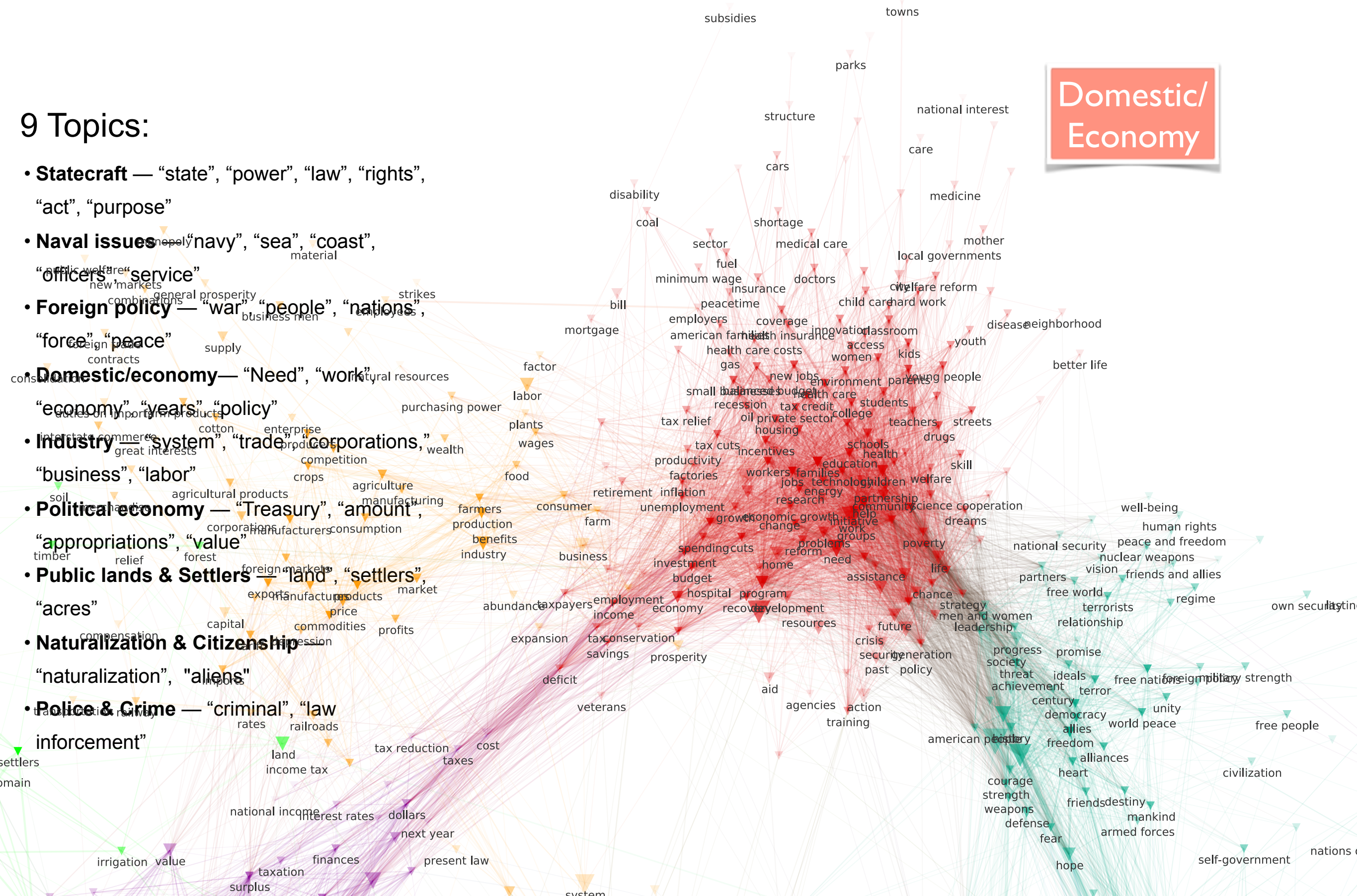
Political Economy

Semantic Map

Domestic/
Economy

9 Topics:

- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Public lands & Settlers** — “land”, “settlers”, “acres”
- **Naturalization & Citizenship** — “naturalization”, “aliens”
- **Police & Crime** — “criminal”, “law enforcement”



Semantic Map

9 Topics:

- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Public lands & Settlers** — “land”, “settlers”, “acres”
- **Naturalization & Citizenship** — “naturalization”, “aliens”
- **Police & Crime** — “criminal”, “law enforcement”

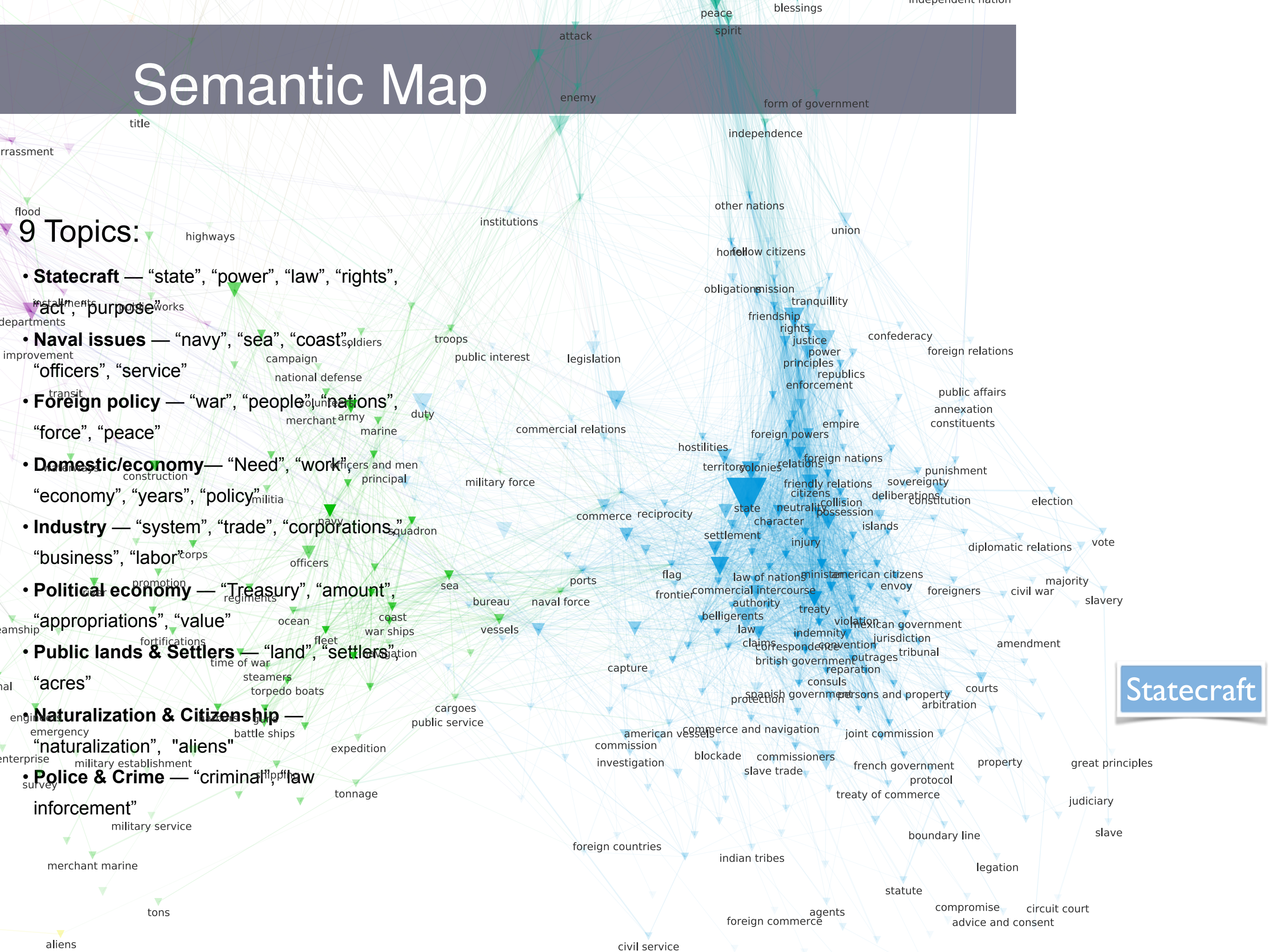
Foreign
Policy

Police &
Crime

Semantic Map

9 Topics:

- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Public lands & Settlers** — “land”, “settlers”, “acres”
- **Naturalization & Citizenship** — “naturalization”, “aliens”
- **Police & Crime** — “criminal”, “law”, “enforcement”



Statecraft

Semantic Map

9 Topics:

- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Public lands & Settlers** — “land”, “settlers”, “acres”
- **Naturalization & Citizenship** — “naturalization”, “aliens”
- **Police & Crime** — “criminal”, “law enforcement”

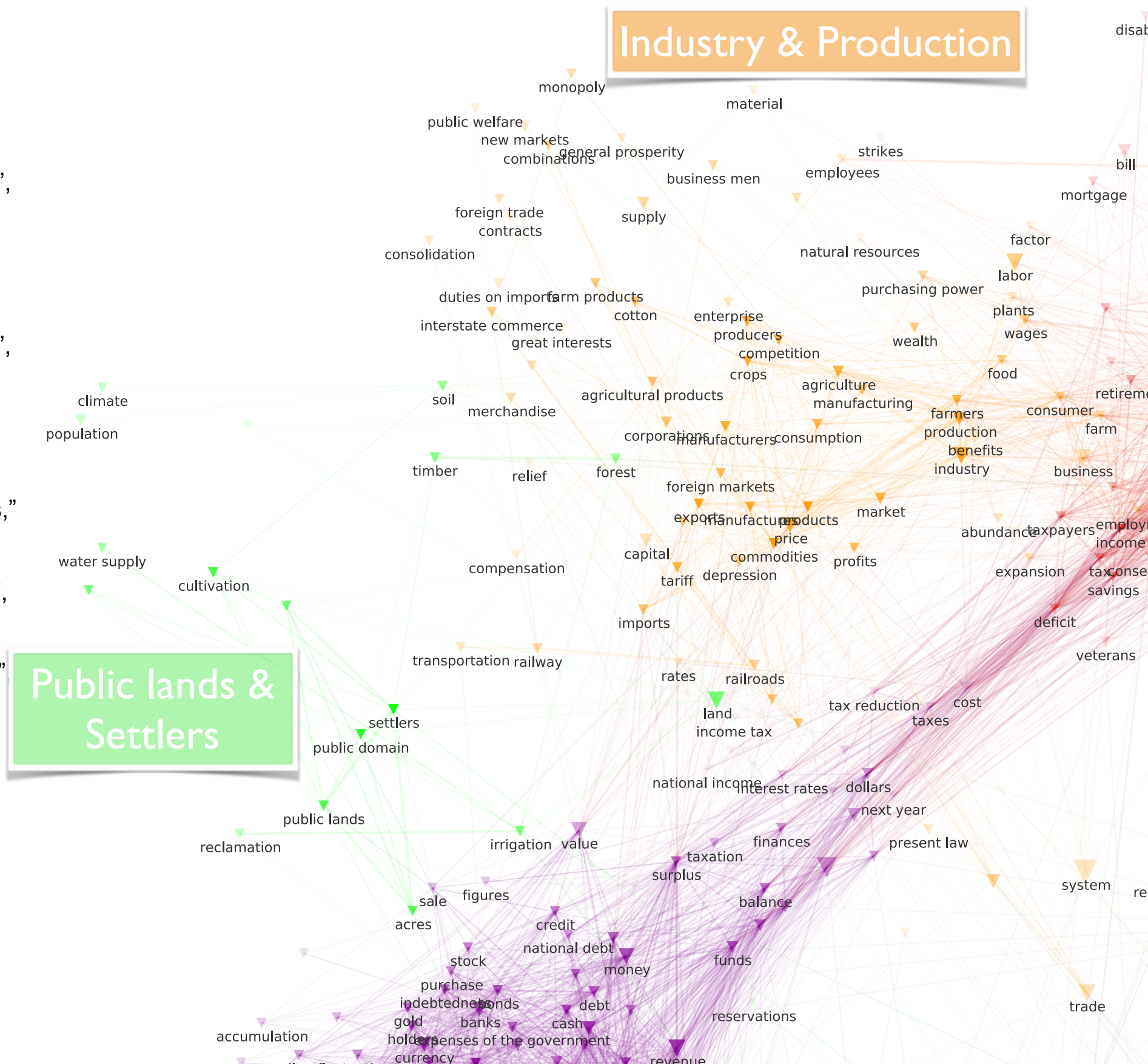
Naturalization & Citizenship

Naval Issues

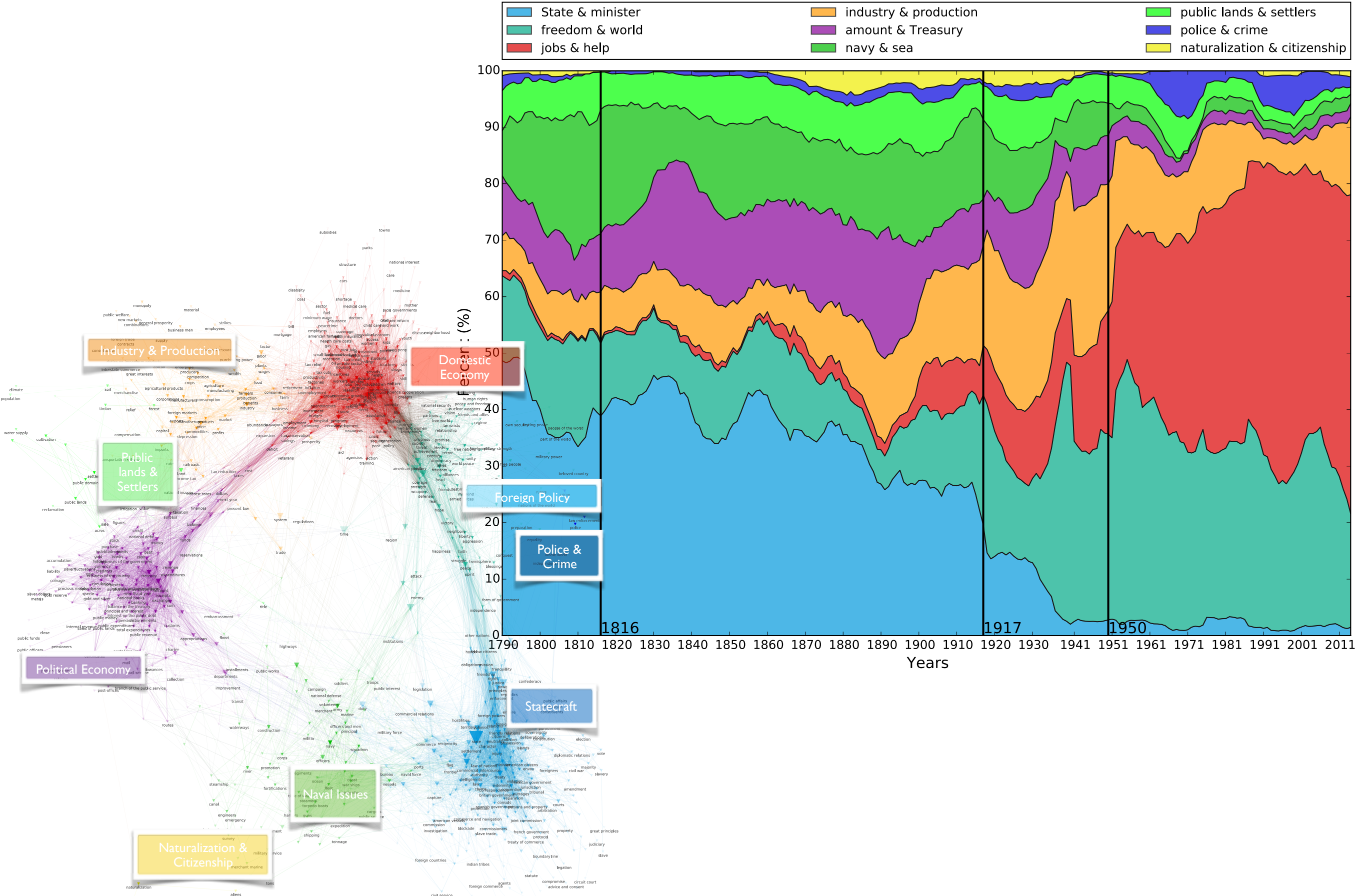
Semantic Map

9 Topics:

- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Public lands & Settlers** — “land”, “settlers” “acres”
- **Naturalization & Citizenship** — “naturalization”, “aliens”
- **Police & Crime** — “criminal”, “law enforcement”



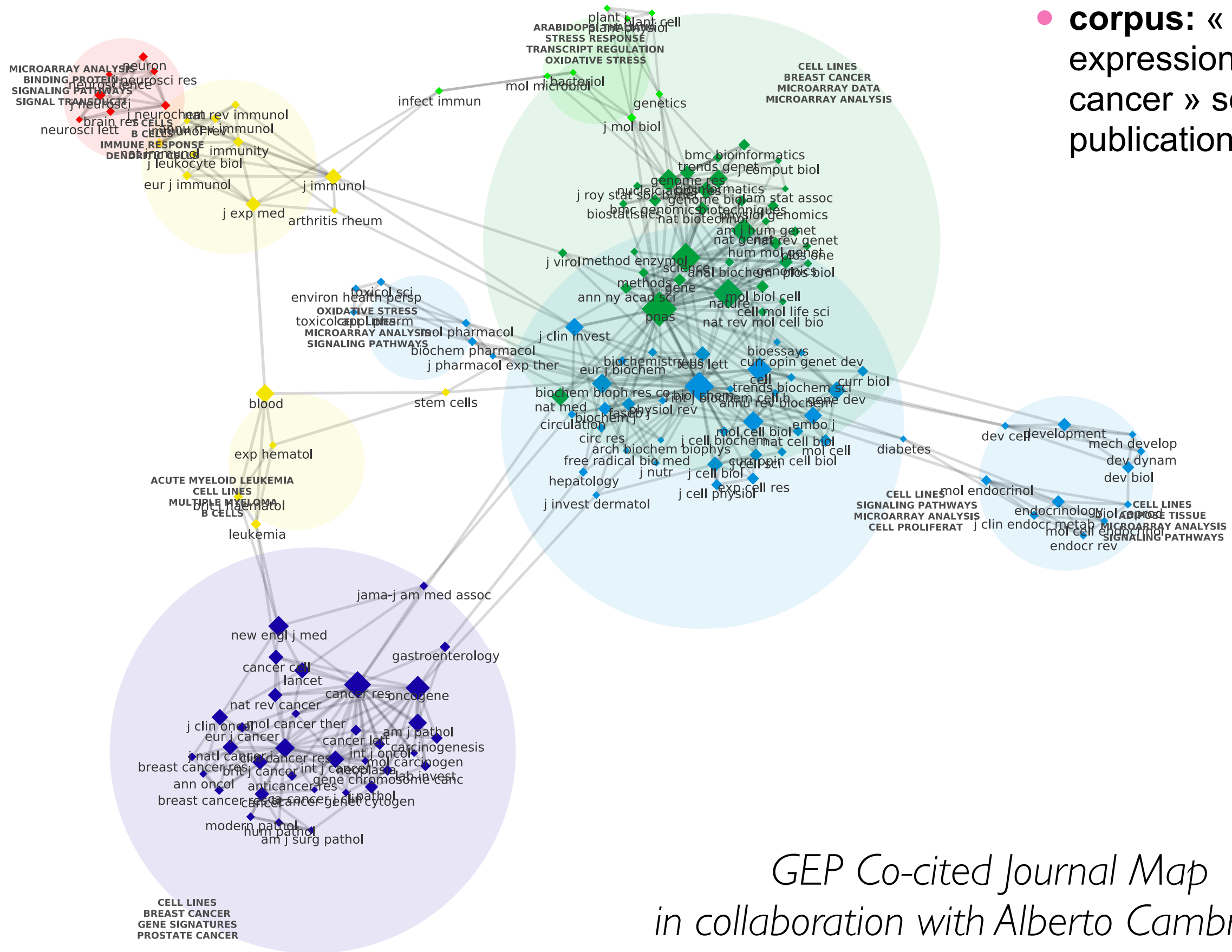
Categories in Time



From heterogeneous structure to dynamics

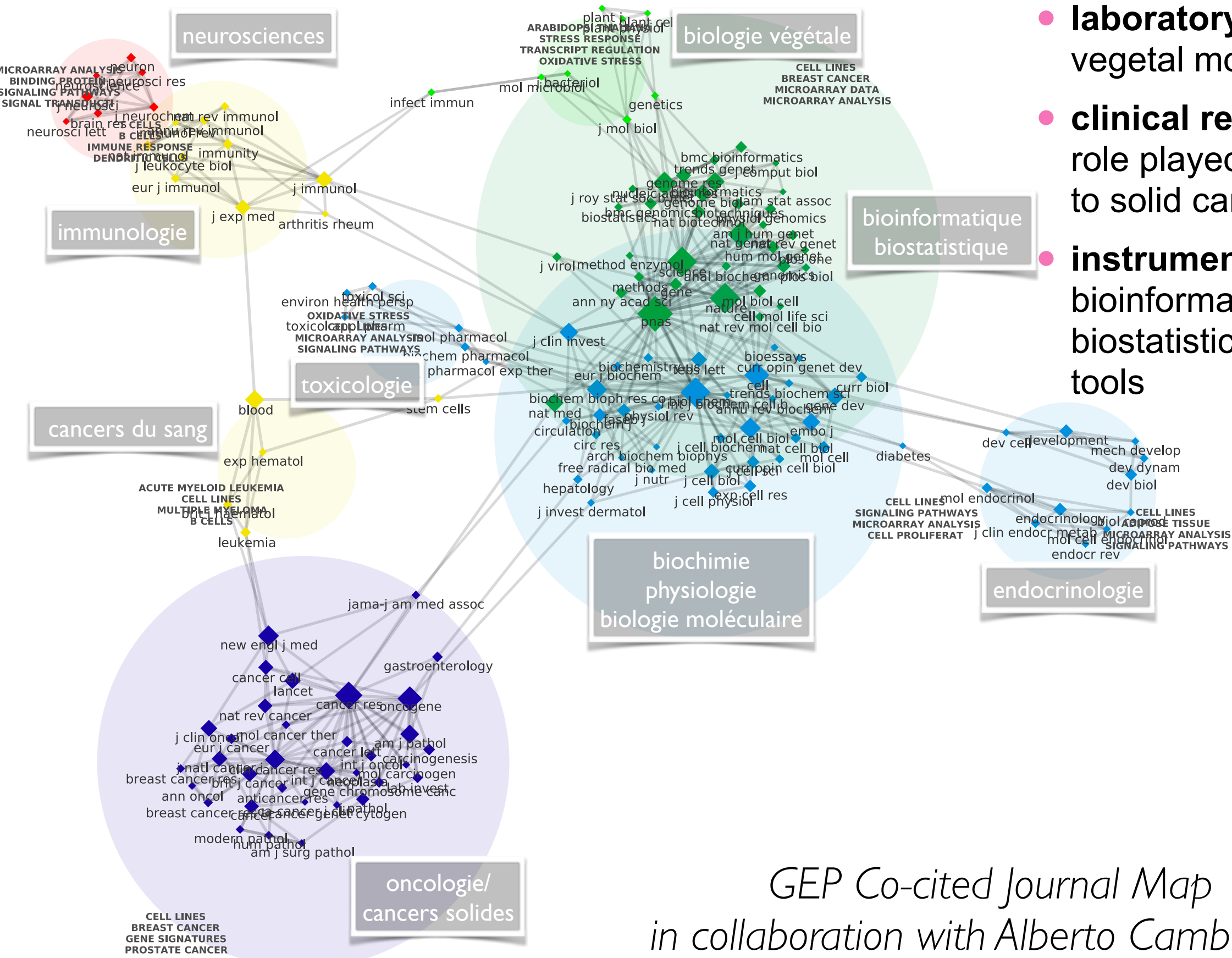
Translational Research Triangle

- **corpus:** « Gene expression profile & cancer » scientific publications



*GEP Co-cited Journal Map
in collaboration with Alberto Cambrosio*

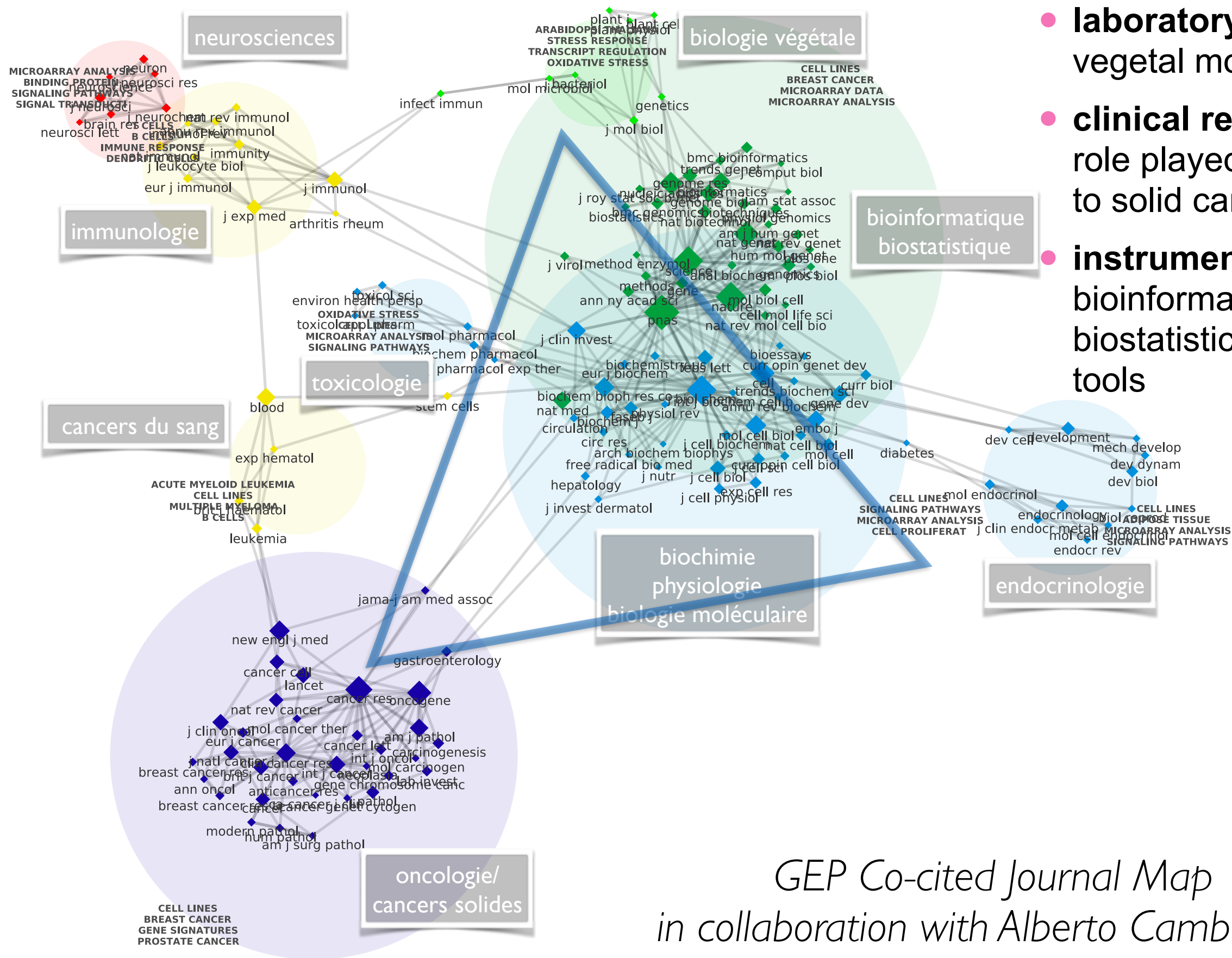
Translational Research Triangle



- **laboratory research** - vegetal model organism
- **clinical research** - central role played by application to solid cancer
- **instrumental research** - bioinformatics and biostatistics as necessary tools

*GEP Co-cited Journal Map
in collaboration with Alberto Cambrosio*

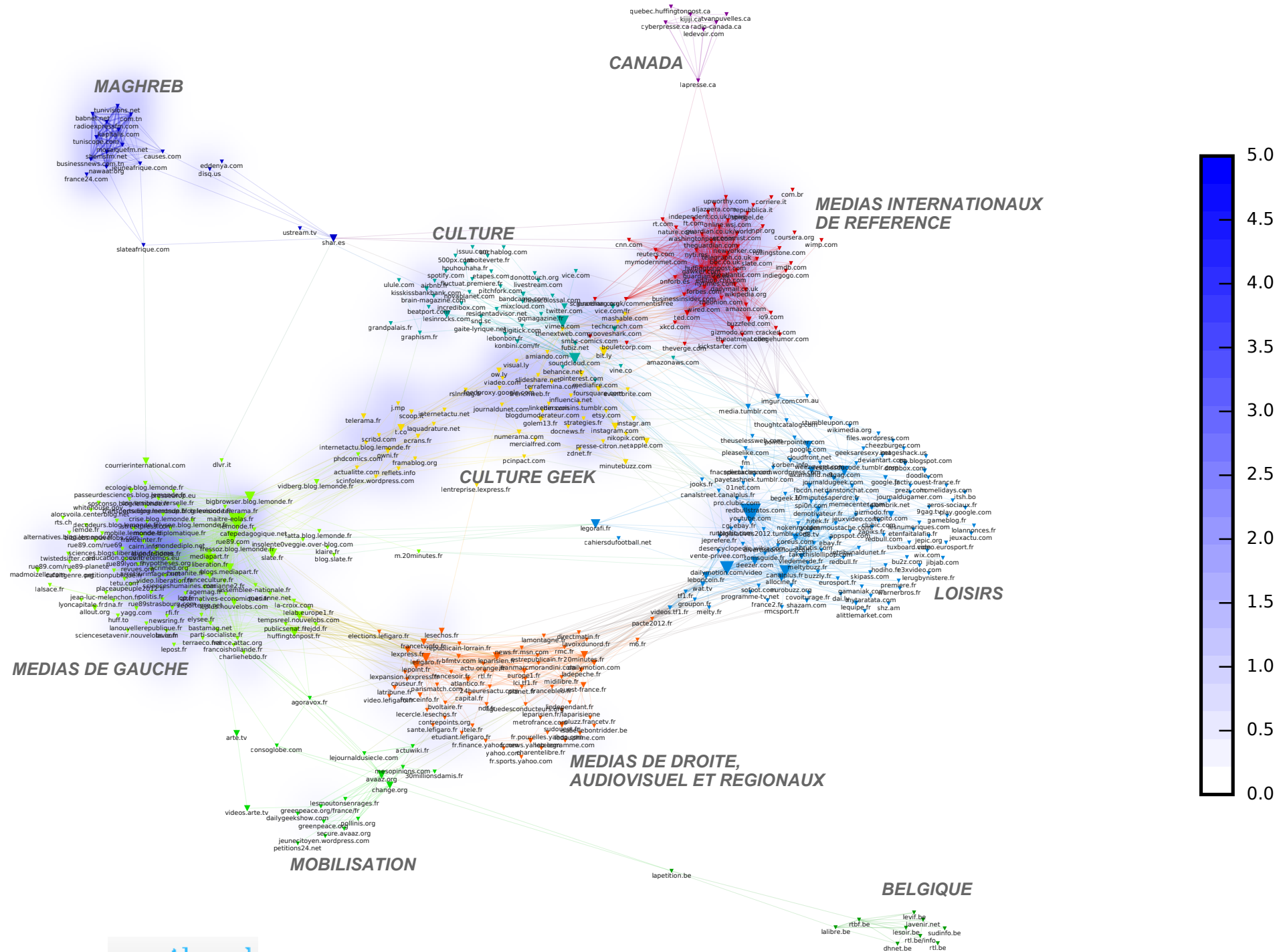
Translational Research Triangle



- **laboratory research** - vegetal model organism
- **clinical research** - central role played by application to solid cancer
- **instrumental research** - bioinformatics and biostatistics as necessary tools

*GEP Co-cited Journal Map
in collaboration with Alberto Cambrosio*

Mixing dimensions of analysis

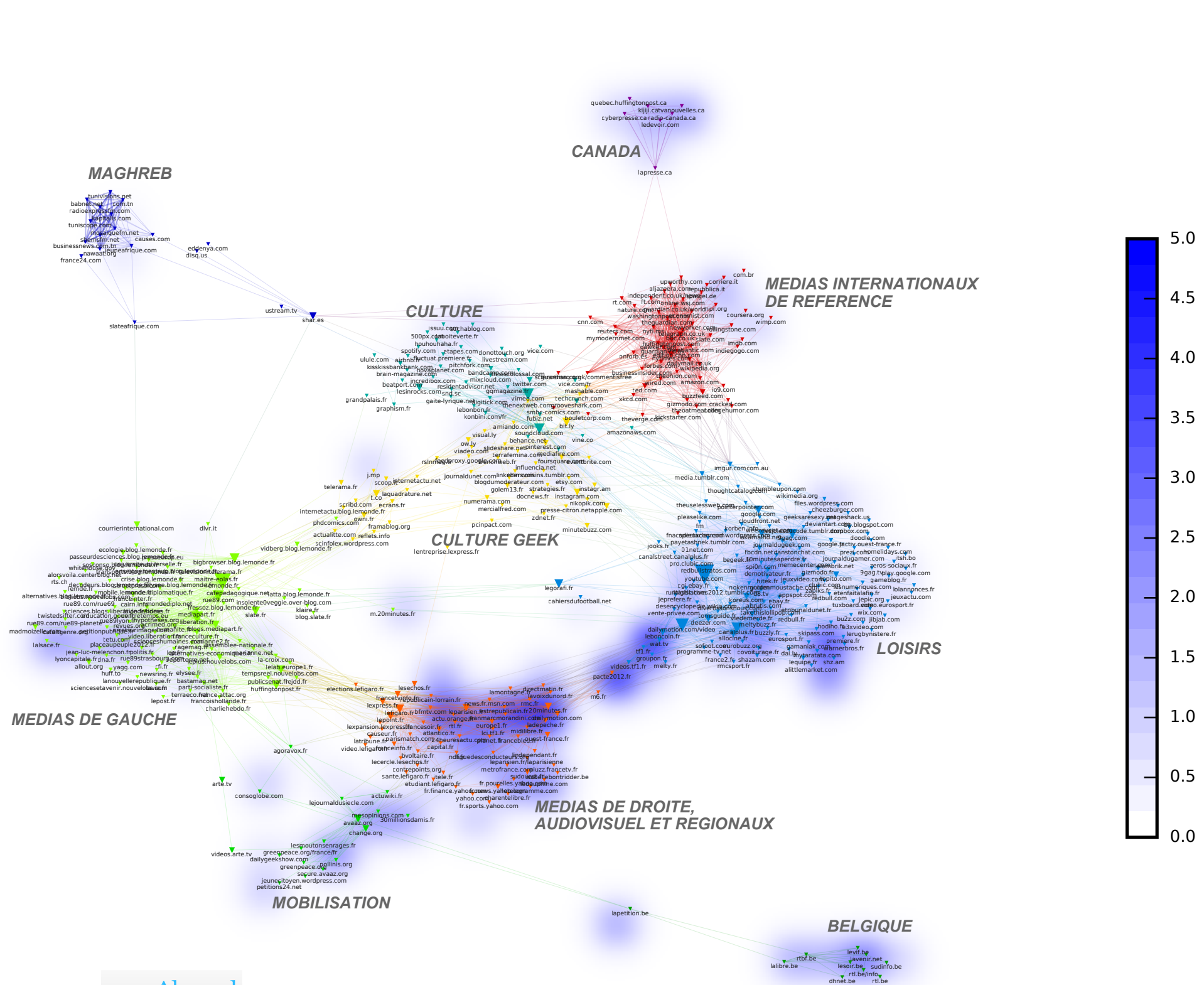


app. Algopol

CORTEXT

Facebook citation network

Mixing dimensions of analysis

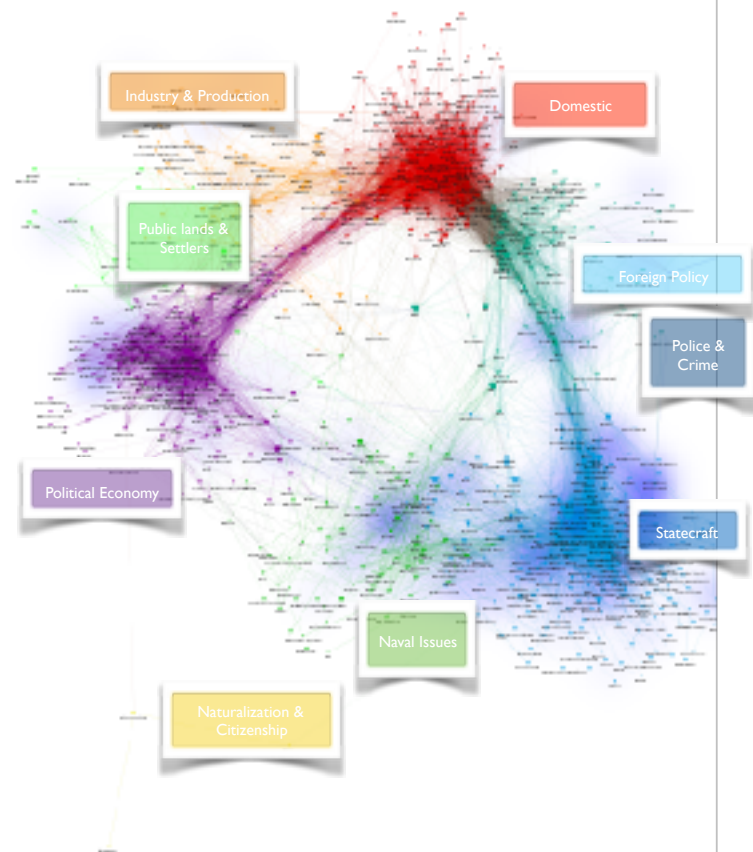


app. Algopol

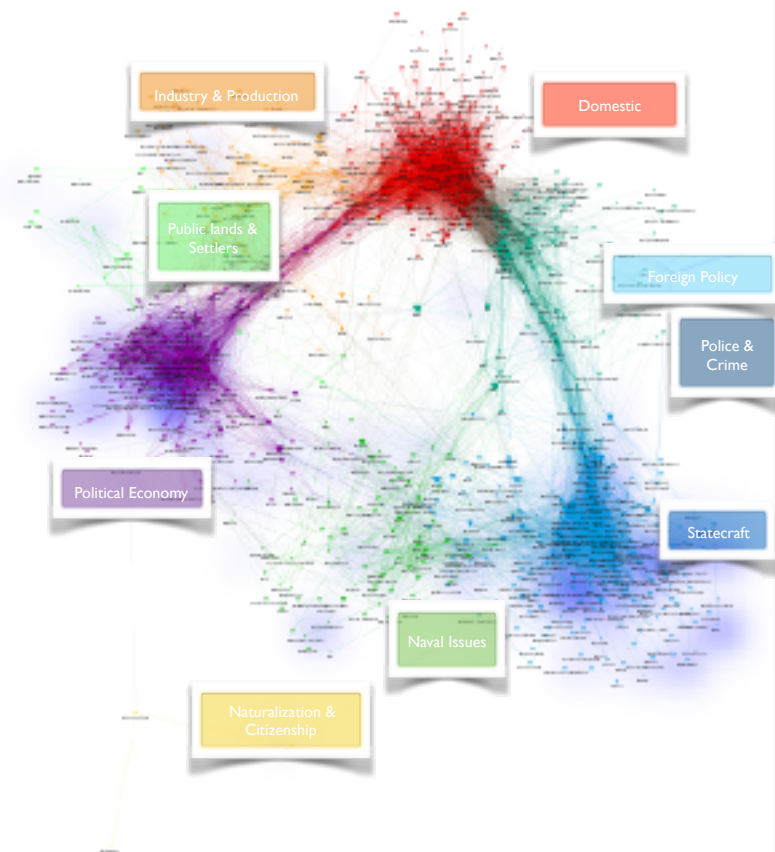
CORTEXT

Mapping individual presidents

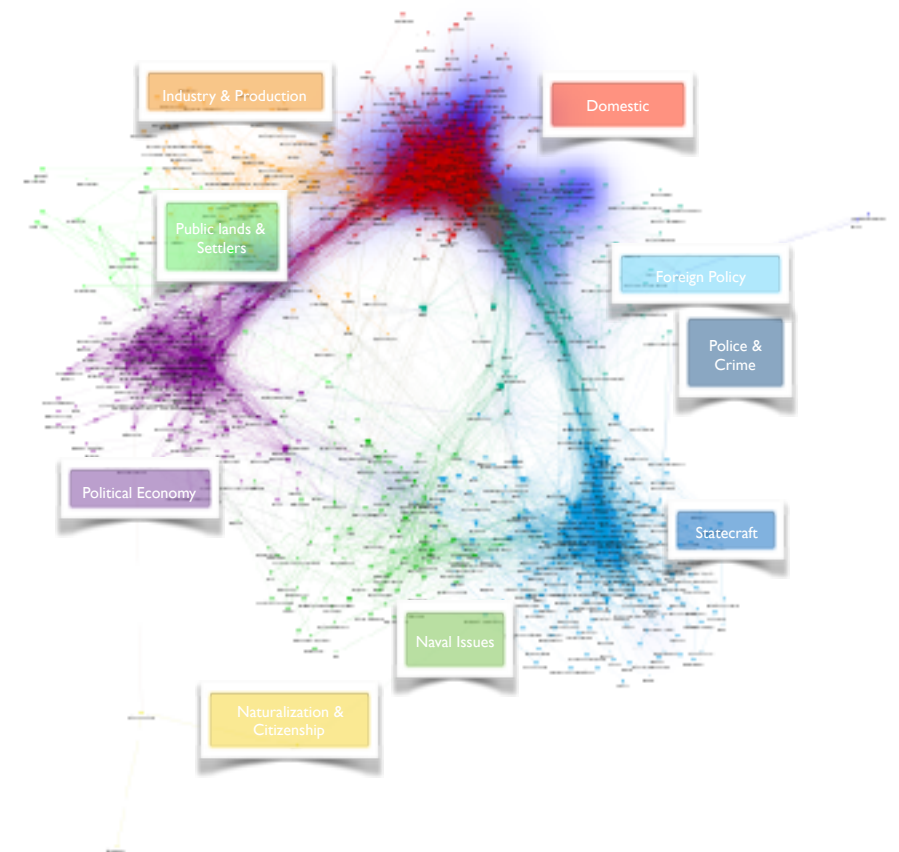
George Washington, 0-3



Abraham Lincoln, 0-3



Ronald Reagan, 0-3



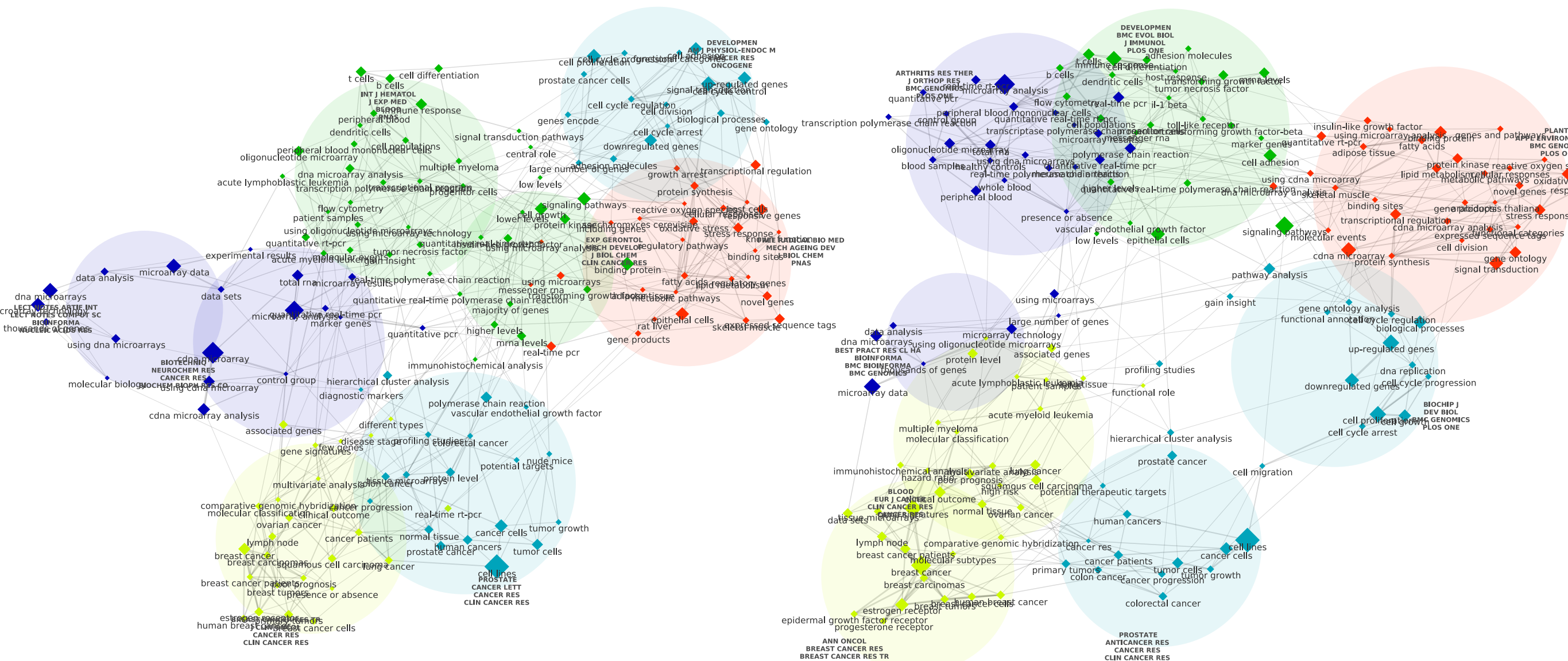
Comparing political discourses across time

Dynamical mapping

Inter-temporal matching between two maps

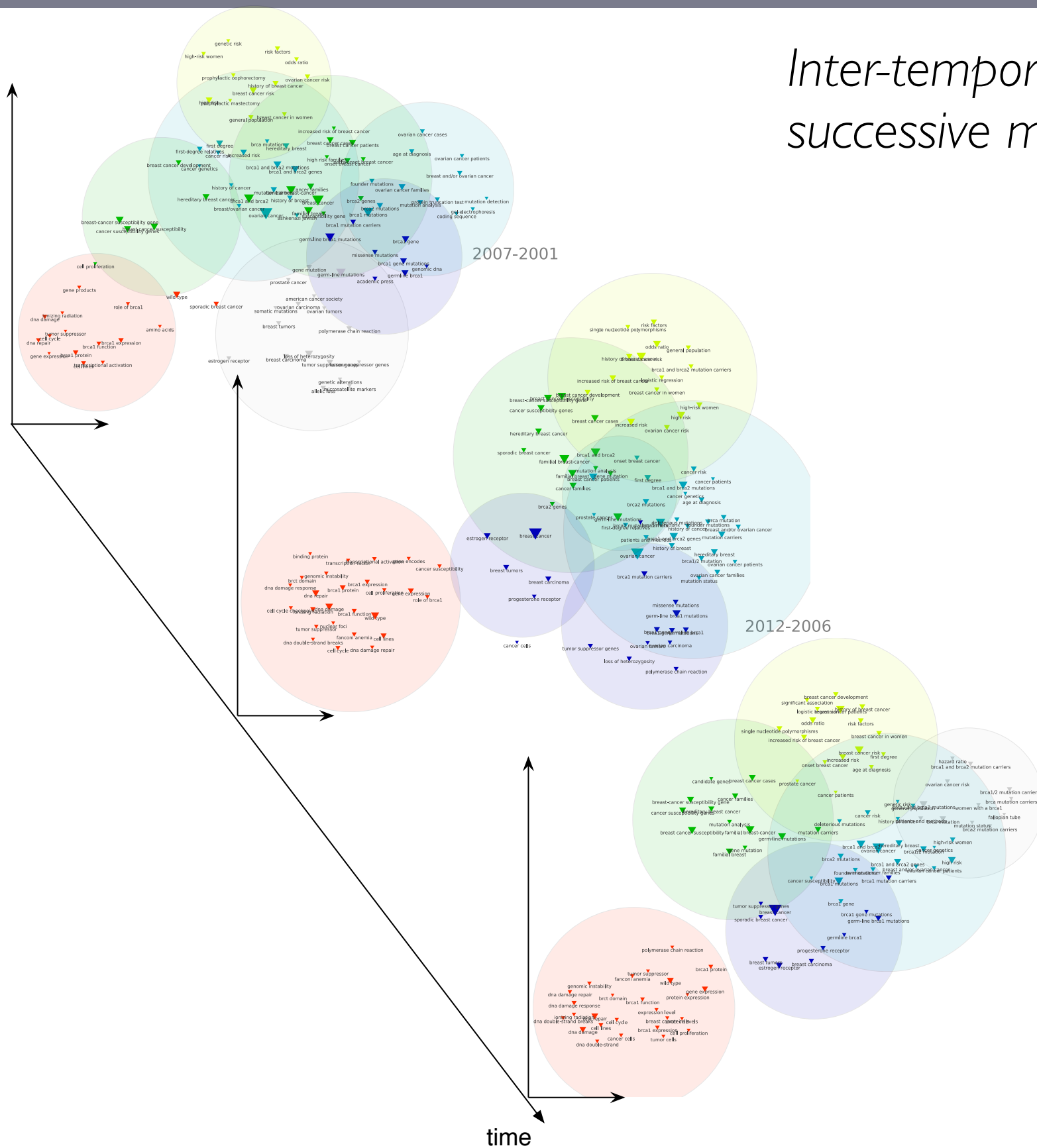
2001-2005

2006-2010



2002-1996

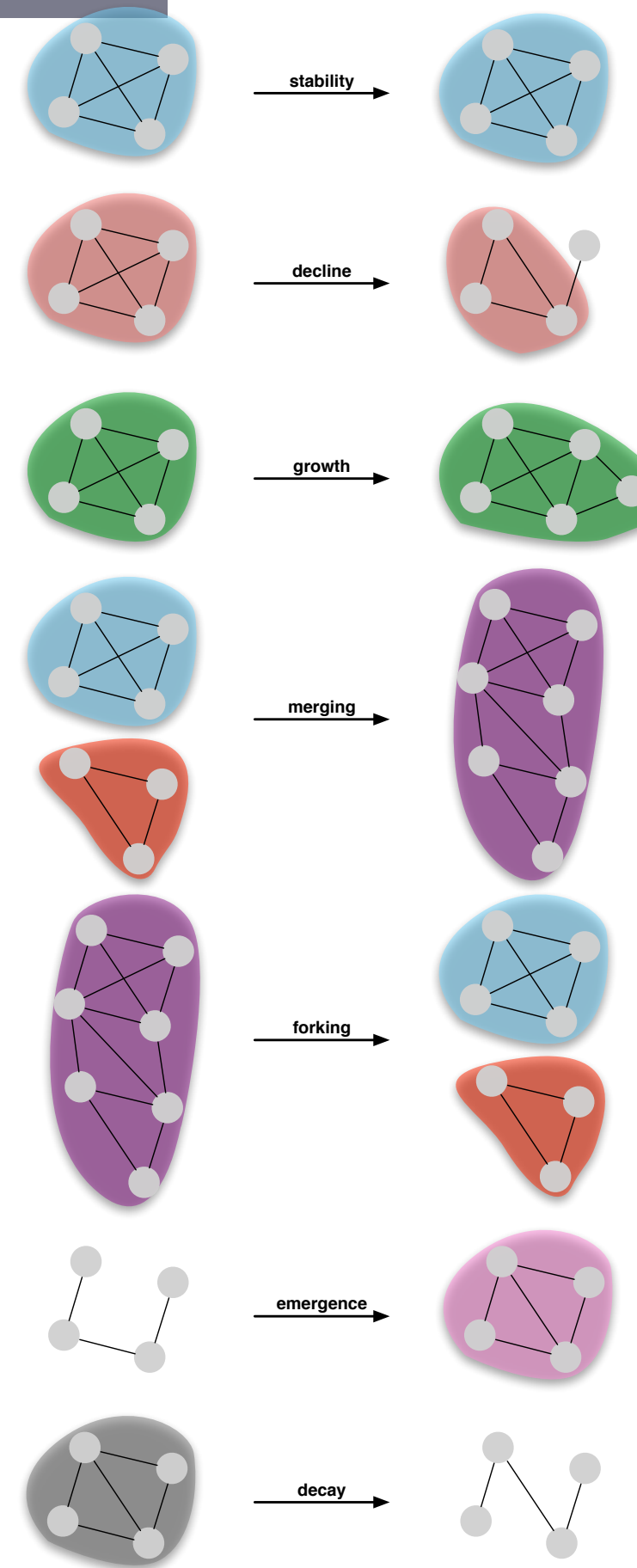
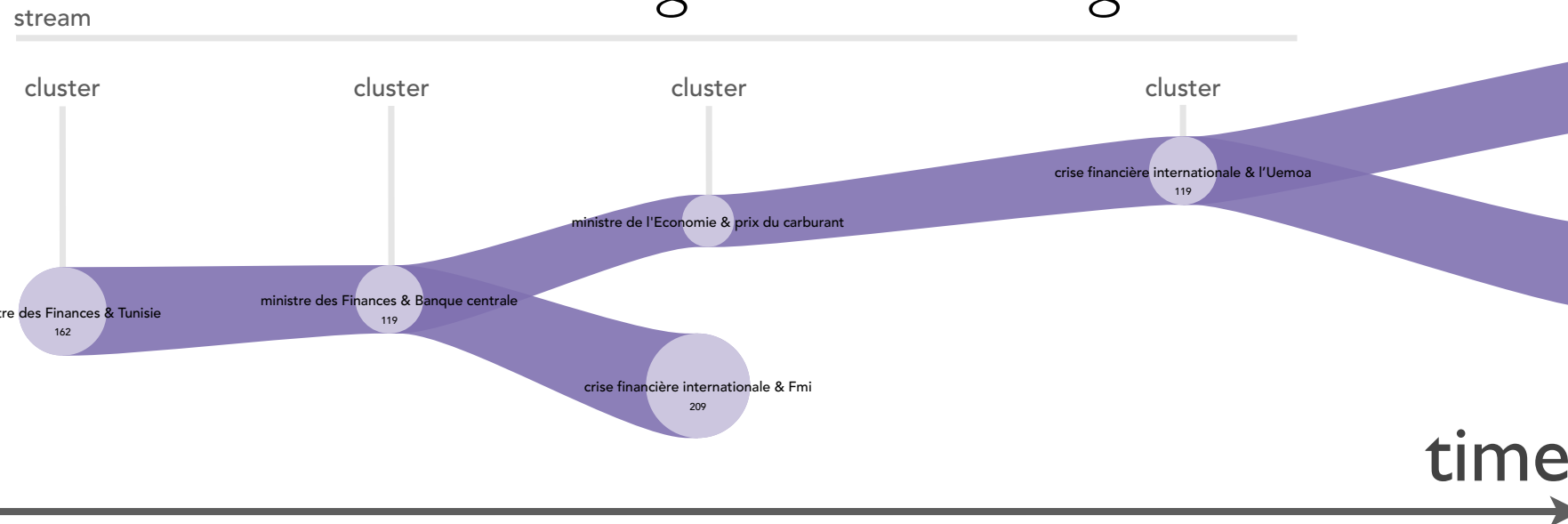
Inter-temporal matching between successive maps



Stream Reconstruction

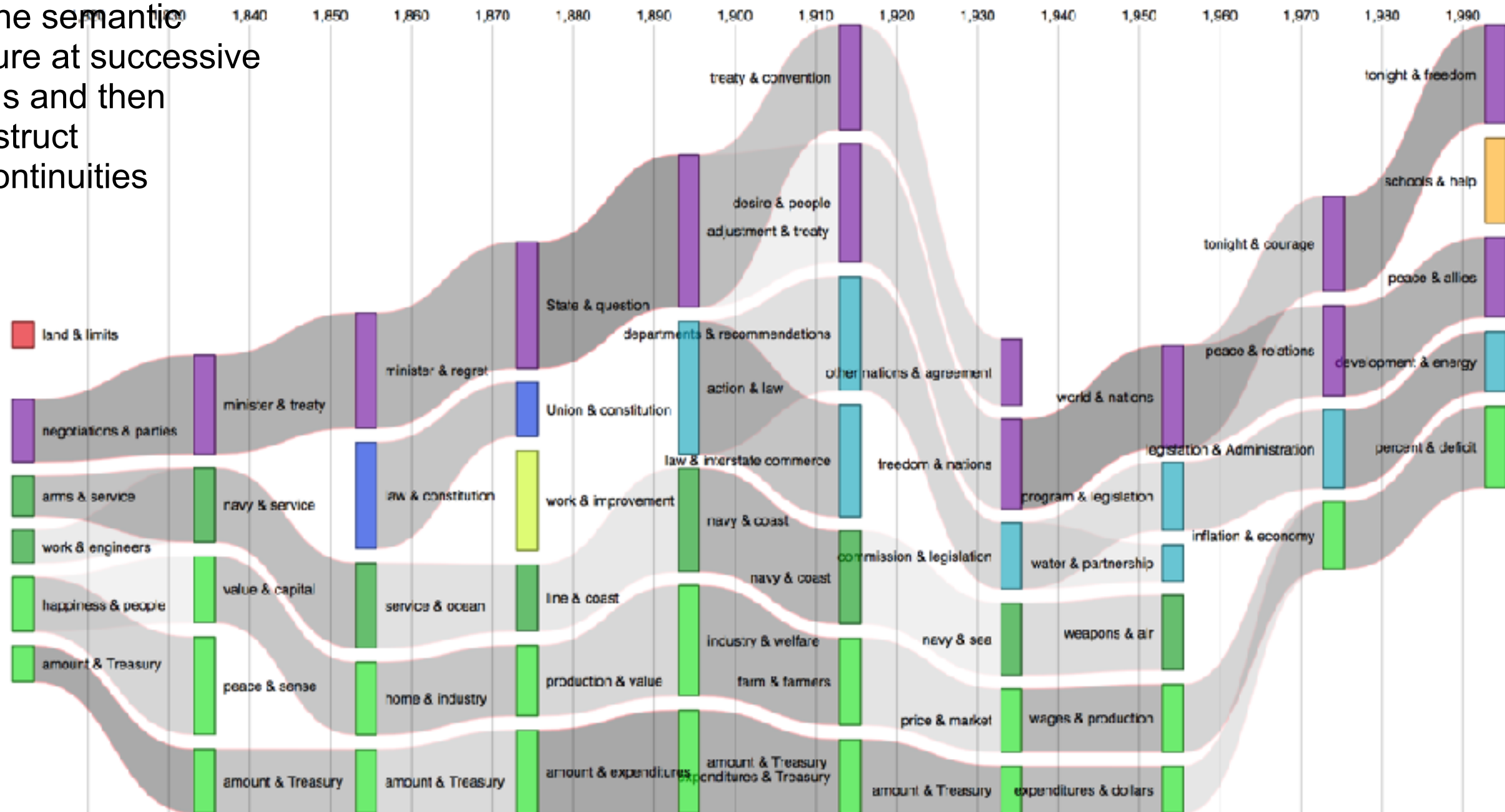
*Possible events
at mesoscopic
level*

*Clusters are intertemporally grouped
into streams according to their lineage*



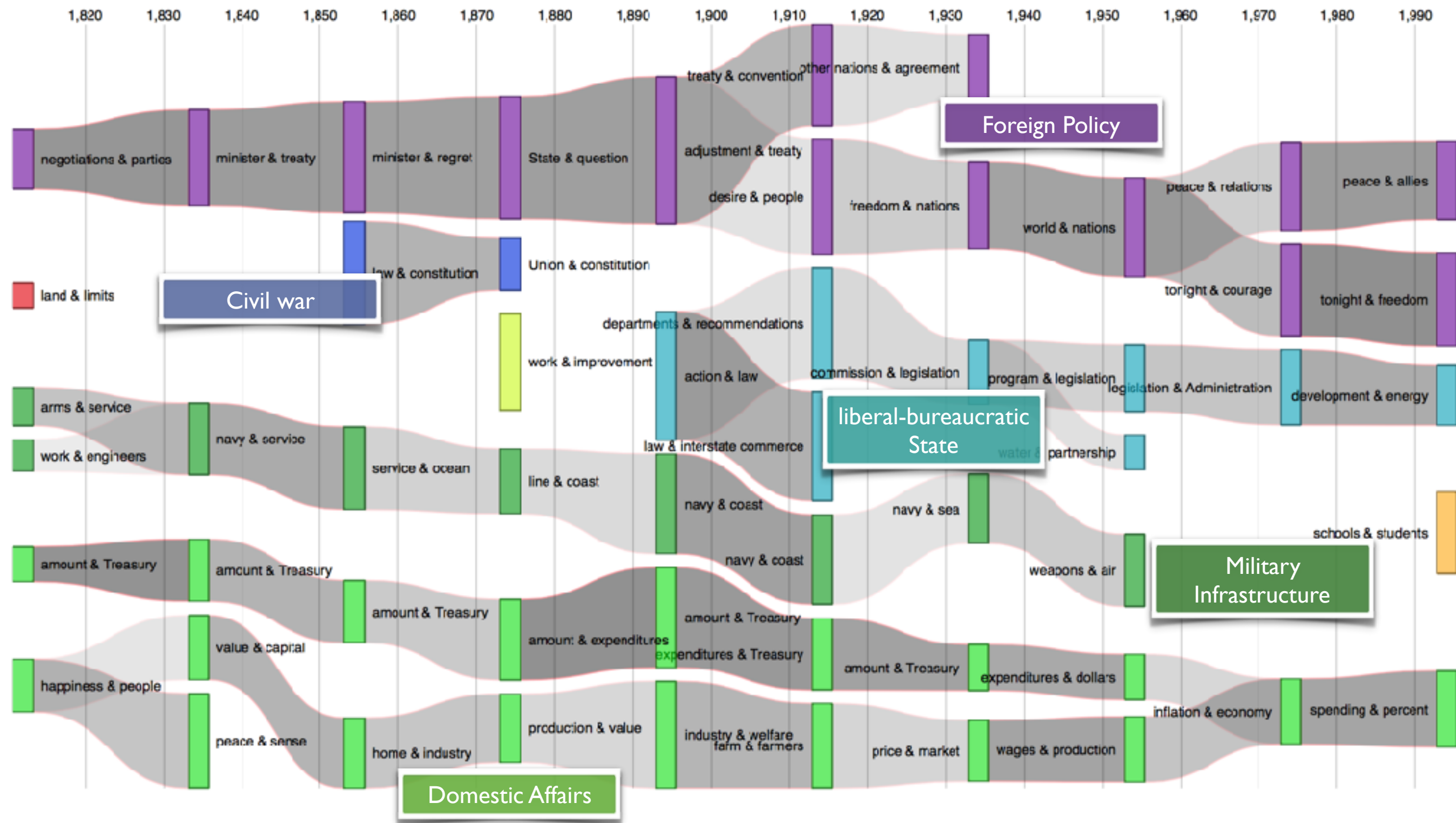
River Networks

- Capture topics with a delimited time span
- Conversely, river networks map the semantic structure at successive periods and then reconstruct (dis)continuities

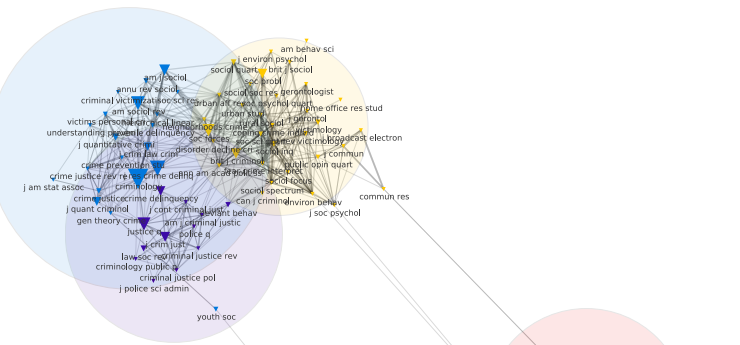


River Network

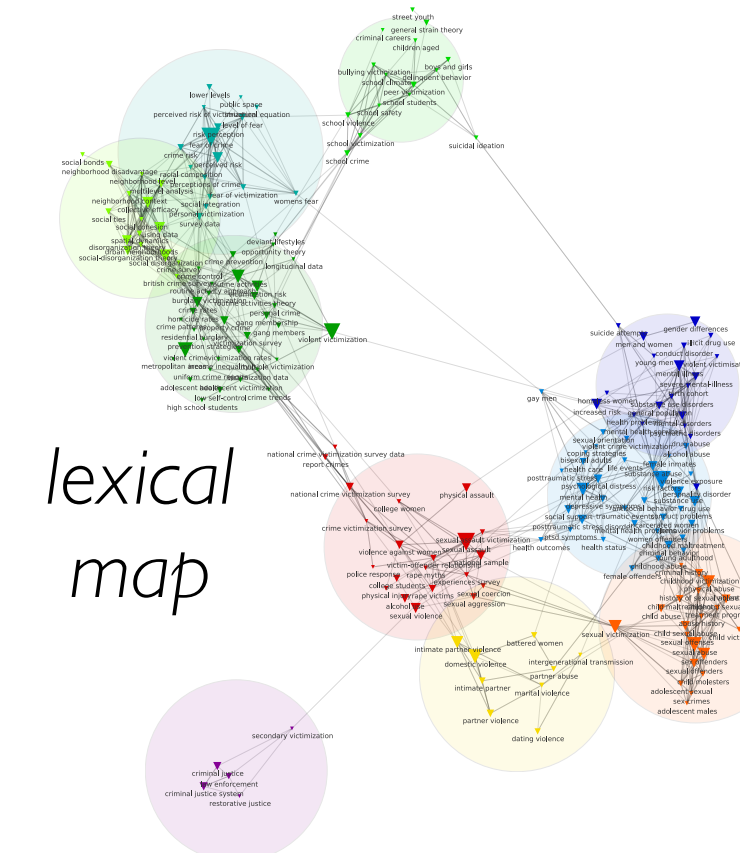
River Network



Mixing dimensions of analysis



1975-2013



victim-offender relationship & sexual assault

criminal justice system & law enforcement

social disorganization & disorganization theory

fear of crime & perceived risk

routine activities & property crime

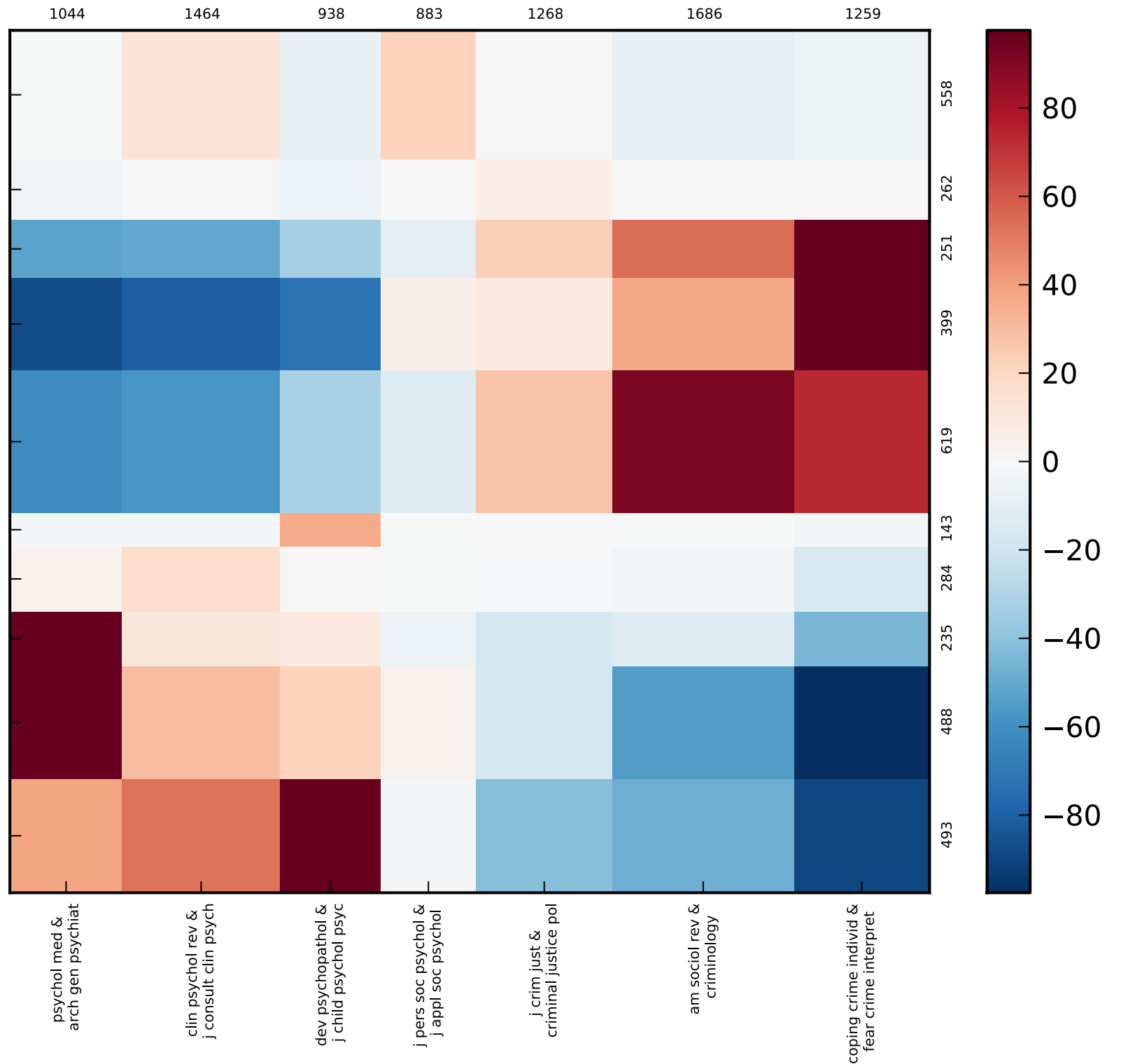
school students & peer victimization

intimate partner violence & domestic violence

mental illness & birth cohort

mental health & substance abuse

sexual abuse & sex offenders

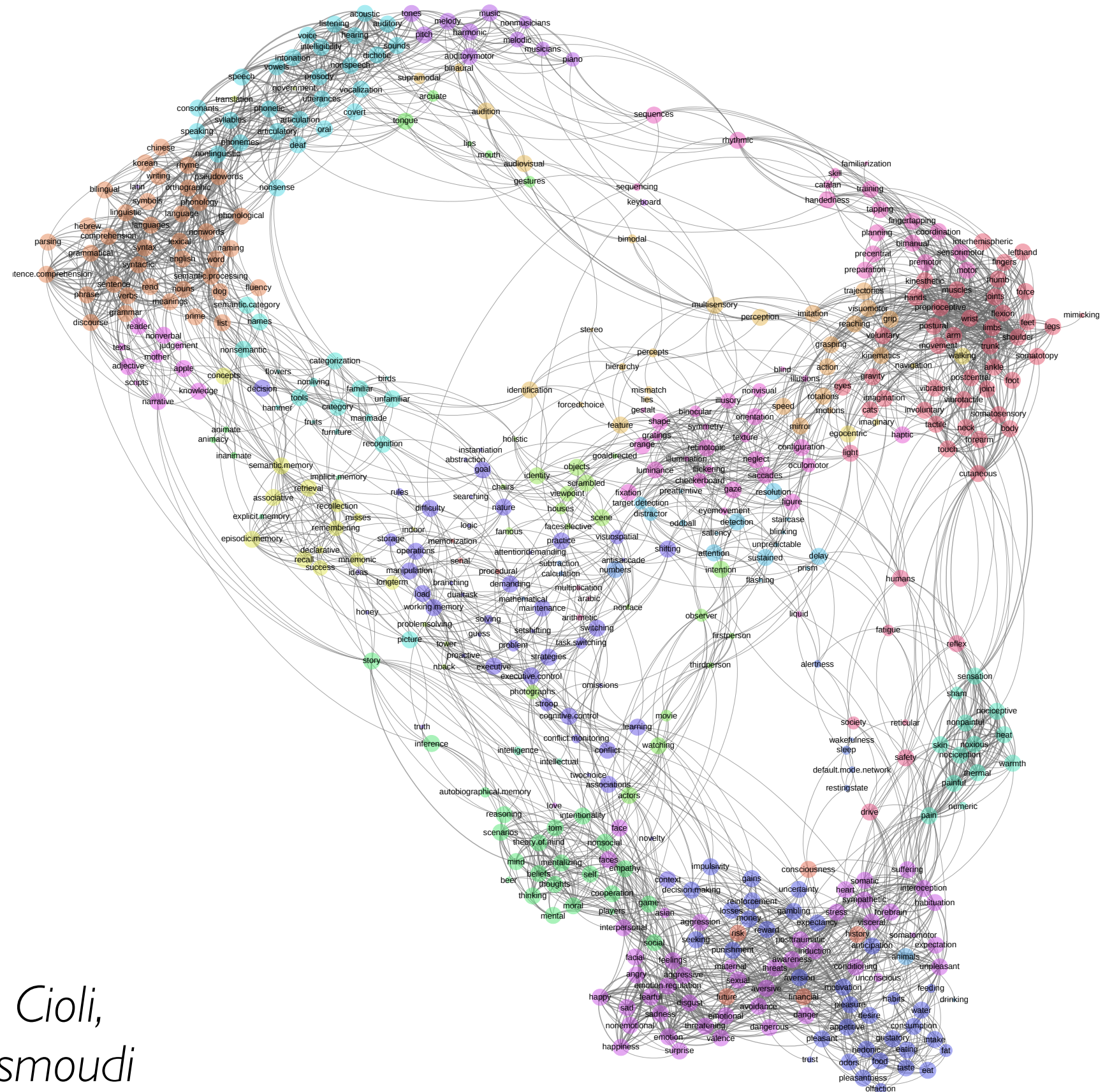


crime science corpus

Knowledge Discovery

Cognitive tasks Map

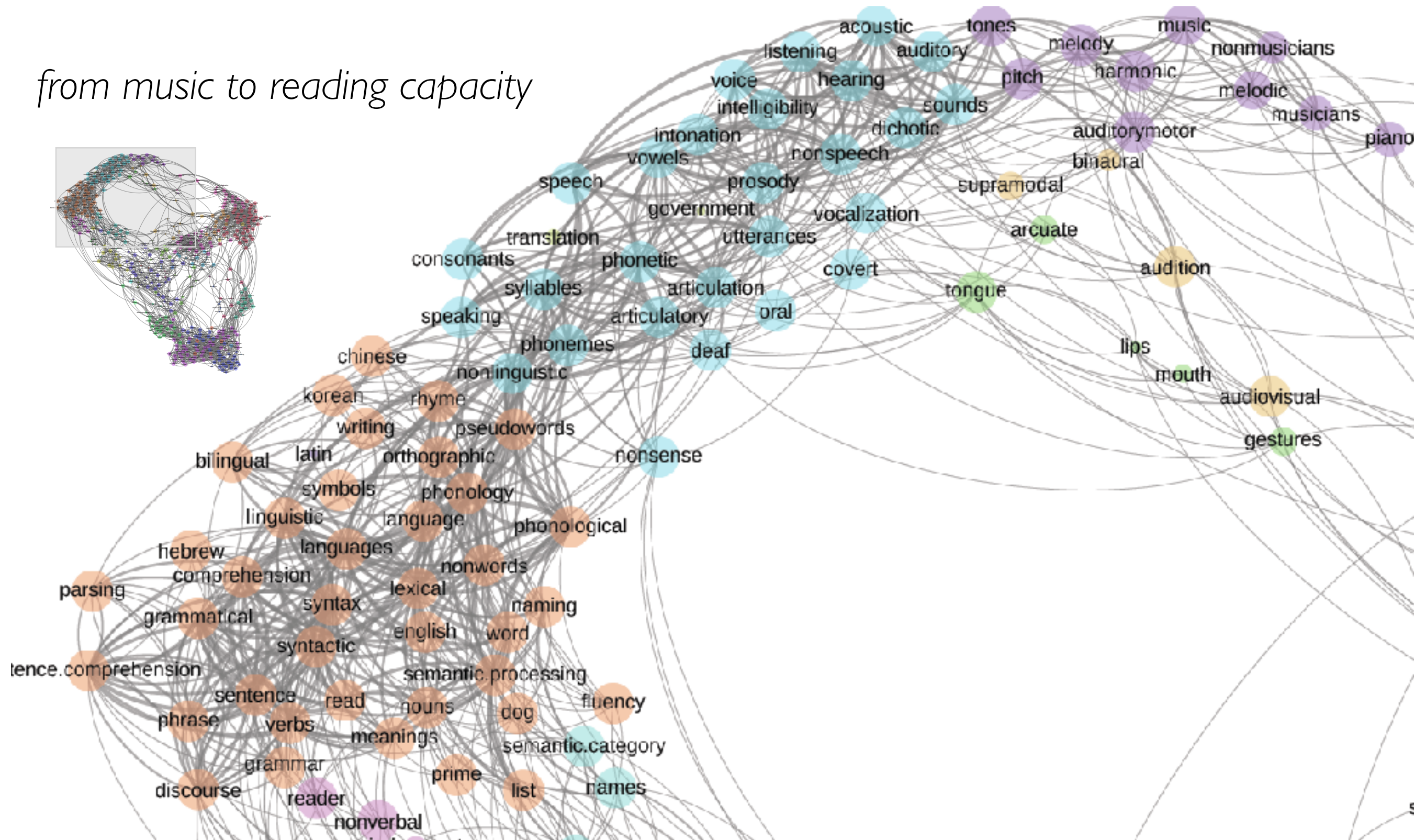
- ~500 cognitive tasks are mapped considering their joint appearance in 4,200 articles in neuro-imagery (fMRI)
- Main cognitive modules can be retrieved along with their relations



*in collaboration with Claudia Cioli,
Yves Burnod and Salma Mesmoudi*

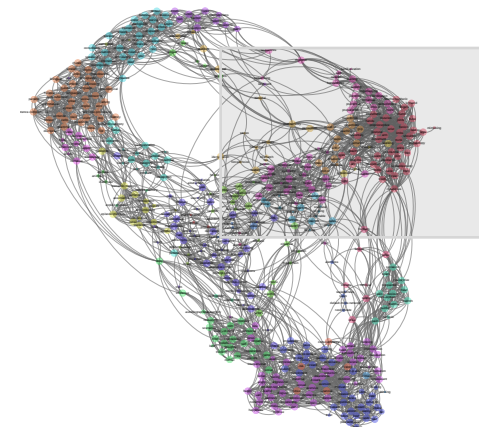
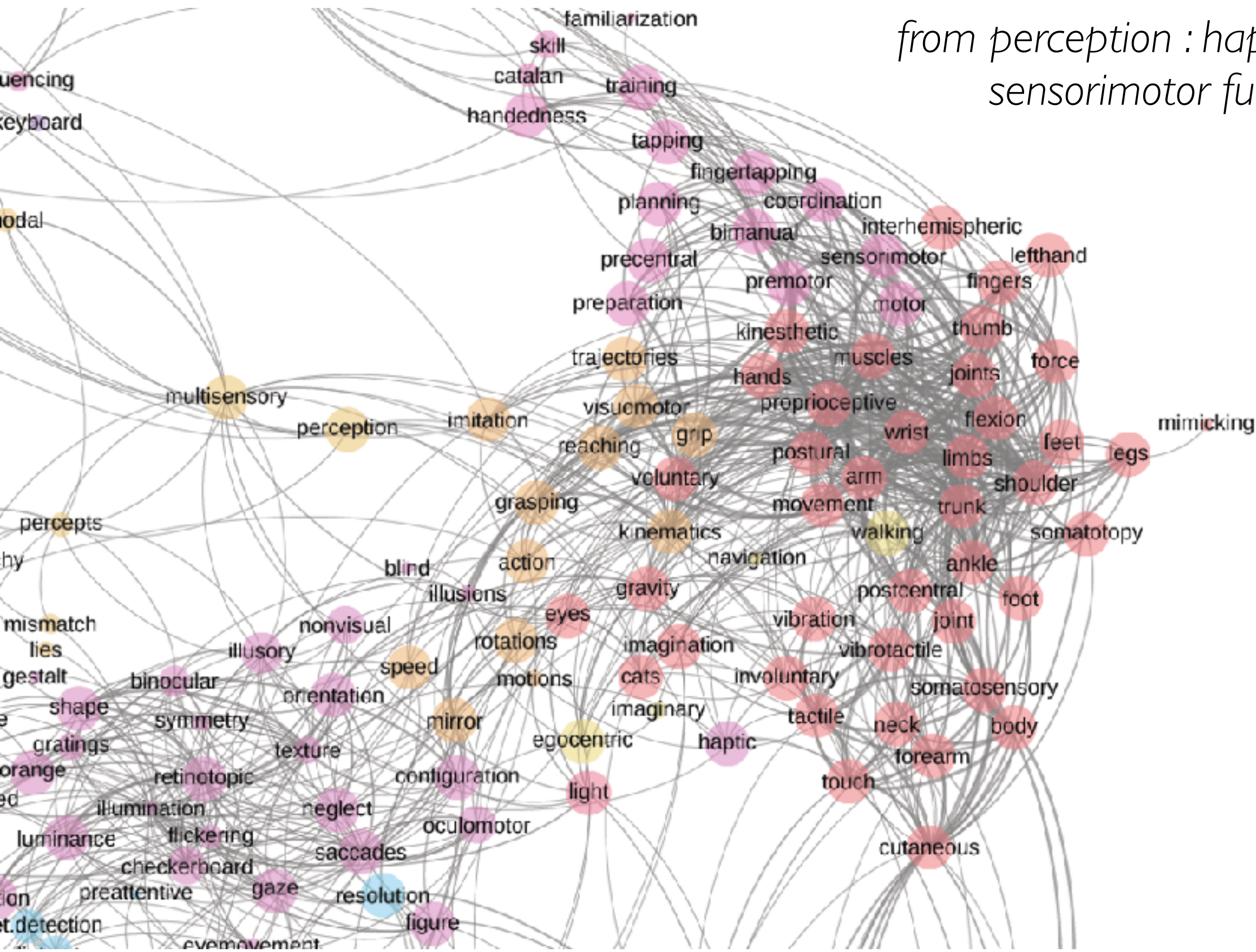
Cognitive tasks Map

from music to reading capacity

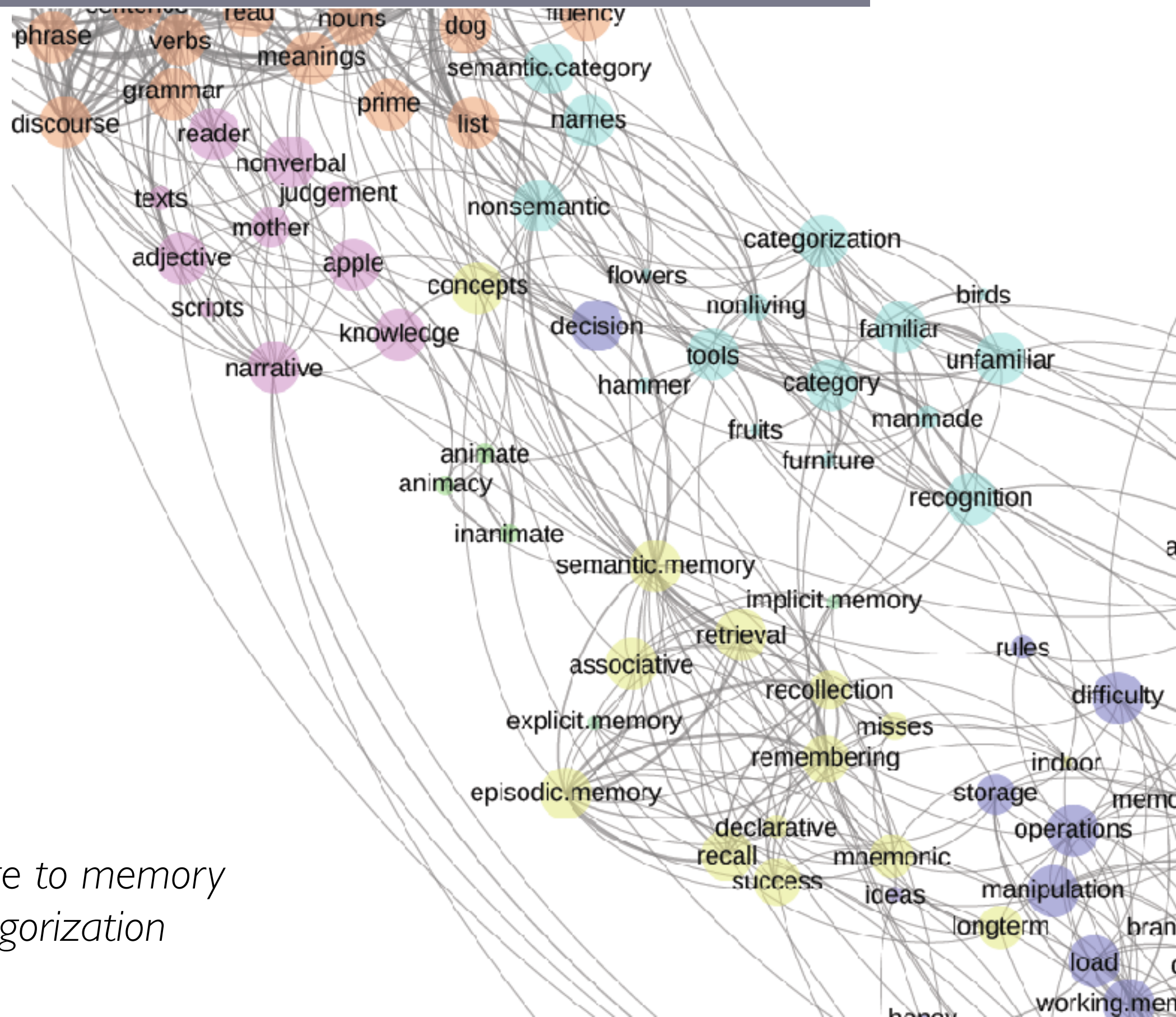
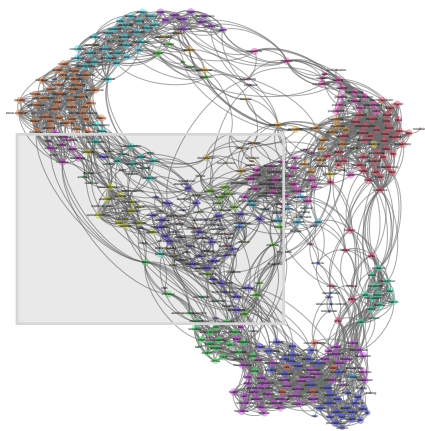


Cognitive tasks Map

from perception : haptic, visual to
sensorimotor functions

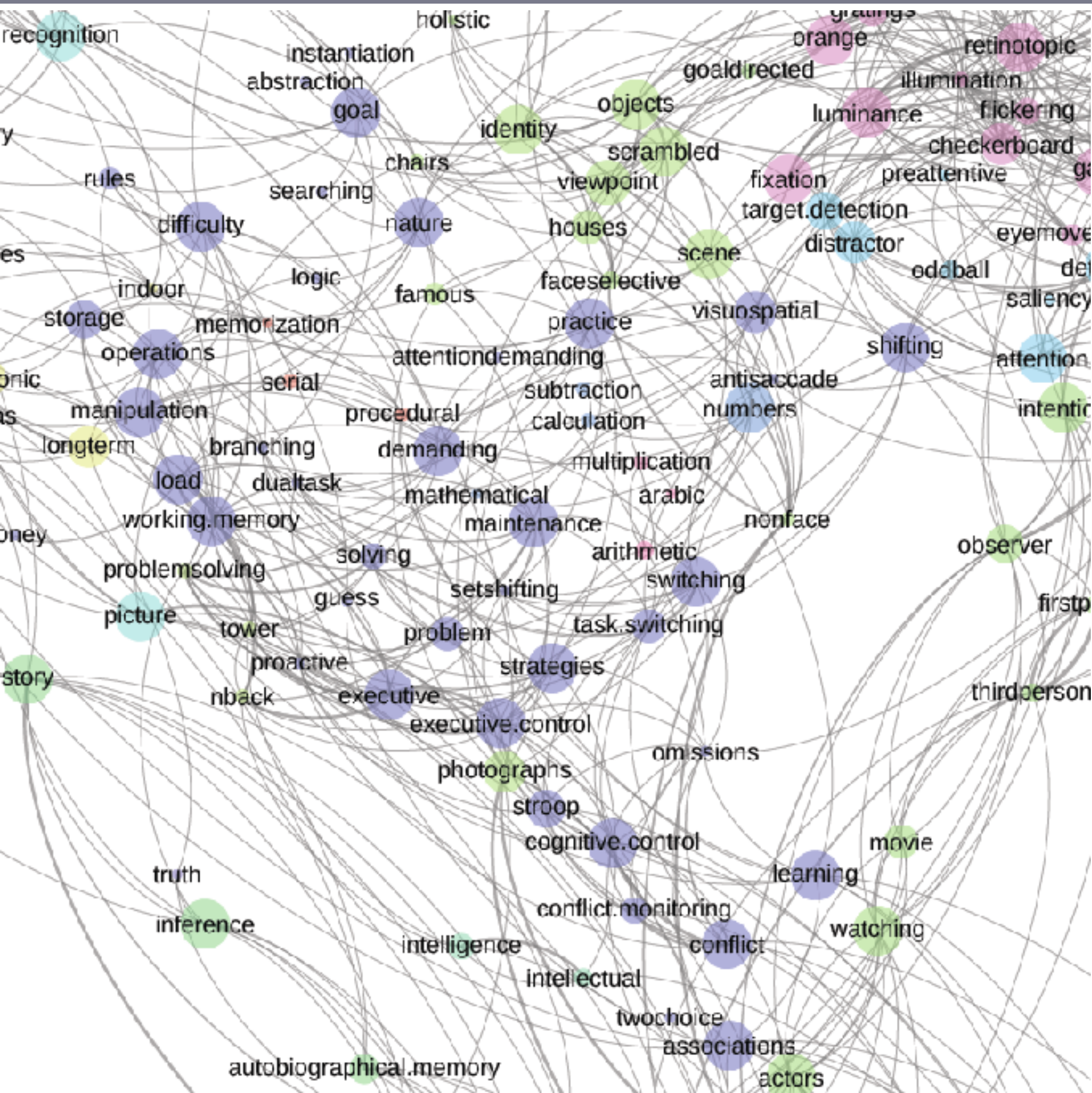


Cognitive tasks Map

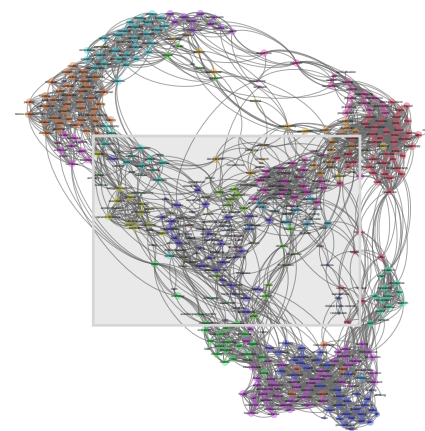


*from language to memory
and categorization*

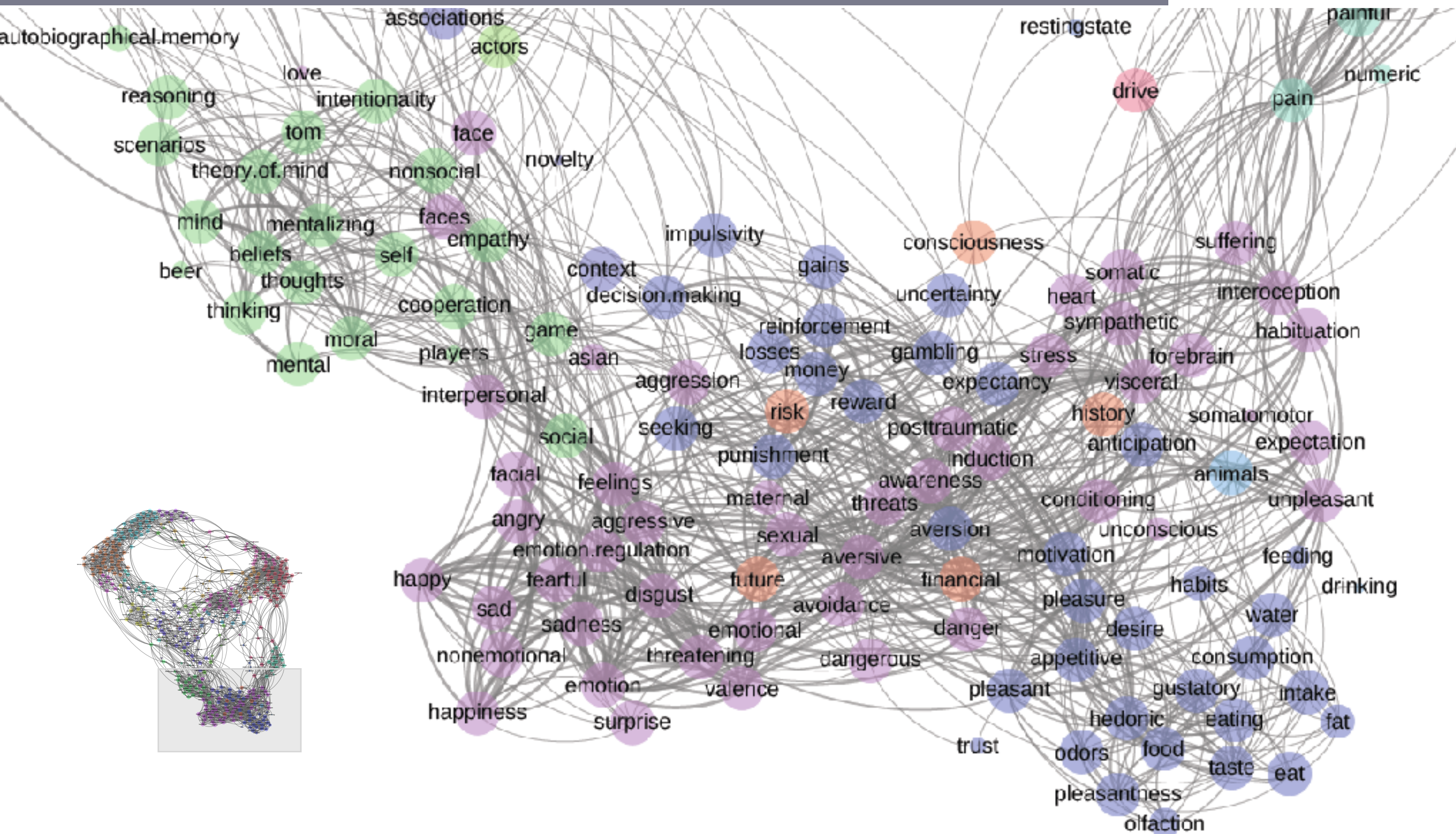
Cognitive tasks Map



*Problem solving and
planing connecting language,
and visual functions*

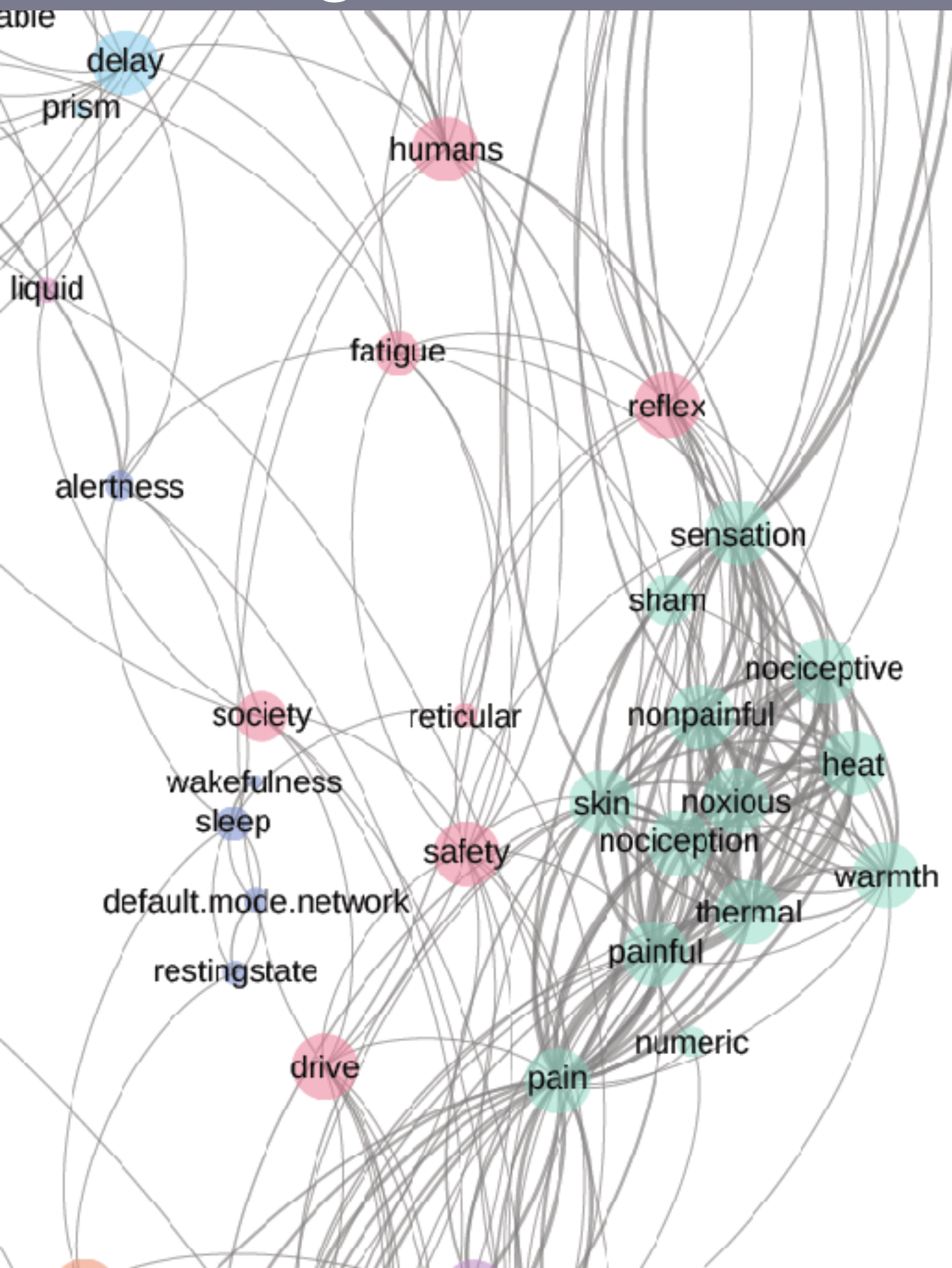


Cognitive tasks Map

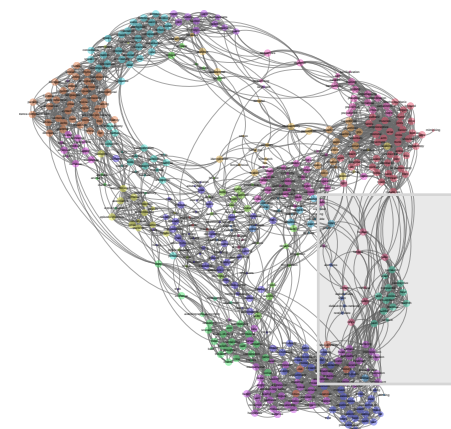


*Theory of Mind,
and emotions*

Cognitive tasks Map

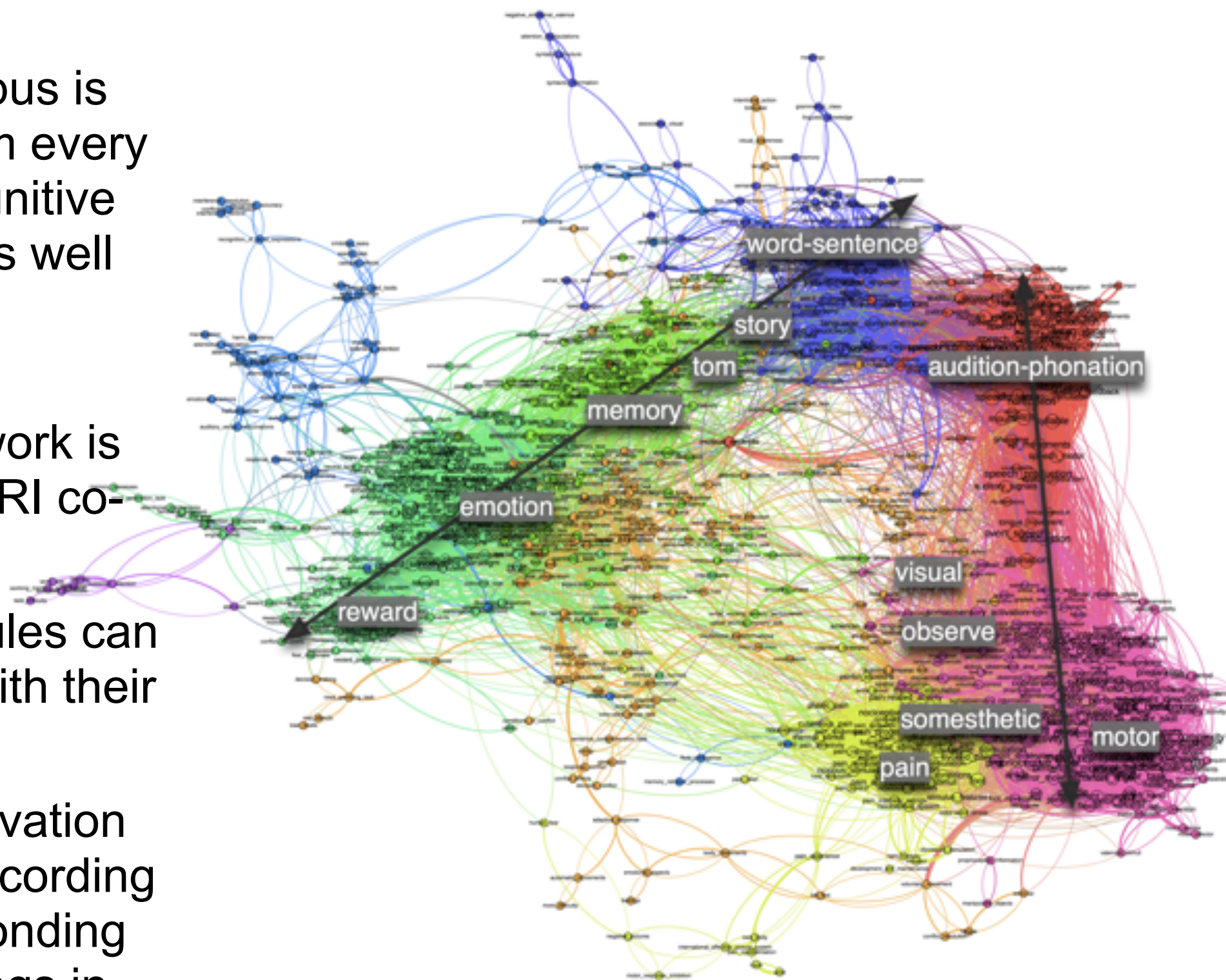


*Pain connecting emotions
to perceptions*



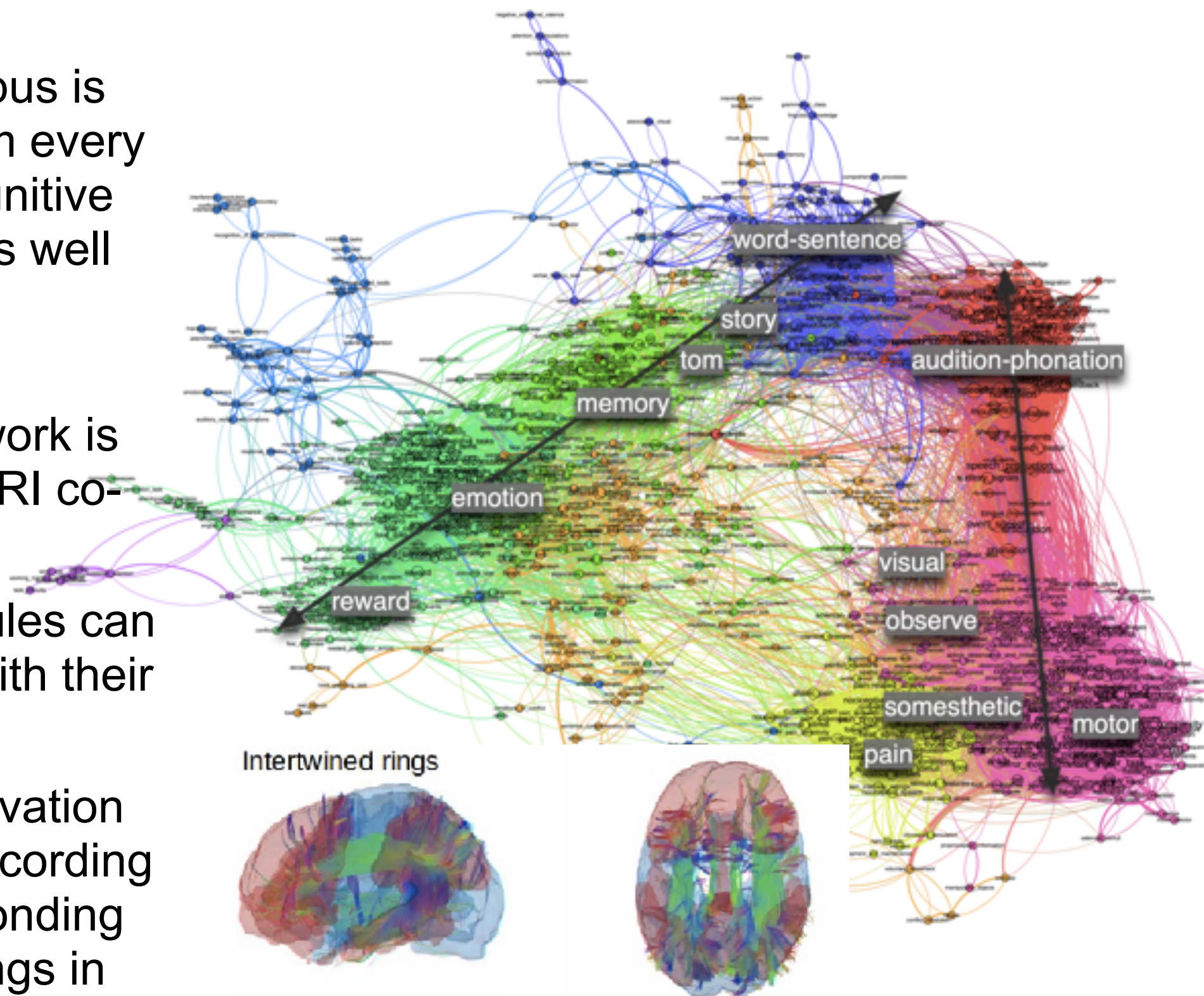
Cognitive tasks Map

- A neuroimager corpus is mined to extract from every fMRI articles the cognitive tasks being tested as well as the brain regions activated
- Cognitive tasks network is built according to fMRI co-activation pattern
- Main cognitive modules can be retrieved along with their relations
- The resulting co-activation map is structured according to two axes corresponding to two intertwined rings in the cortex



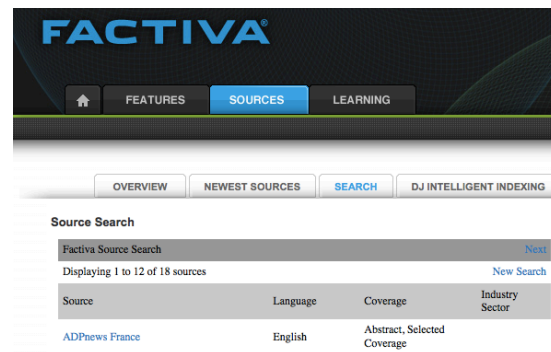
Cognitive tasks Map

- A neuroimager corpus is mined to extract from every fMRI articles the cognitive tasks being tested as well as the brain regions activated
- Cognitive tasks network is built according to fMRI co-activation pattern
- Main cognitive modules can be retrieved along with their relations
- The resulting co-activation map is structured according to two axes corresponding to two intertwined rings in the cortex



Case Study on Food Security issue media framing

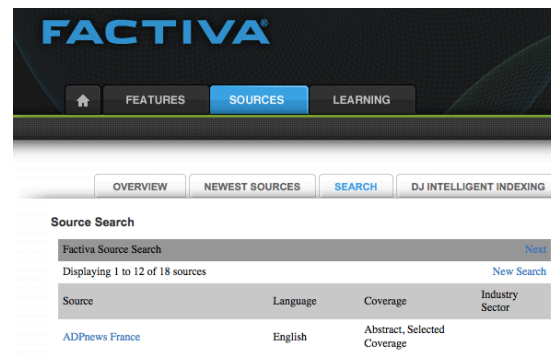
Case Study: food security



query=
"sécurité alimentaire"

~22,000 articles in
french speaking
journals over
7 years

Case Study: food security



query=
"sécurité alimentaire"

~22,000 articles in
french speaking
journals over
7 years

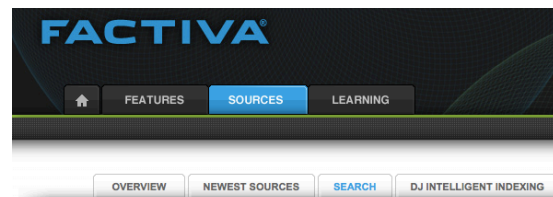
- morphosyntactic analysis
- chunking of grammatical patterns
- statistical filters

NLP tools

- named entities extraction for
countries

Terms list

Frames dynamics: food security



Source Search

Factiva Source Search [New Search](#)

Displaying 1 to 12 of 18 sources

Source	Language	Coverage	Industry Sector
ADPnews France	English	Abstract, Selected Coverage	

query=
"sécurité alimentaire"

~22,000 articles in
french speaking
journals over
7 years

- morphosyntactic analysis
- chunking of grammatical patterns
- statistical filters

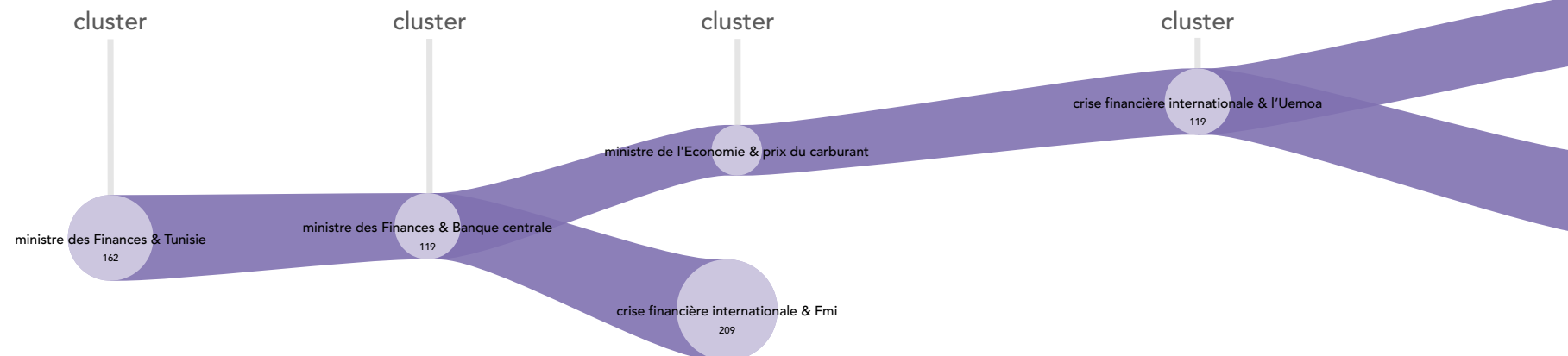
NLP tools

- named entities extraction for
countries

Terms list

indexation of extracted terms
and semantic network computation

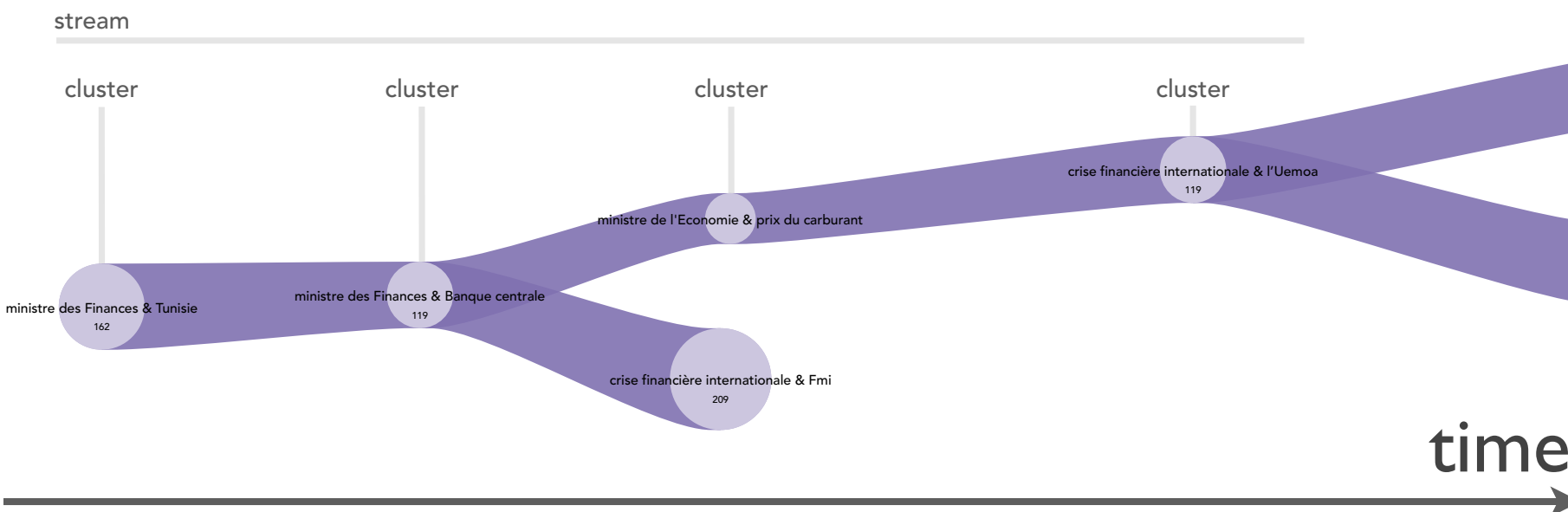
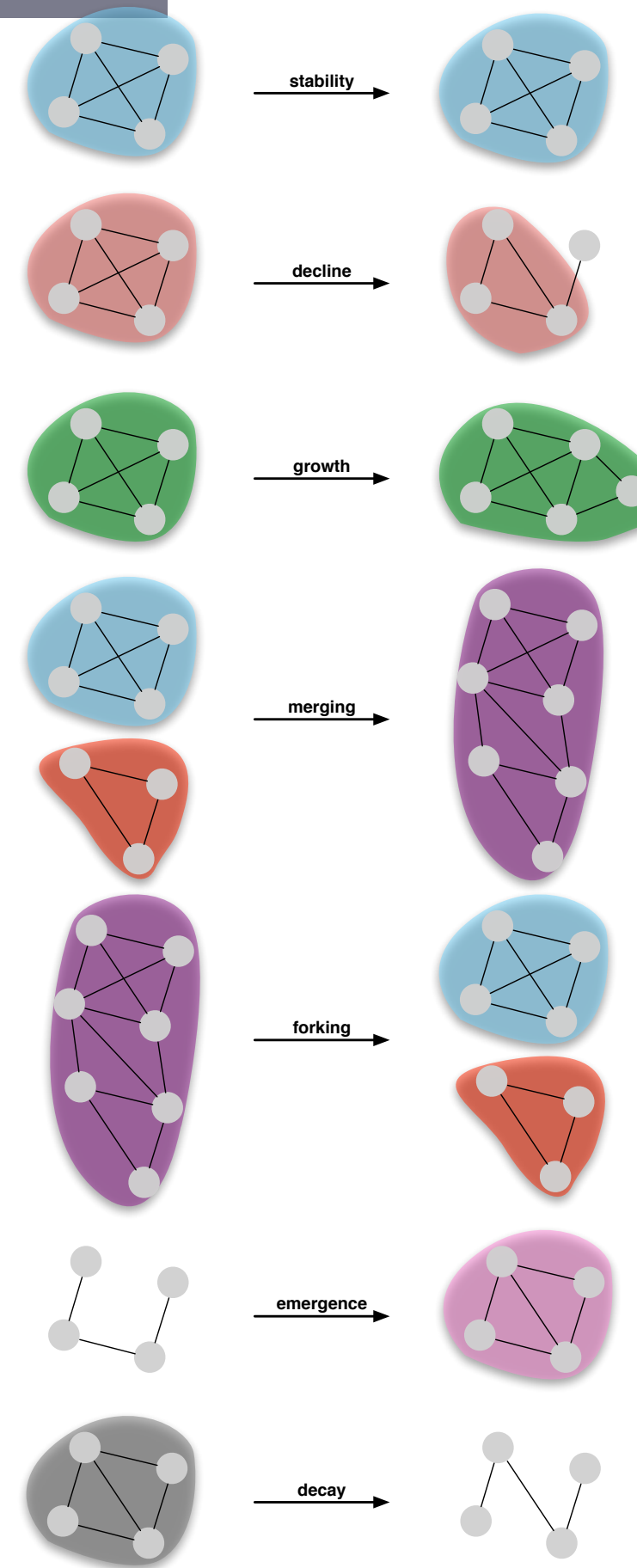
stream



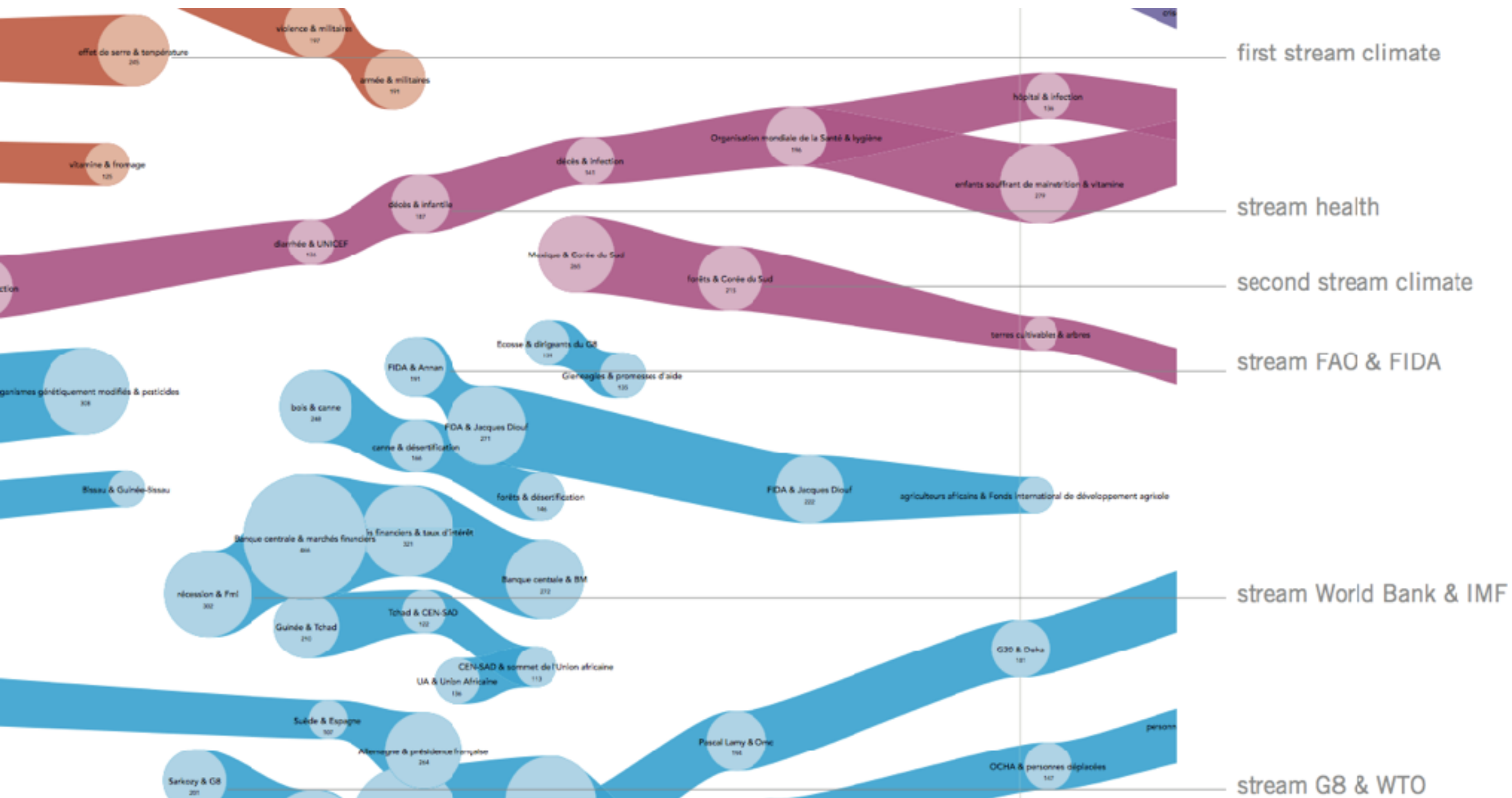
Stream Reconstruction

- **transformations** of attractors in time
- Attractors are intertemporally grouped into **streams** according to their lineage

Possible events at mesoscopic level



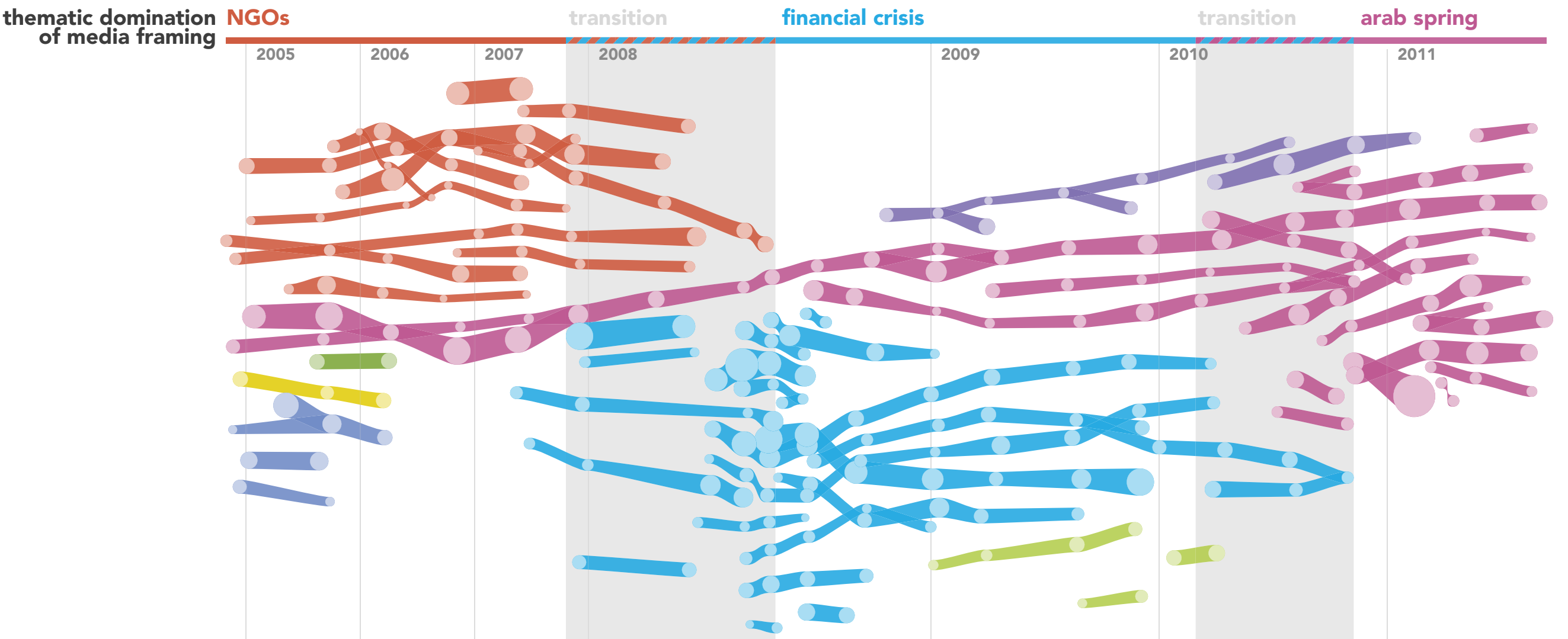
Frames dynamics: food security



web platform:
<http://pulseweb.veilleedynamique.com>

*Proximity metrics between streams
shape the global landscape of
Frames Dynamics*

Media Frames of World Food Security



2004 - 2008

food insecurity in humanitarian crises

caused by natural disasters(floods/droughts), or wars and conflicts

food security & agricultural policies

2008 - 2010

food insecurity & poverty: the '08 financial crisis

food riots spreading, increased number of undernourished people, management of food as a global issue by international institutions

2010 - 2011

food insecurity: a cause of social unrest from global crisis to local consequences
food insecurity: an health problem
continuity of infant malnutrition issues

*qualitative analysis by
Lise Cornilleau*

Major changes of the sociotechnical conditions in which publics are enacted

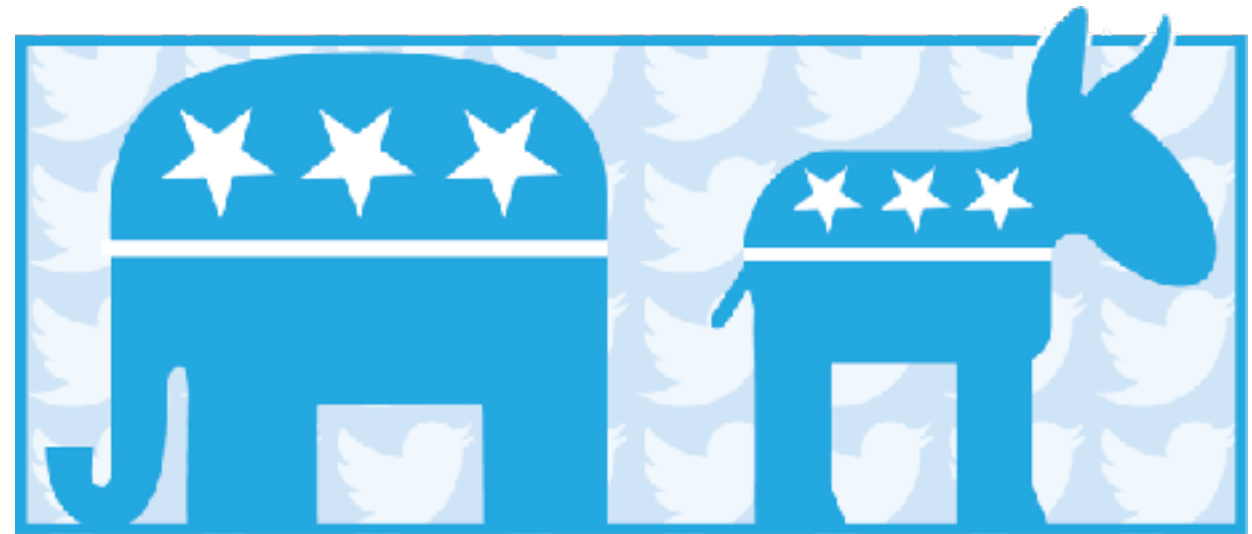


Can we still consider these collectives emerging from such digital artifacts as "publics"?

Social media enthusiasts : Digital protests

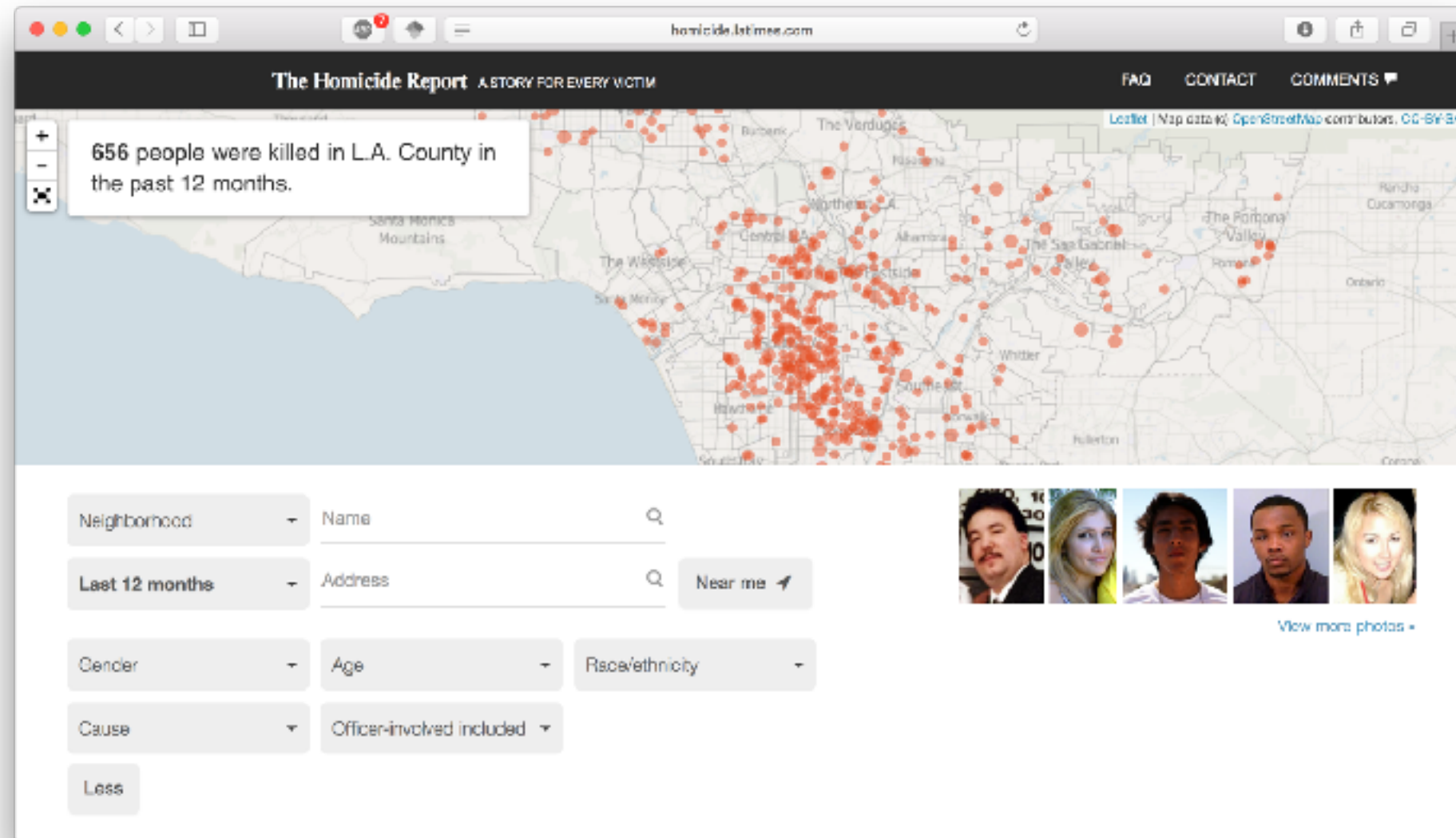


Pessimist observers: the risk of fragmentation (Sunstein, 2001)



Claims

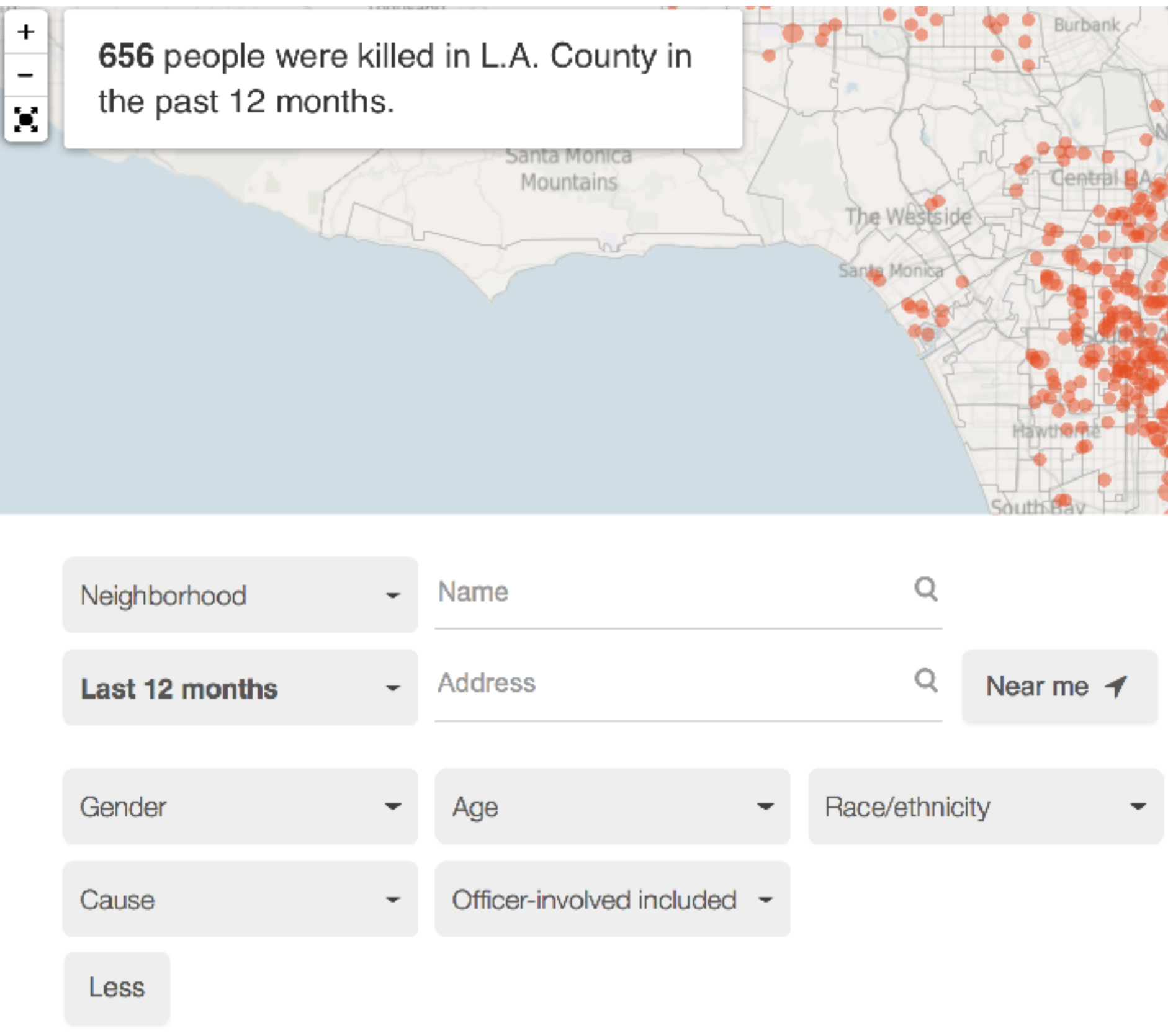
- Quantitative analysis of online traces is a profitable way to describe and explain how these collectives unfold



Los Angeles Times' "Homicide Report"

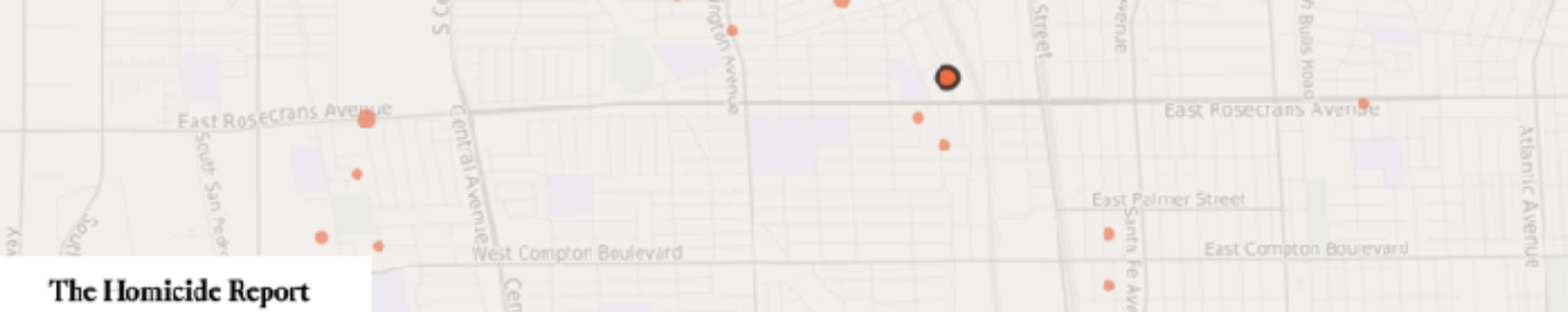
- Digital platforms enact heterogeneous publics that challenge collectives produced by traditional mass media.

Empirical setting




- 1461 homicides spanning over two years from January 2010 to December 2012
- Structured information about every victim (name, age, gender, ethnicity), homicide circumstances and location

Empirical setting



The Homicide Report



Donald Kelly, 29
POSTED FEB. 23, 2011, 9:06 A.M.

Donald Kelly, a 29-year-old black man, was shot and killed Tuesday, Feb. 22, in the 1300 block of East Culver Avenue in Compton, according to Los Angeles County coroner's records.

Sheriff's deputies responded to an "illegal shooting call" and found Kelly lying on the floor in a pool of blood, according to a Sheriff's Department news release.

Kelly was taken by paramedics to a hospital, where he died. According to coroner's records, he was shot in the chest.

Two reader comments

“ Donald....its been almost three years and we are still missing ur spirit. Our life changed the day u left....missing u red....

— stephanie
Jan. 23, 2014 at 3:06 p.m.

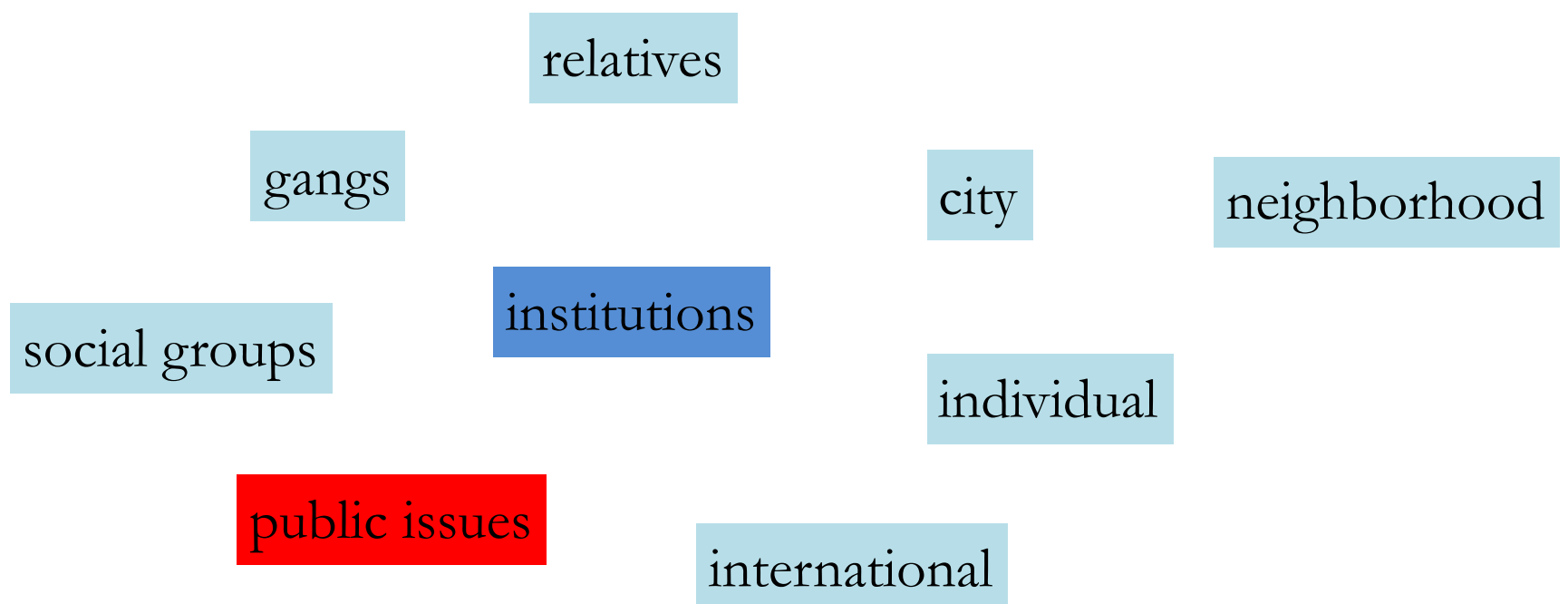
- 1461 homicides spanning over two years from January 2010 to December 2012
- Structured information about every victim (name, age, gender, ethnicity), homicide circumstances and location
- 18 648 comments

Data pre-processing

Comments are coded:

- Actors/collectives mentioned in the comment

“Mighty Mike, disobeying orders from **cops** should not be enough of a reason to pull the trigger and kill the suspect. I guarantee you if **cops** were held **accountable** for their actions just like the average **citizen** they will not be so quick to pull the trigger.



Data pre-processing

Comments are coded:

- Actors/collectives mentioned in the comment
- Link between the author and the victim

“ I love and miss you Cousin your always in my Heart ”

Address to the victim

Familial link

Friendship link

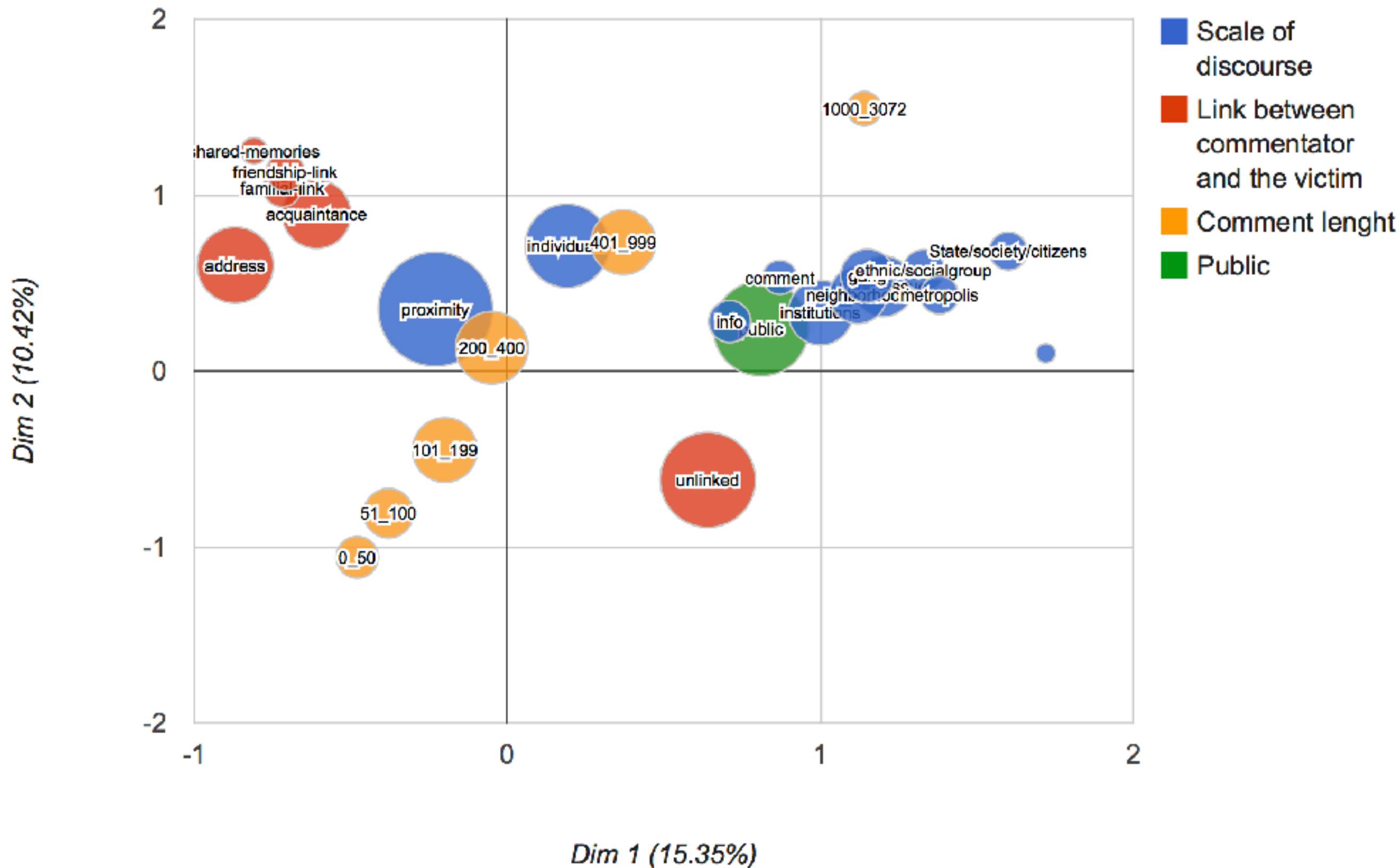
no link

Acquaintance

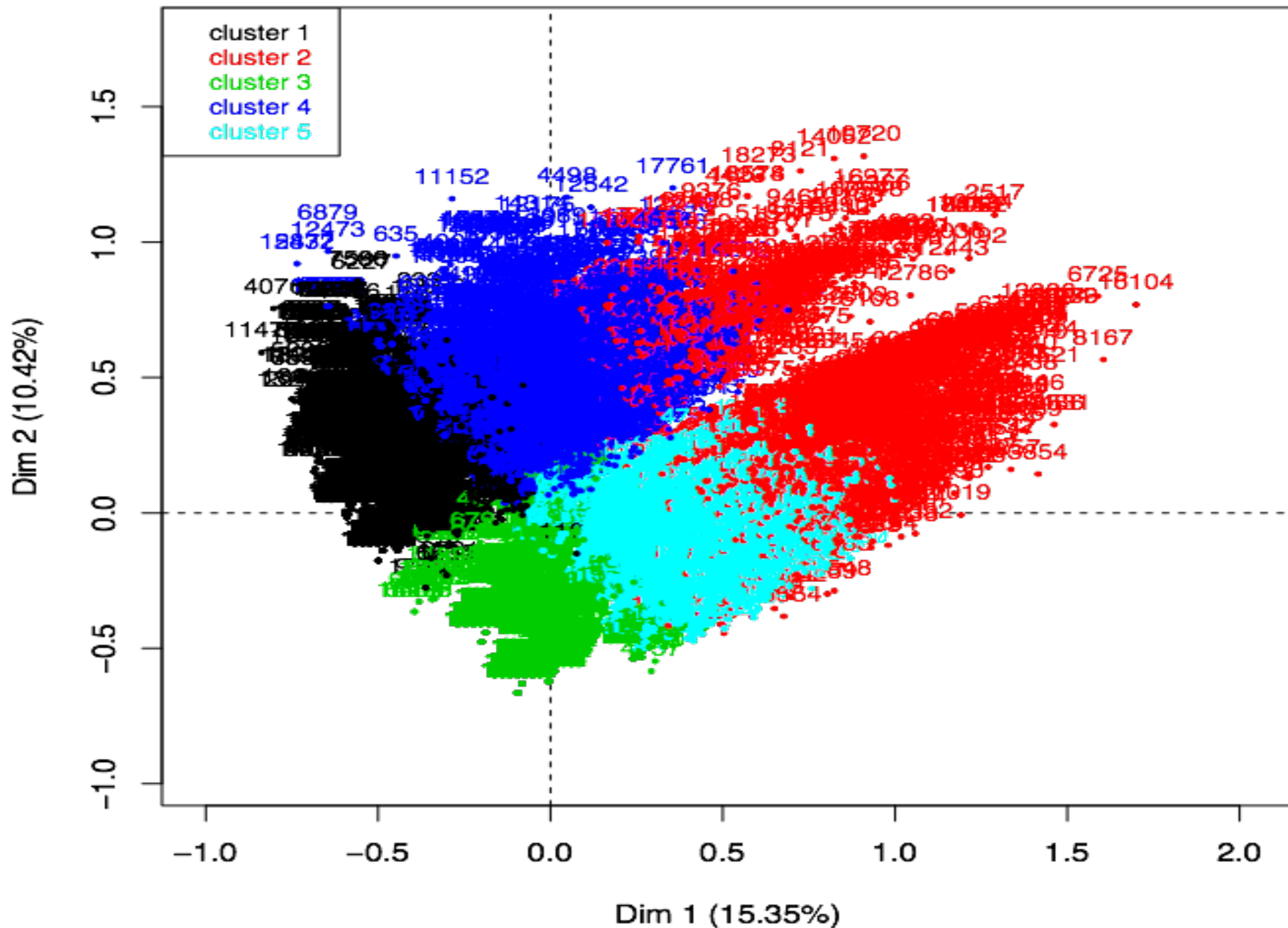
Shared memories

Understanding Coverage

LA Times Homicide Report - MCA analysis of comments



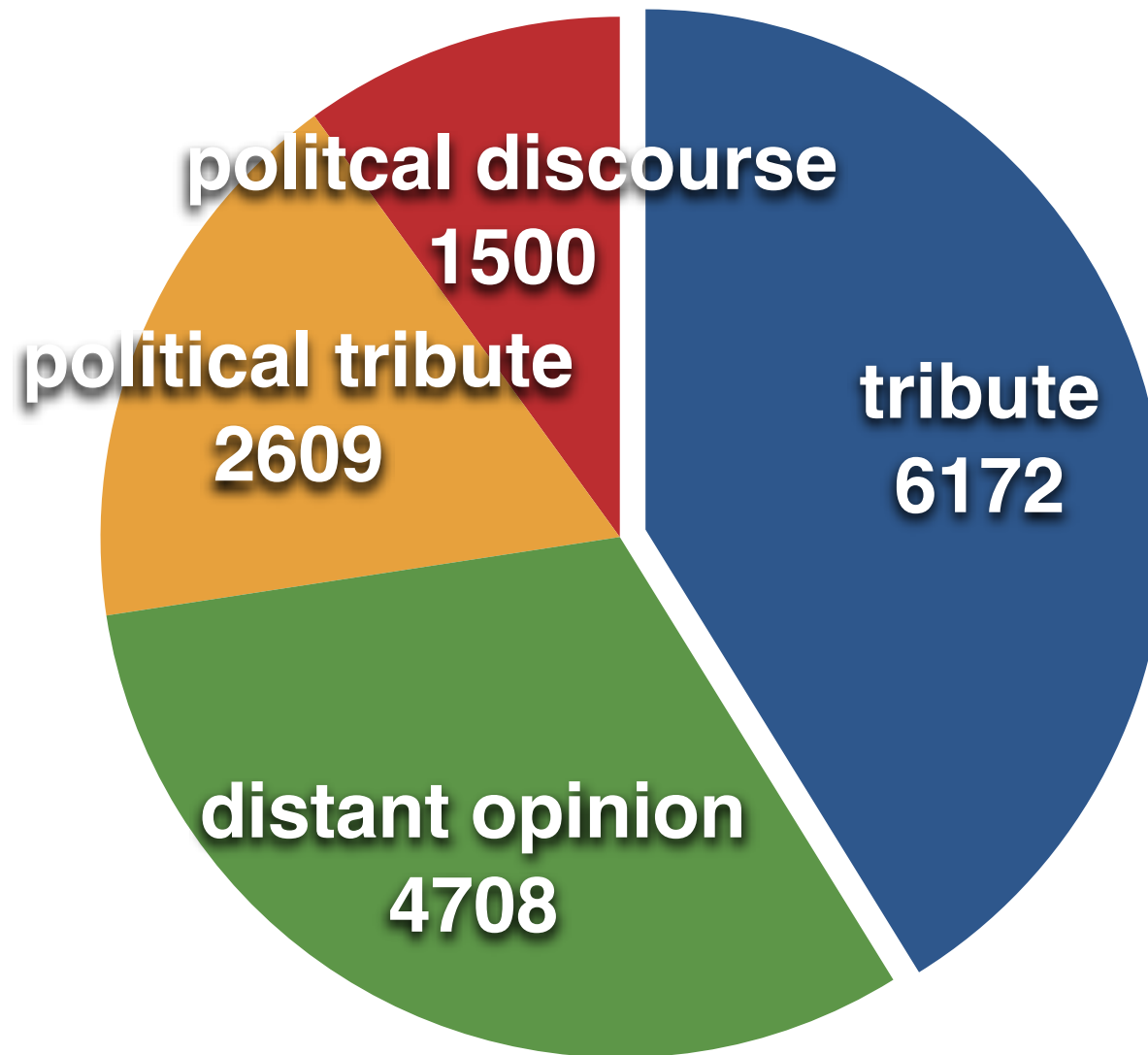
Understanding Coverage



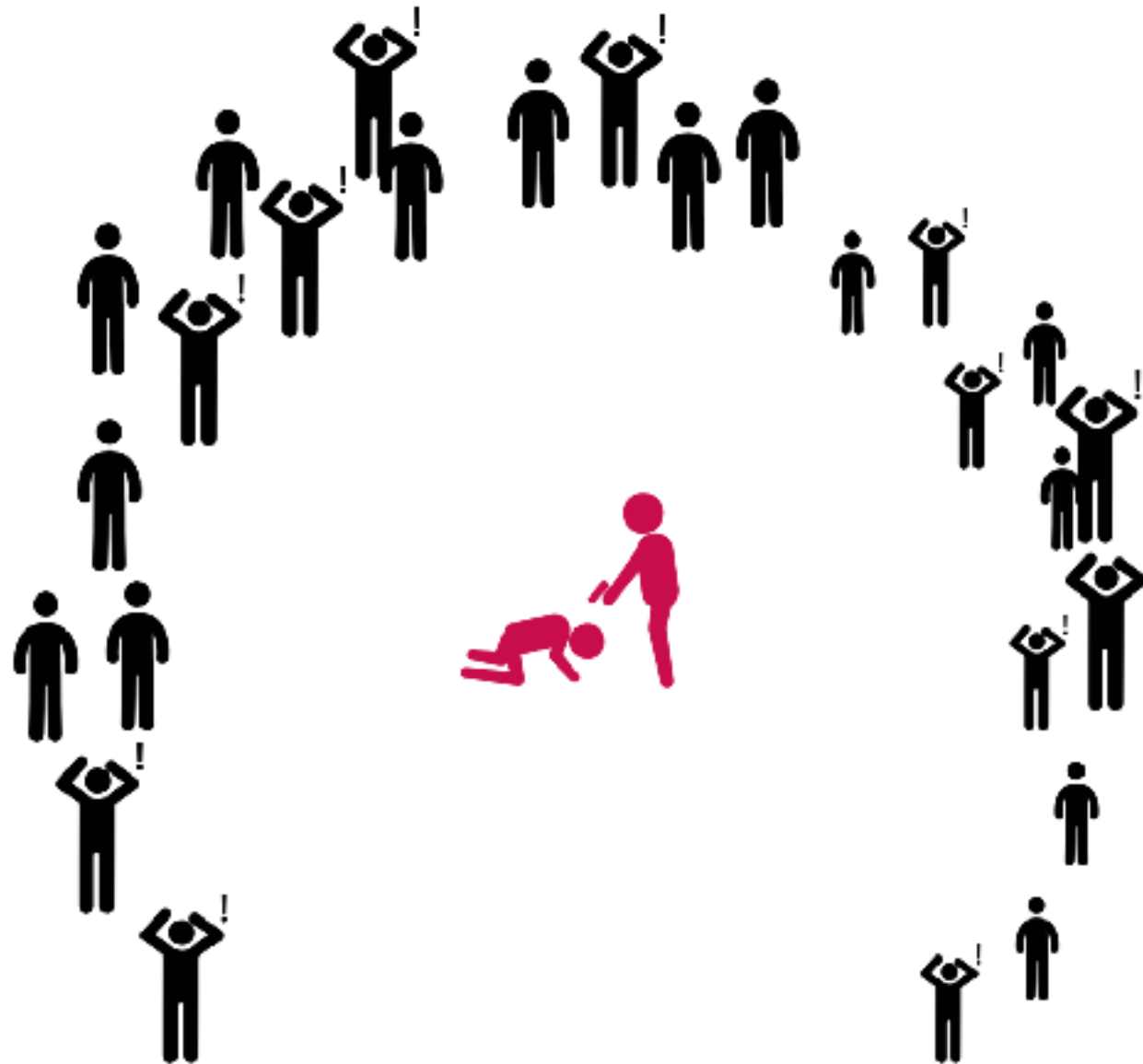
Data pre-processing

Comments are coded:

- Actors/collectives mentioned in the comment
- Link between the author and the victim
- Comments are clustered in 4 groups



1 - The “Invisible Coliseum”



- Typical of mass media
- A small number of selected occurrences
- Large sets of distant individuals

Problem n°1 : We know little about the authors of comments

Participation is locally-based:

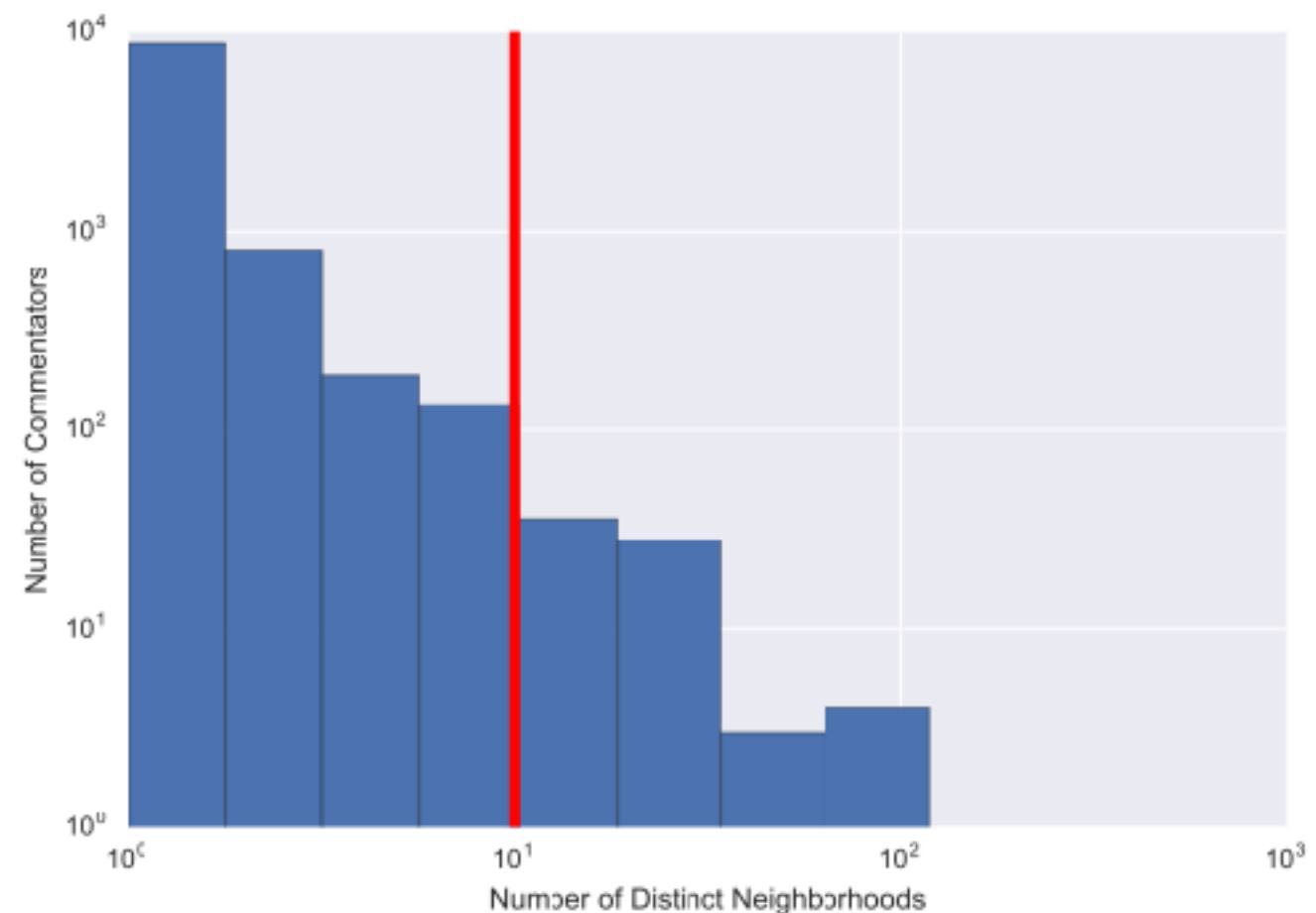
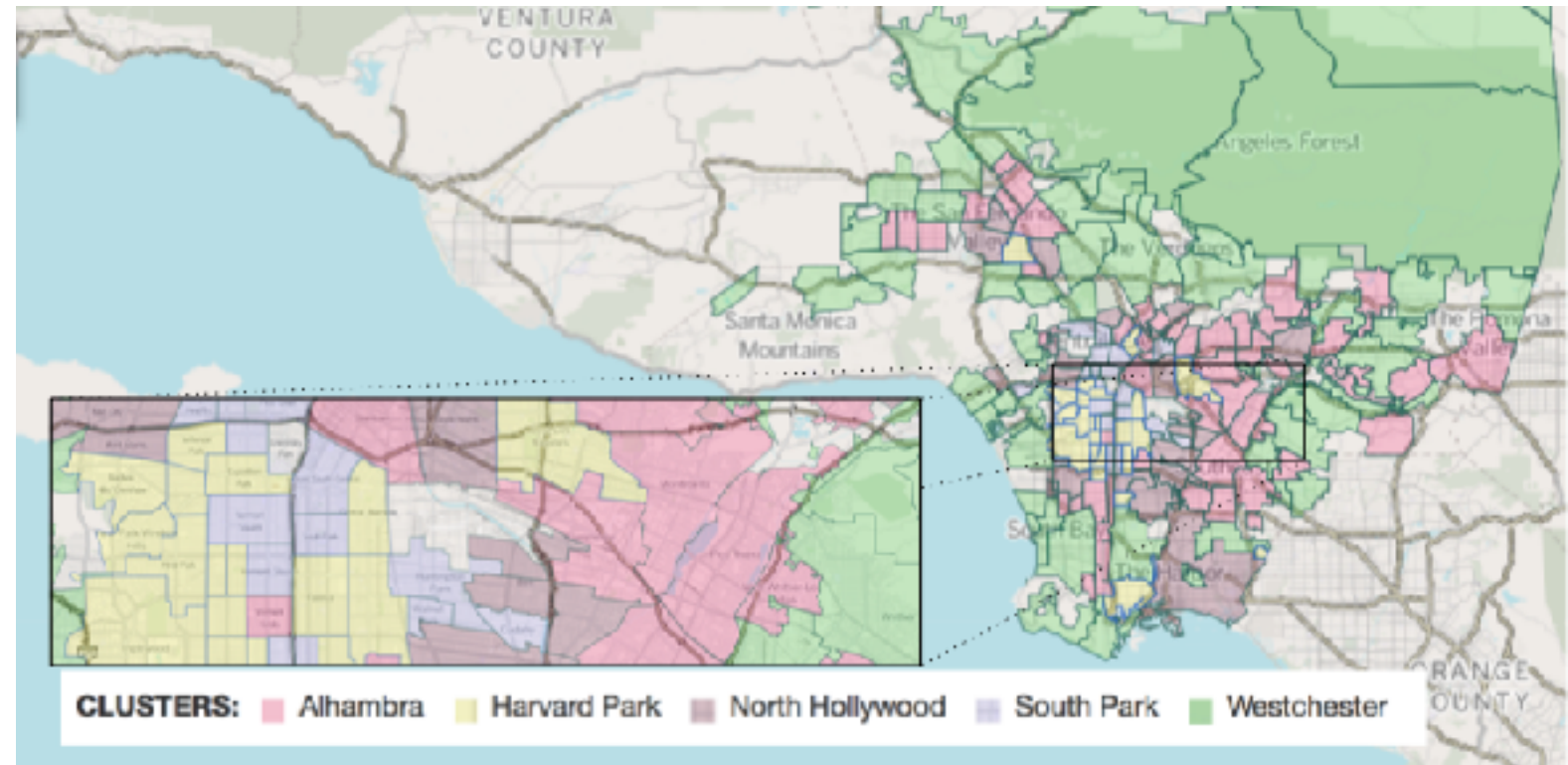
88% of authors comment homicides that occur in a single neighborhood

Methodological choices:

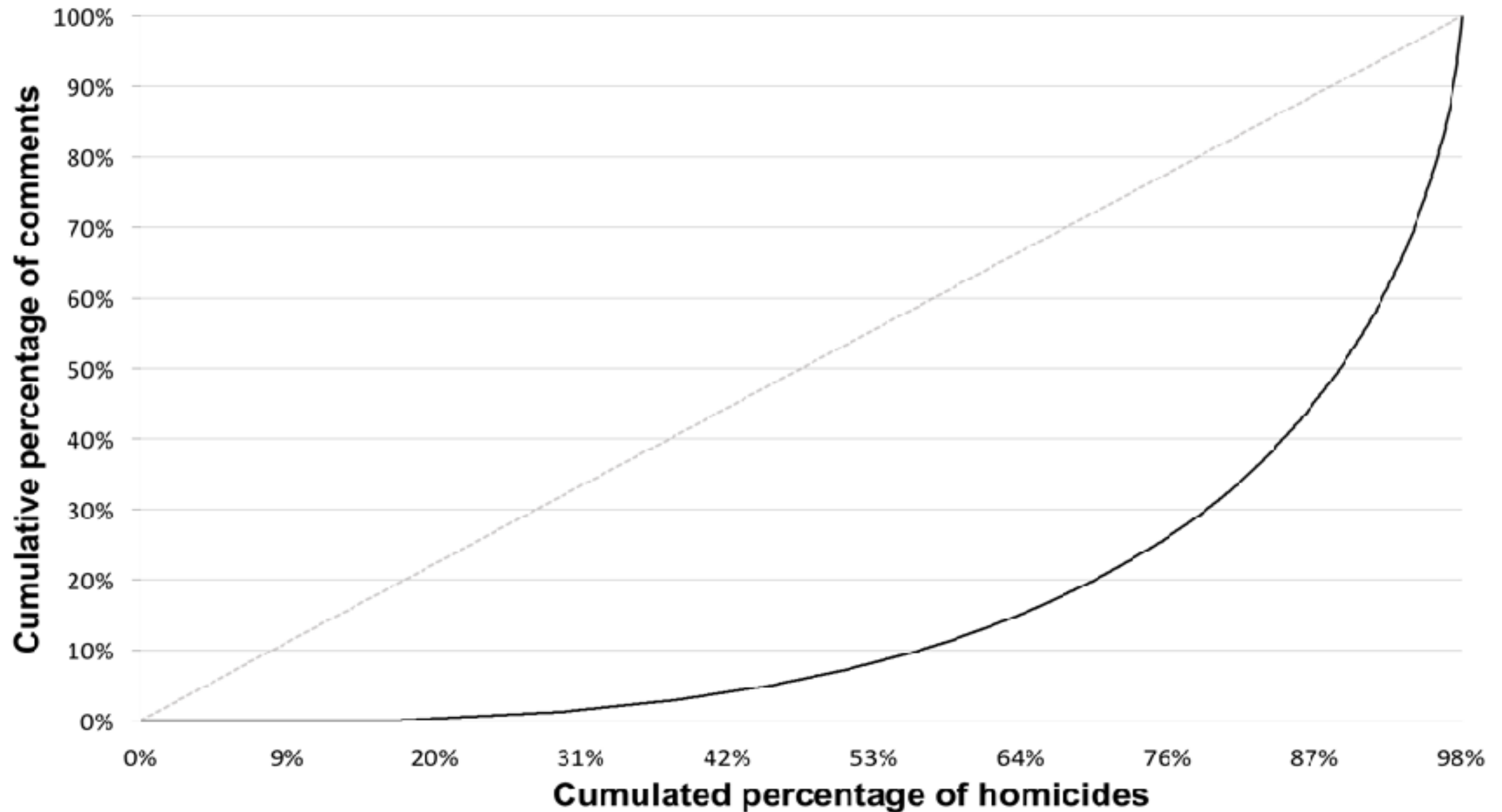
- Collecting data about neighborhoods is a way to add some information about authors
- Two population of contributors : local authors /metropolitan authors

Data pre-processing

- **Neighborhoods** are grouped in 5 different clusters
- **9985 authors** are either categorized as local posters or metropolitan posters (49)

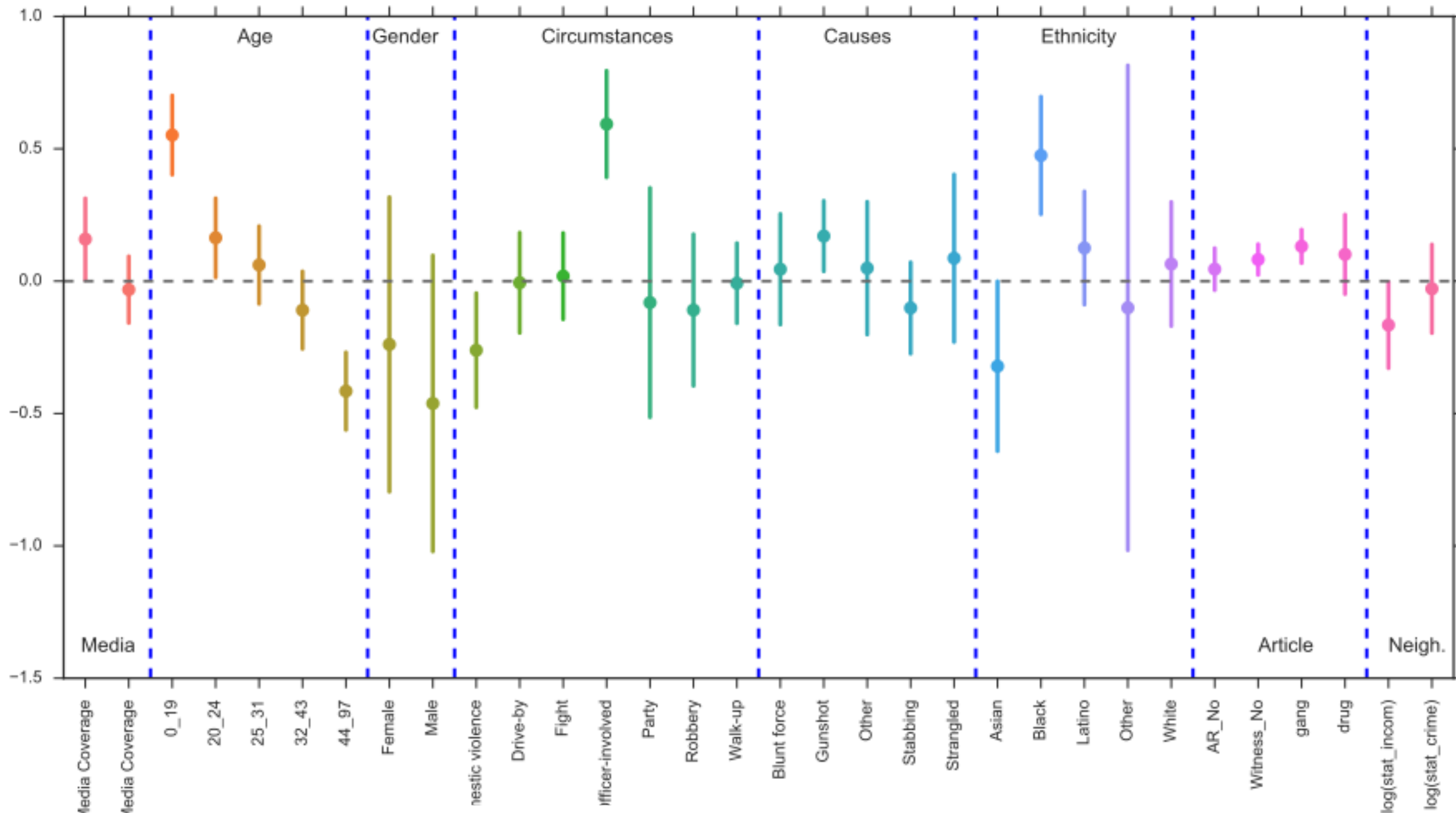


Unbalanced Attention to Homicides



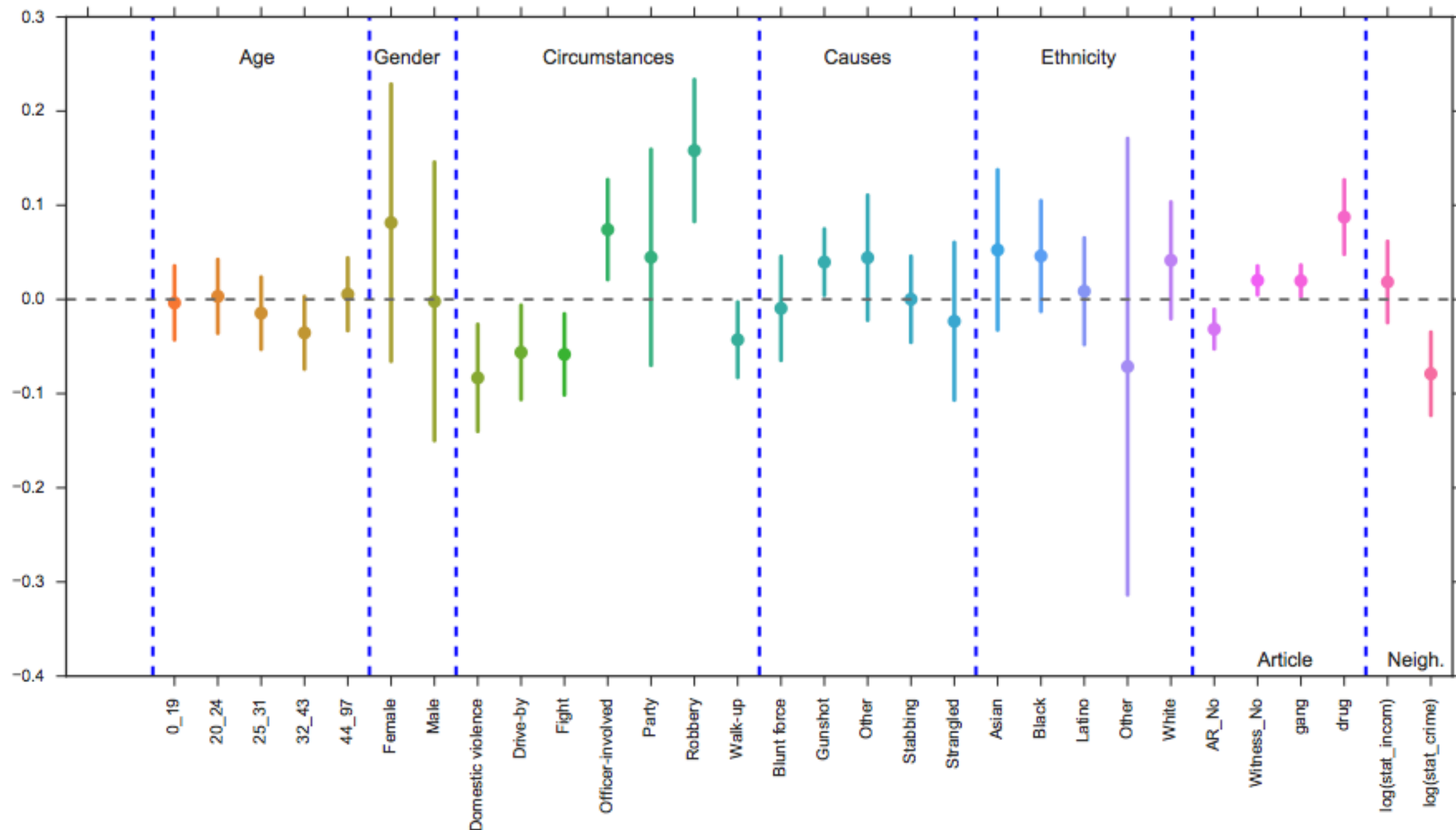
80% of comments focus on 30% of homicides

Understanding Participation



Regression analysis to predict the number of comments

Understanding Participation



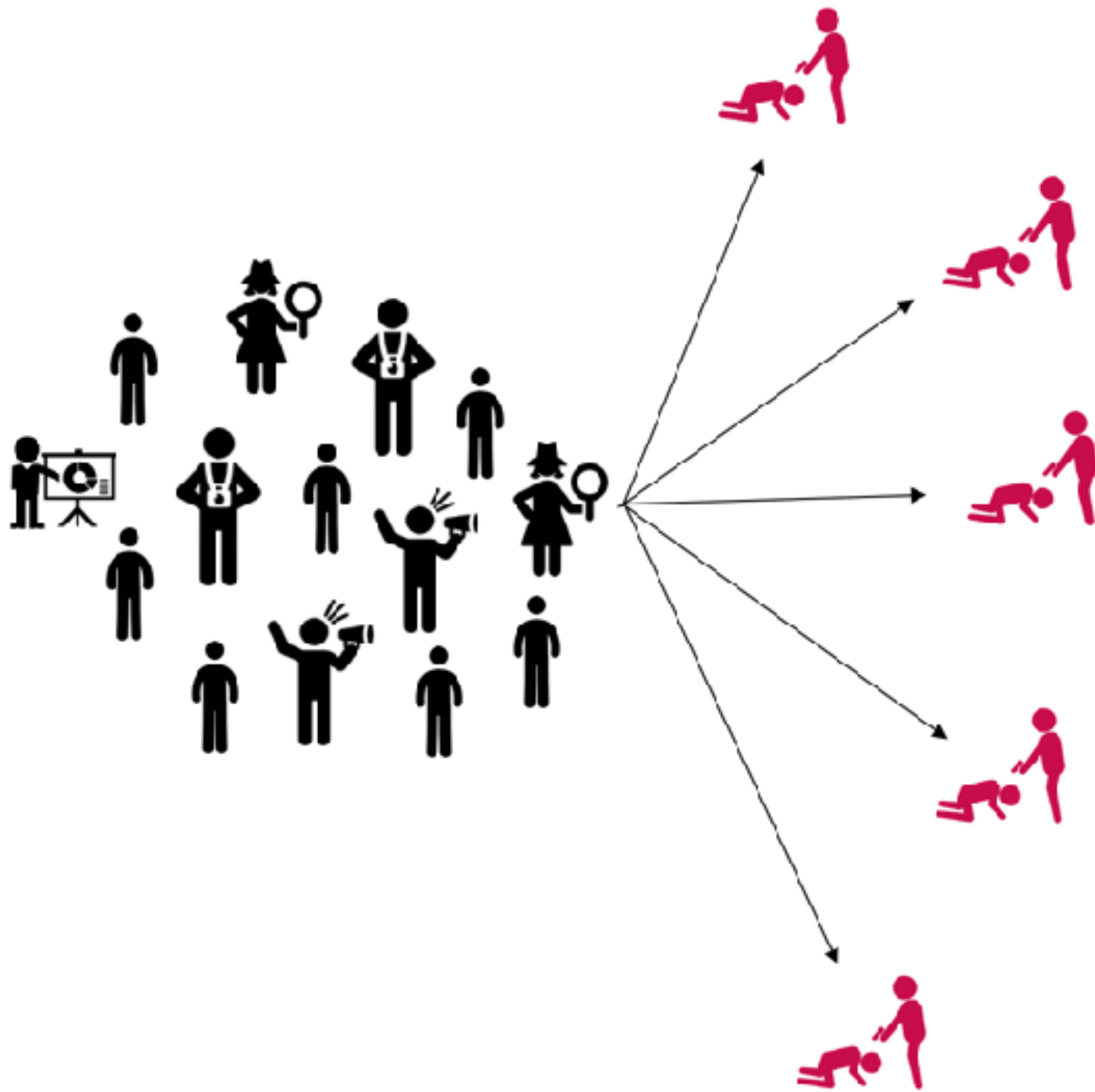
Conversely to predict homicides media coverage (print)

Understanding Coverage

	Printed LA Times coverage	Homicide report coverage
positive effet	Robbery Officer-involved	Young victims Black Victims Officer-Involved Gang-related
negative effect	Street Violence High crime rate neighborhood	Senior Victims Rich Neighborhoods

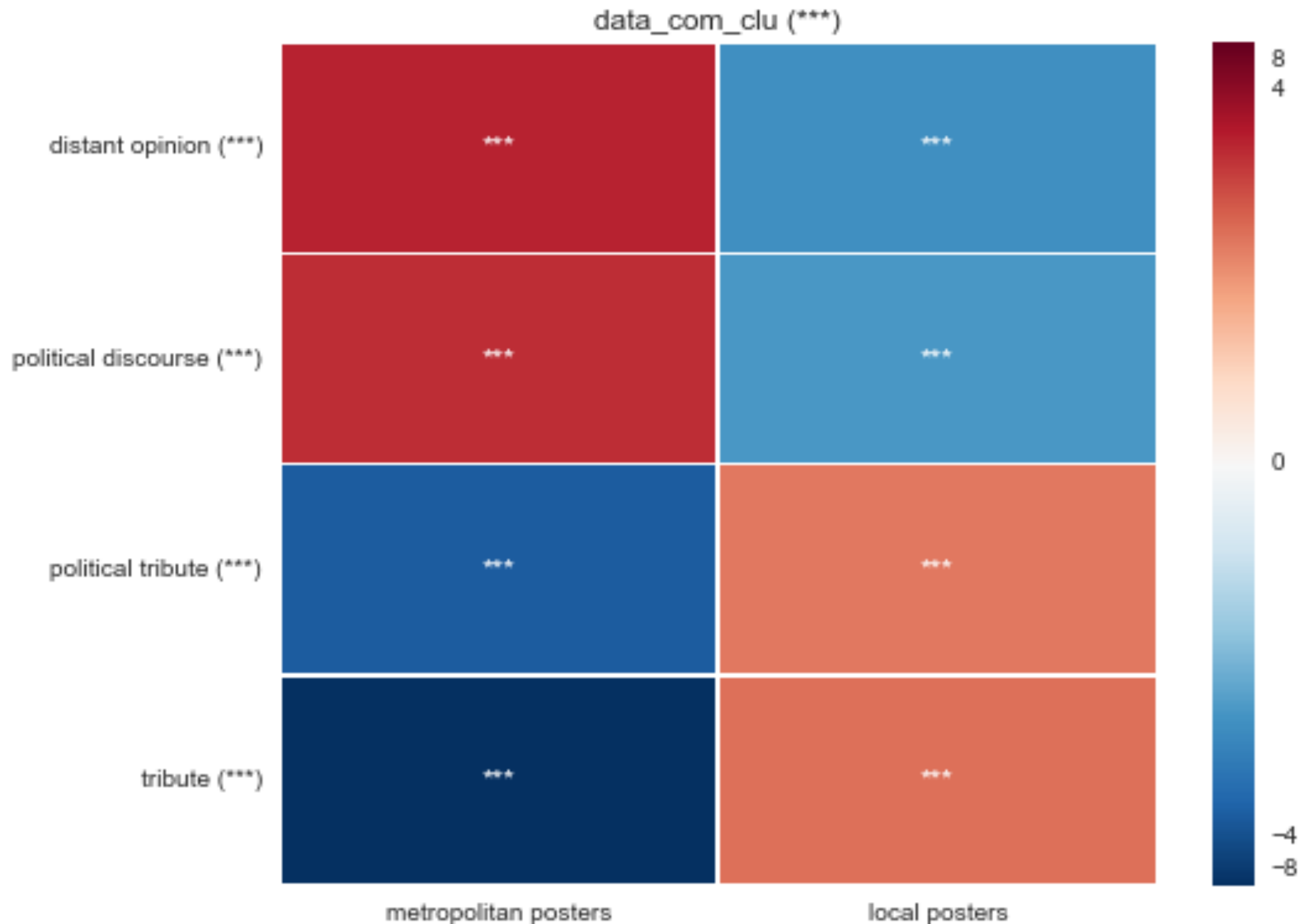
(Regression Analysis Results)

2 - The collective of inquiry



- Mobilization of activists
- A process of collective interpretation

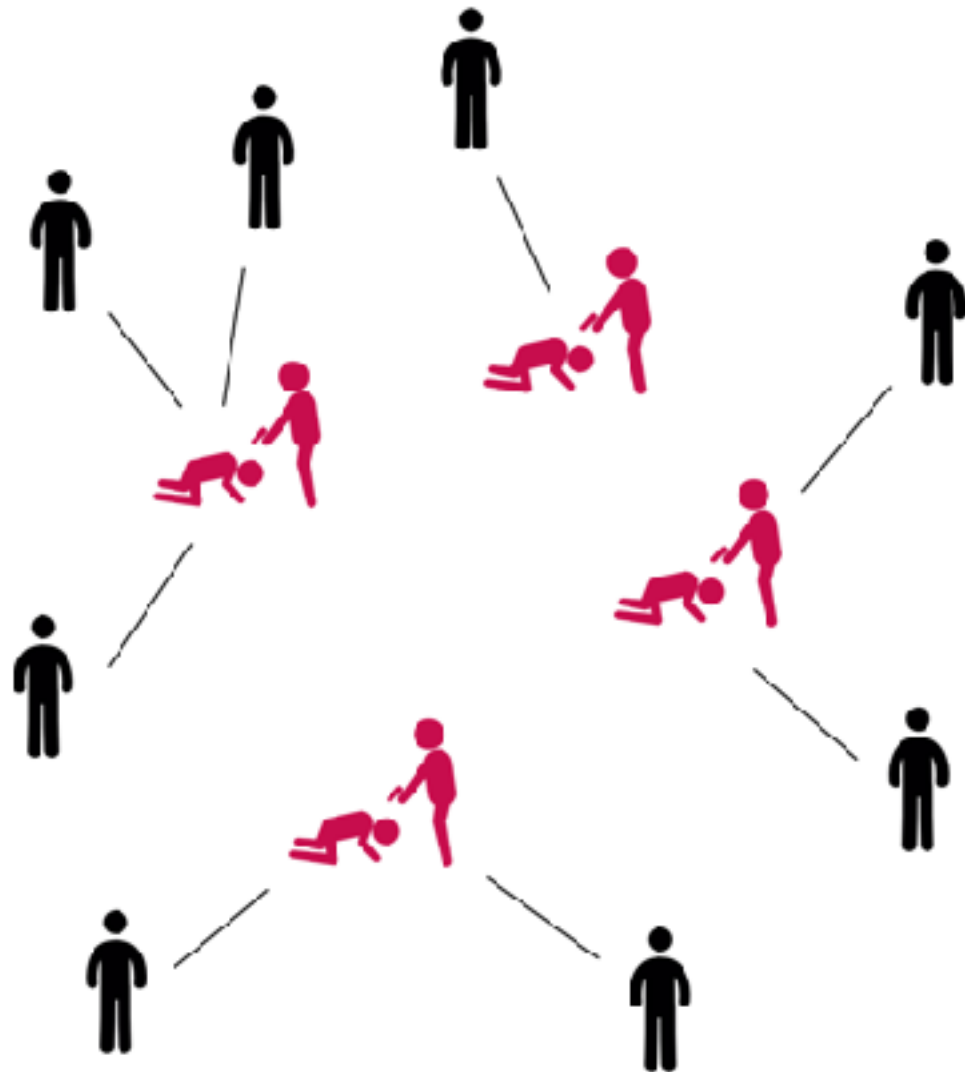
Different users, different discourses



“The under privilege people, which are for the most part minorities, are the ones who get the short end of the stick. Many have to take deals in order to avoid serving more time for a crime they did not commit. I avoid playing the race card, but when a white homeless man gets killed by cops in fullerton, some cops are actually held accountable. I'm all for justice, but when a cop kills a hispanic or black justice is not served. **Syscom3**, you like stats and I was wondering if you happen to have the demographic breakdown of people who are charged with a crime and are actually found guilty. I'm almost certain the White people get more of a break than minorities. It's sad but true.”

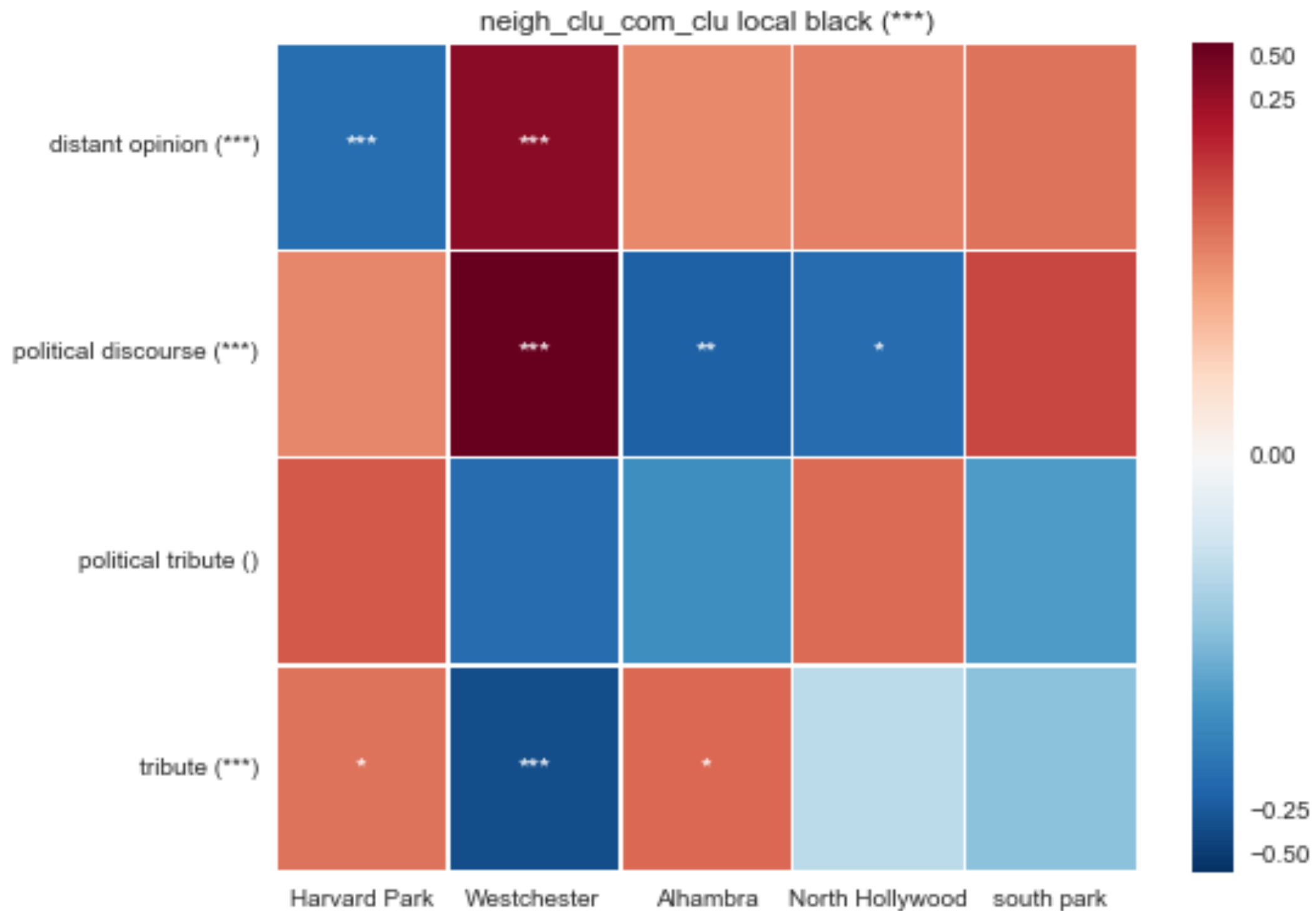
[Jag; Nov. 17, 2012 at 10 a.m.; homicide of Armondo Casillas]

3 – Multitude of Neighbors



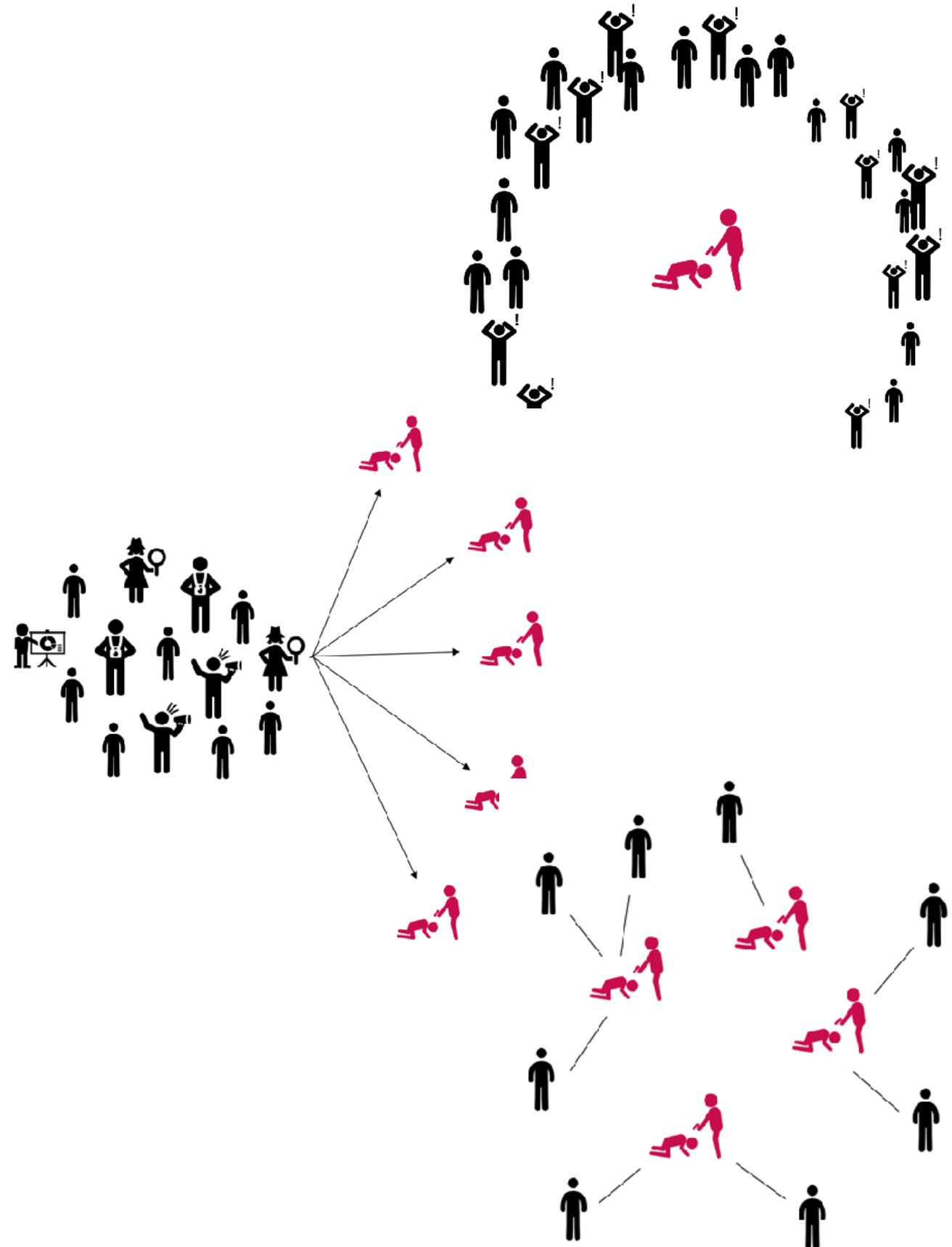
- Large number of occurrences
- Interest based on proximity

Different communities, different discourses

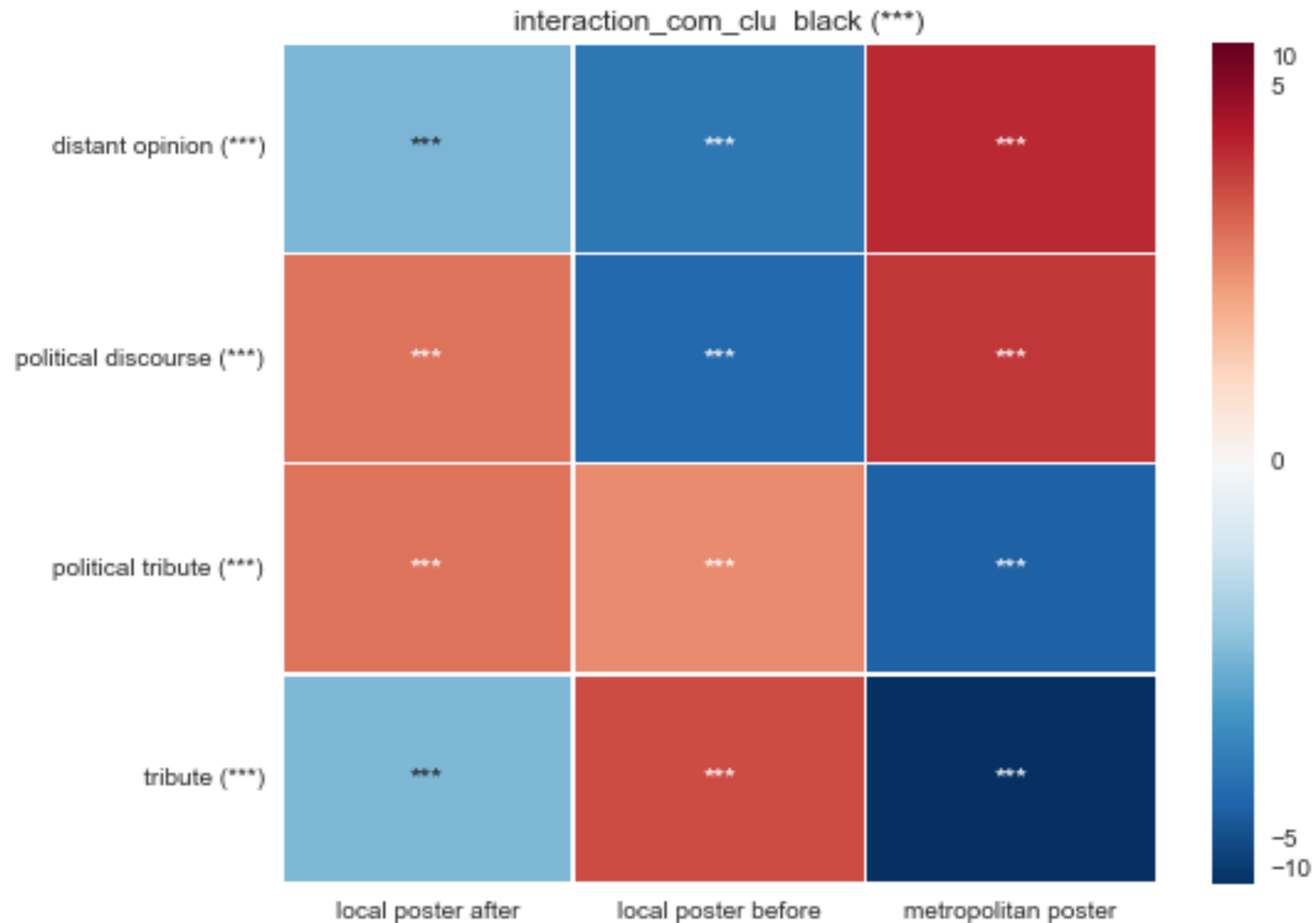


Three Different ways of making public

- The “**Invisible Coliseum**”
- The collective of inquiry
- The multitude of neighbors



Three Publics interacting in the same space



Case study Conclusions

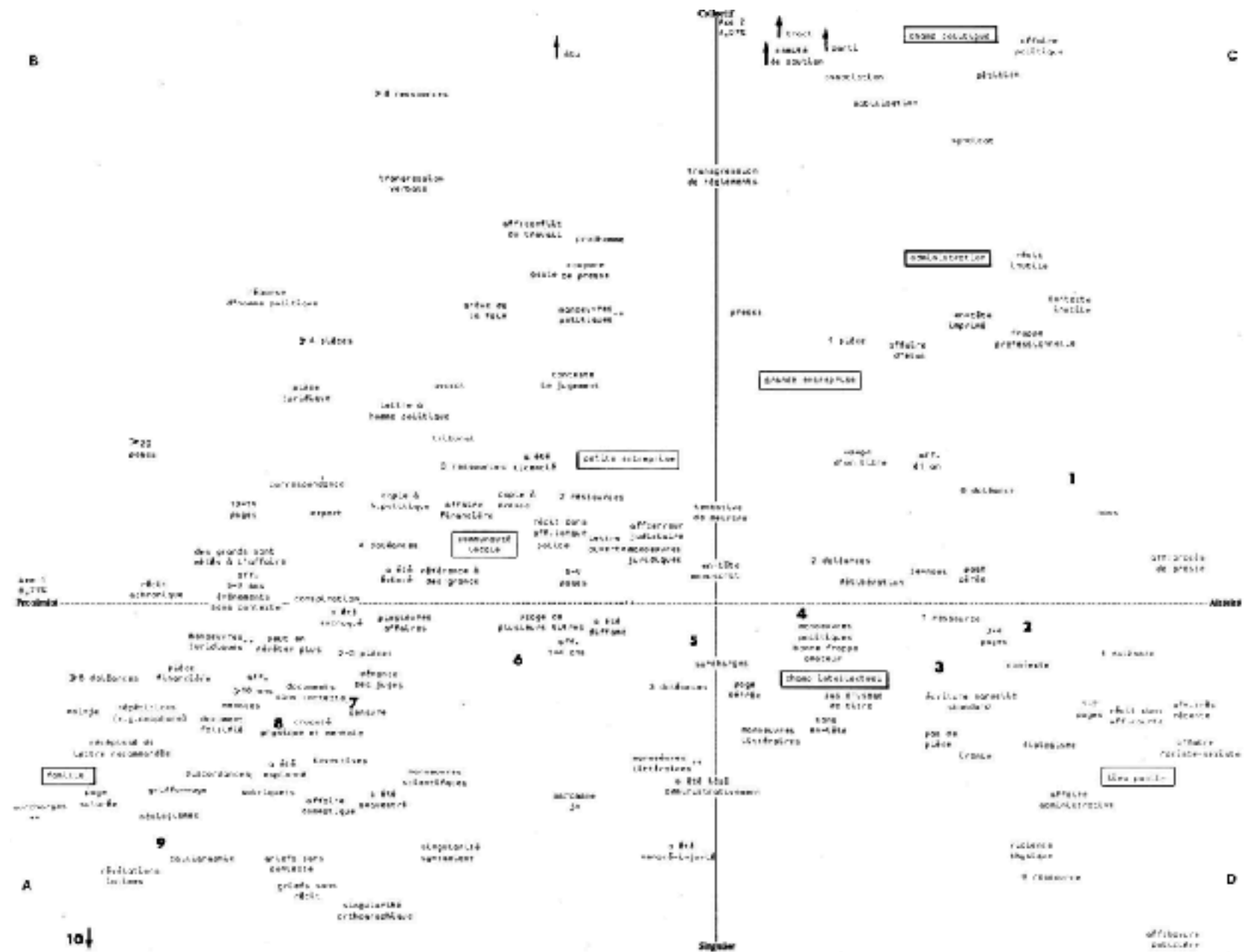
- Digital platforms enact heterogeneous forms of publics, which shares some aspects with traditional publics
- Quantitative analysis of online traces is a profitable way to account how these collectives unfold
- Interaction between those different forms of public

Une stratégie parmi d'autres

- Analyse de correspondance, « La dénonciation » de L. Boltanski
- Enquêter avec Marlowe (F. Chateauraynaud)
- Cartographier les réseaux d'actants au sein des controverses socio techniques (B. Latour)

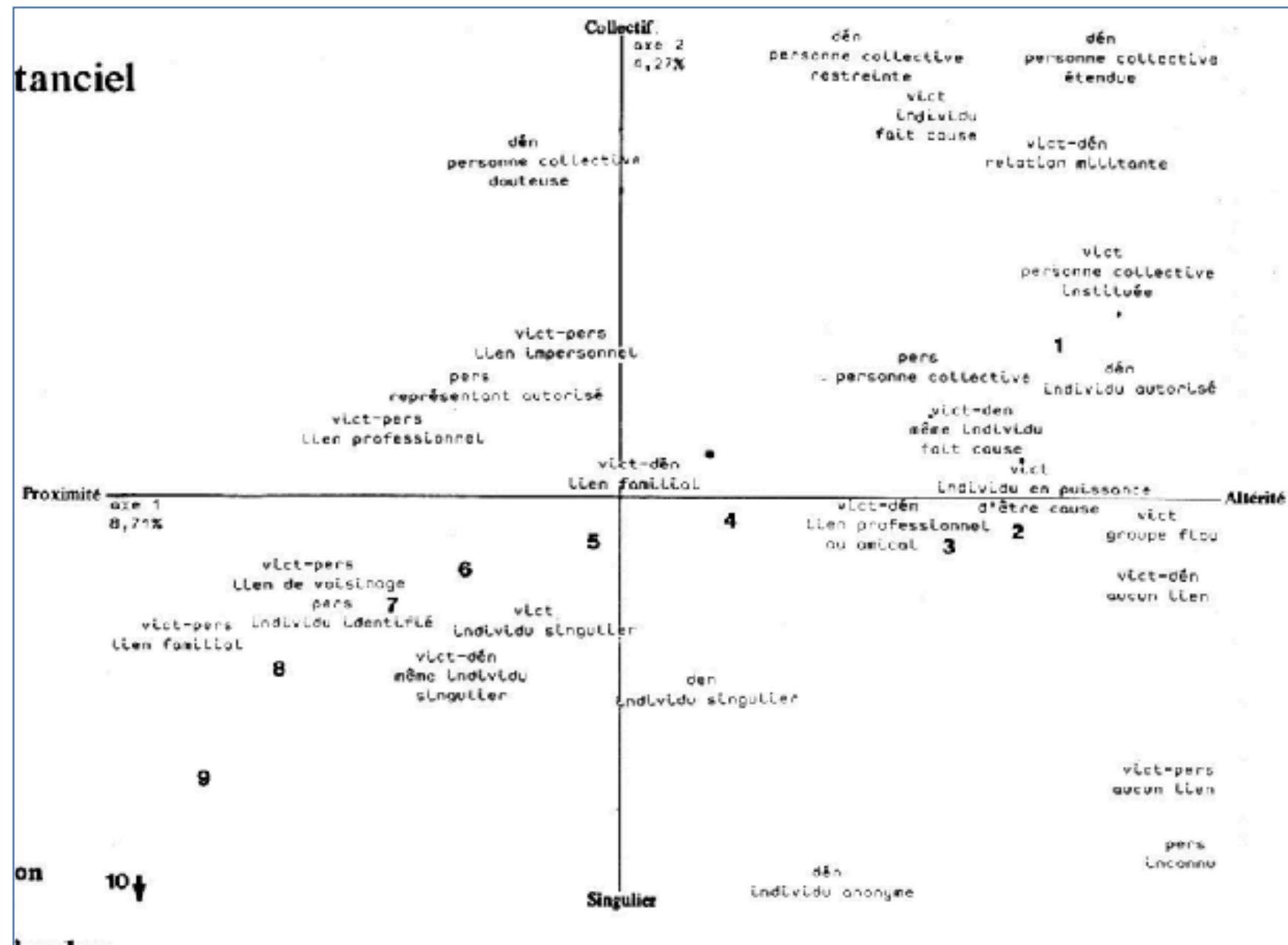
« La dénonciation » de L. Boltanski

- Qu'est ce qui fait qu'une dénonciation publique peut-être jugée comme normale ou anormale ?
- Codage de lettres de dénonciation envoyées au Monde. On code le contenu de l'affaire, la lettre (dans sa dimension graphique, stylistique, réthorique) mais aussi les propriétés sociales de l'auteur. Et on confronte ce codage à des jugements de normalité.

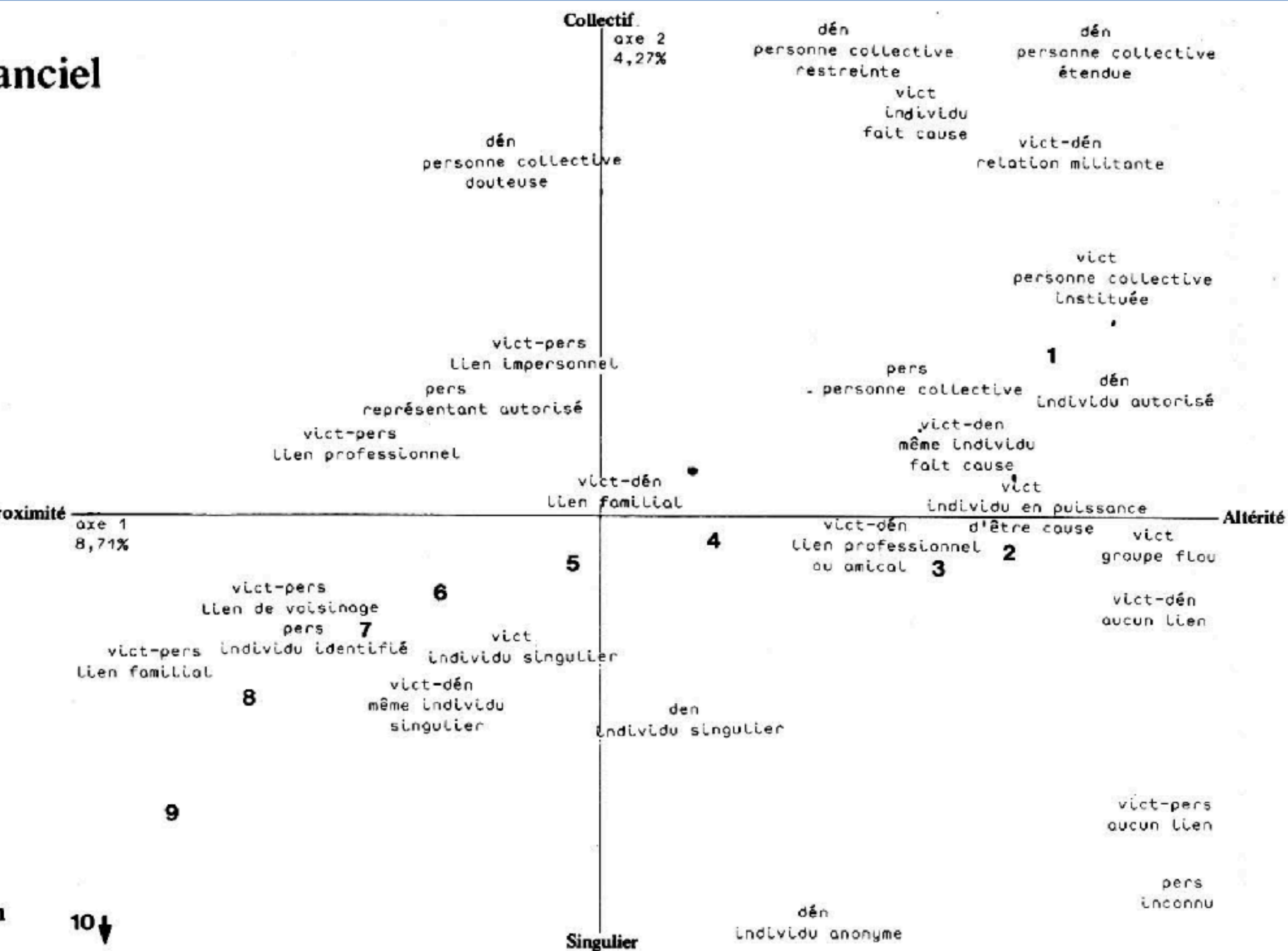


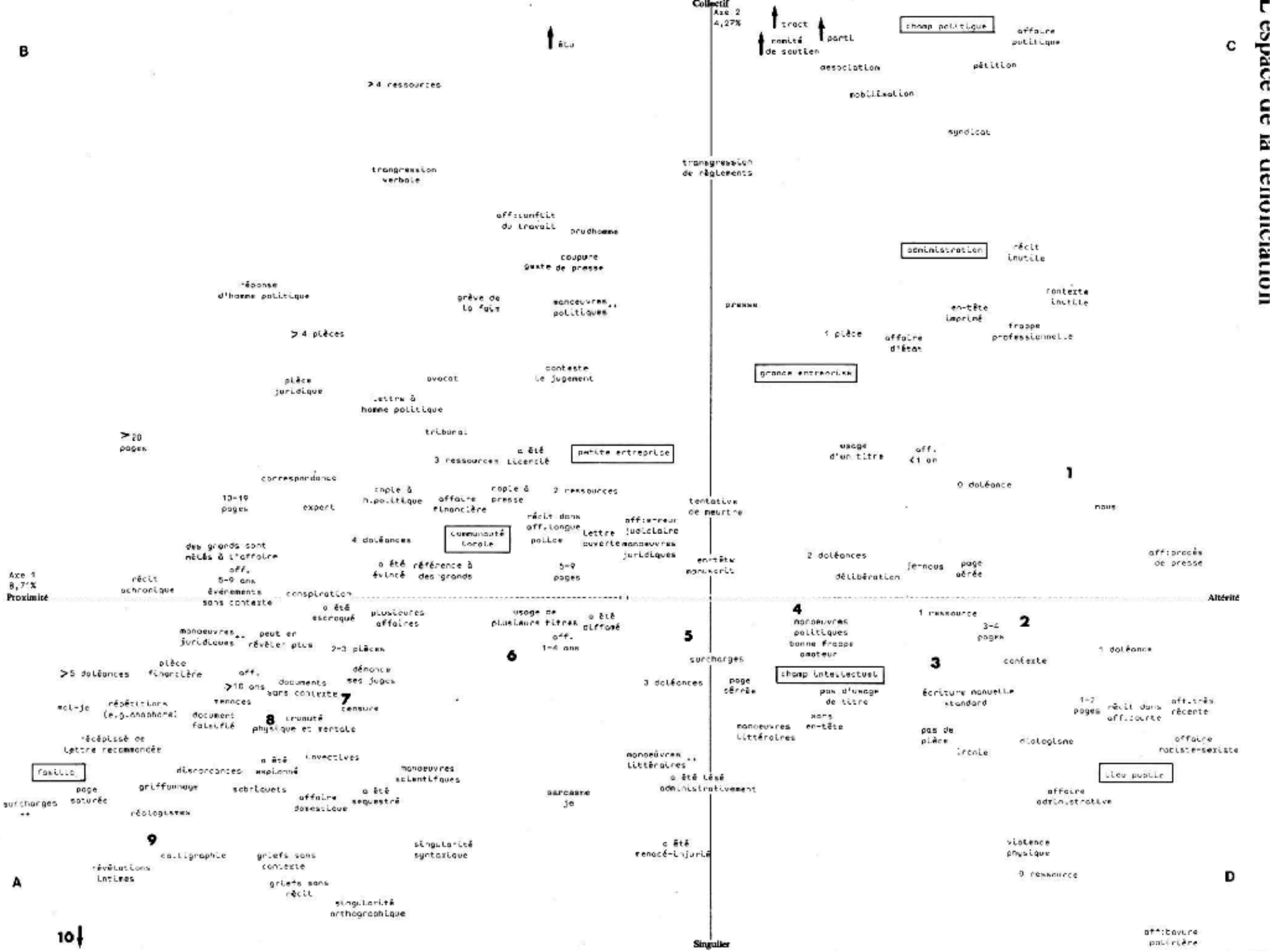
« La dénonciation » de L. Boltanski

- Qu'est ce qui fait qu'une dénonciation publique peut-être jugée comme normale ou anormale ?
- Codage de lettres de dénonciation envoyées au Monde. On code le contenu de l'affaire, la lettre (dans sa dimension graphique, stylistique, rhétorique) mais aussi les propriétés sociales de l'auteur. Et on confronte ce codage à des jugements de normalité.
- Interprétation via système actantiel (dénonciateur, victime, persécuteur, juge)



ancier



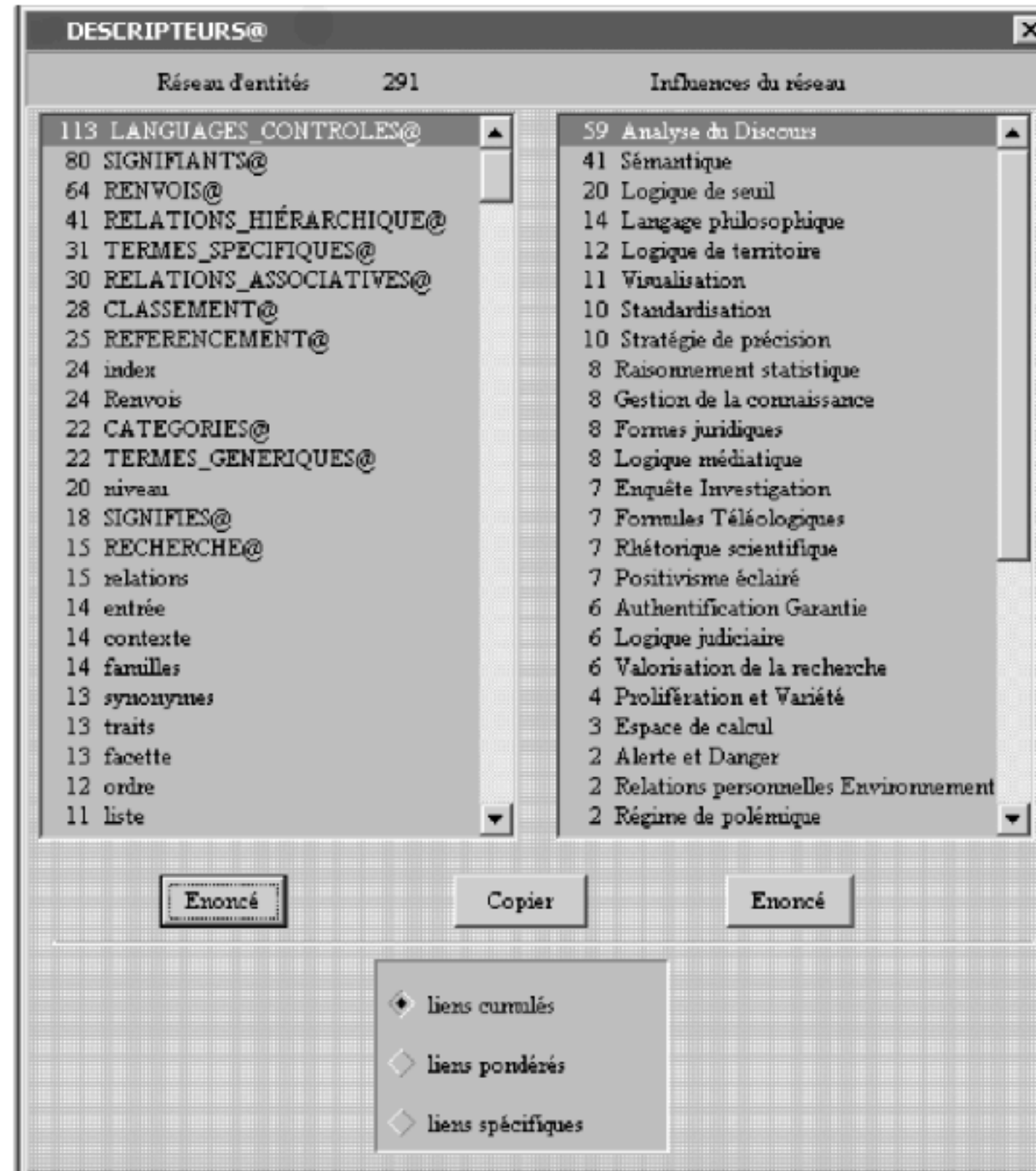


PROSPERO

- PROgramme de Sociologie **Pragmatique**, Expérimentale et Réflexive sur Ordinateur
- Procédures automatisées limitées et toujours sous contrôle – critique forte des approches à base d'AFC comme Alceste (faussetement inductives, séparant résultat et interprétation)
- Corpus normés et relativement impersonnels

PROSPERO – codage des éléments textuels

- Entités: personnes, institutions, valeurs, émotions, lieux, concepts, etc.
- Qualités: qualifications attachées aux entités
- Epreuves: décrivent les relations entre entités (défendre, mobiliser, accuser, etc.)
- Marqueurs
- Mots-outils
- Nombres
- Indéfinis



PROSPERO – éléments recombinaés entre eux

- Expressions qui combinent entités et qualités: « emploi fictif »
- Etres fictifs : ensemble de toutes les variations textuelles utilisées pour faire référence à un actant.
- Catégories ou Ordres du discours: composées de groupe d'entités, de qualités ou d'épreuves, elles permettent de qualifier un texte comme relevant d'un régime de controverse, pamphlétaire, etc.

PROSPERO – les types d'analyse

- Les réseaux et anti-réseaux
- Le temps des textes
- Corpus et sous-corpus

C'ELINE@	148	antisémite	4	Céline est <i>antisémite</i> comme on est poitrinaire...
		vrai	4	morceaux choisis qui sont du <i>vrai</i> Céline...
		admirable	3	faire tranquillement le départ entre le Céline condamnable et le Céline <i>admirable</i> ...
		loin	3	Céline est très <i>loin</i> du vérisme...
		seul	3	elle place désormais Céline <i>seul</i> face à ses lecteurs...
		nouveau	3	comparé le <i>nouveau</i> Céline à un nouvel hybride monstrueux de Rousseau et de Beckett...
		rentré	3	Plus tard, Céline <i>rentré</i> recommençant à publier...
		né	3	Louis-Ferdinand Destouches est <i>né</i> à Courbevoie le 22 mai 1894...
		maboul	3	Et si ce n'était pas une plaisanterie alors il serait, lui Céline, complètement <i>maboul</i> ...
		renfermé	2	toujours garder à l'esprit l'image d'un Céline blessé, solitaire, <i>renfermé</i> dans...

Figure 2. Carte de liens obtenue après projection de la catégorie « Économie de la Promesse et Prophétie de Bonheur » sur un des corpus (ici nanos_chimères)

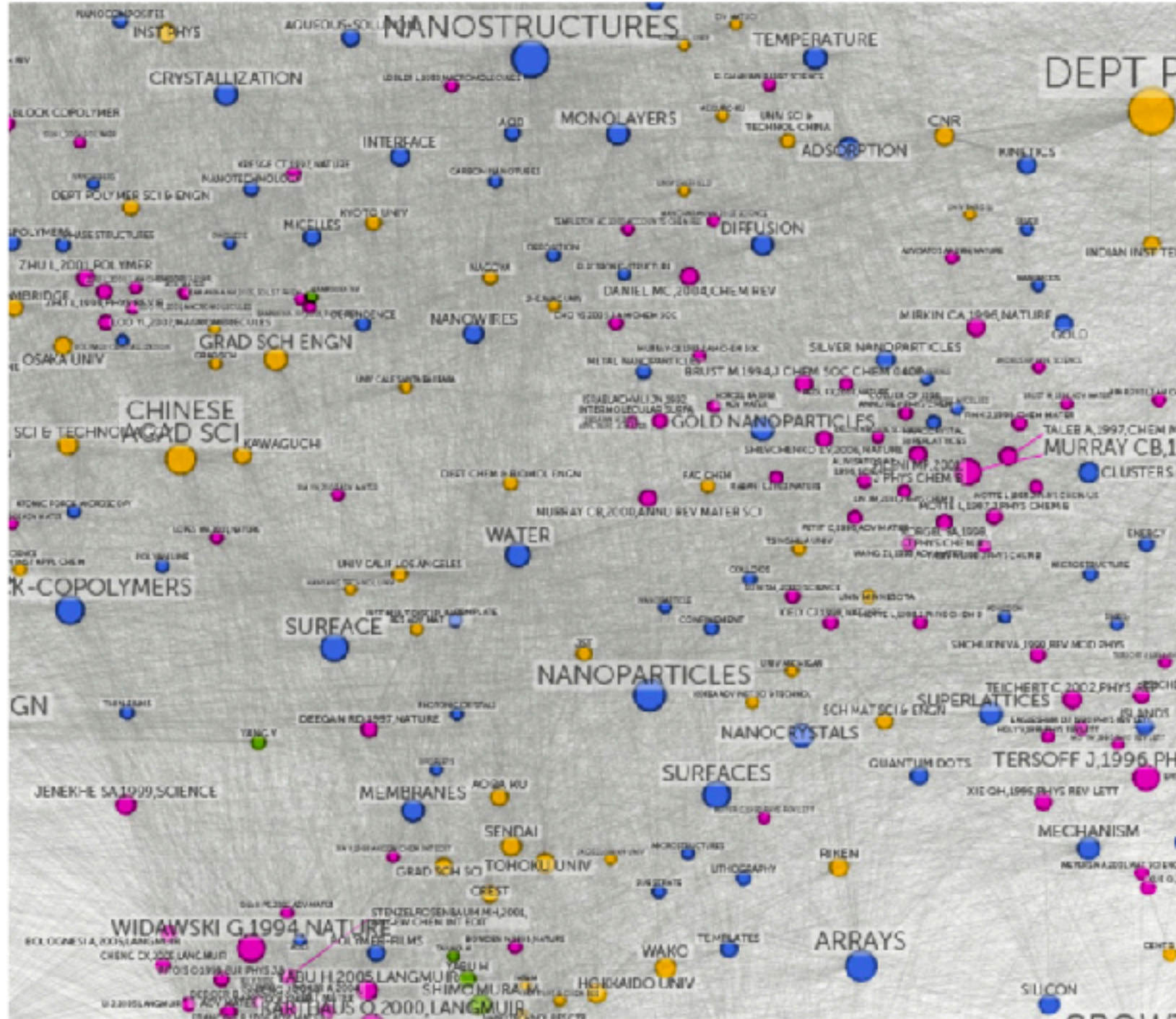


TRAJECTOIRES ARGUMENTATIVES ET CONSTELLATIONS DISCURSIVES

Réseaux, 2014/6 n° 188, p. 121-158

Naviguer dans les controverses

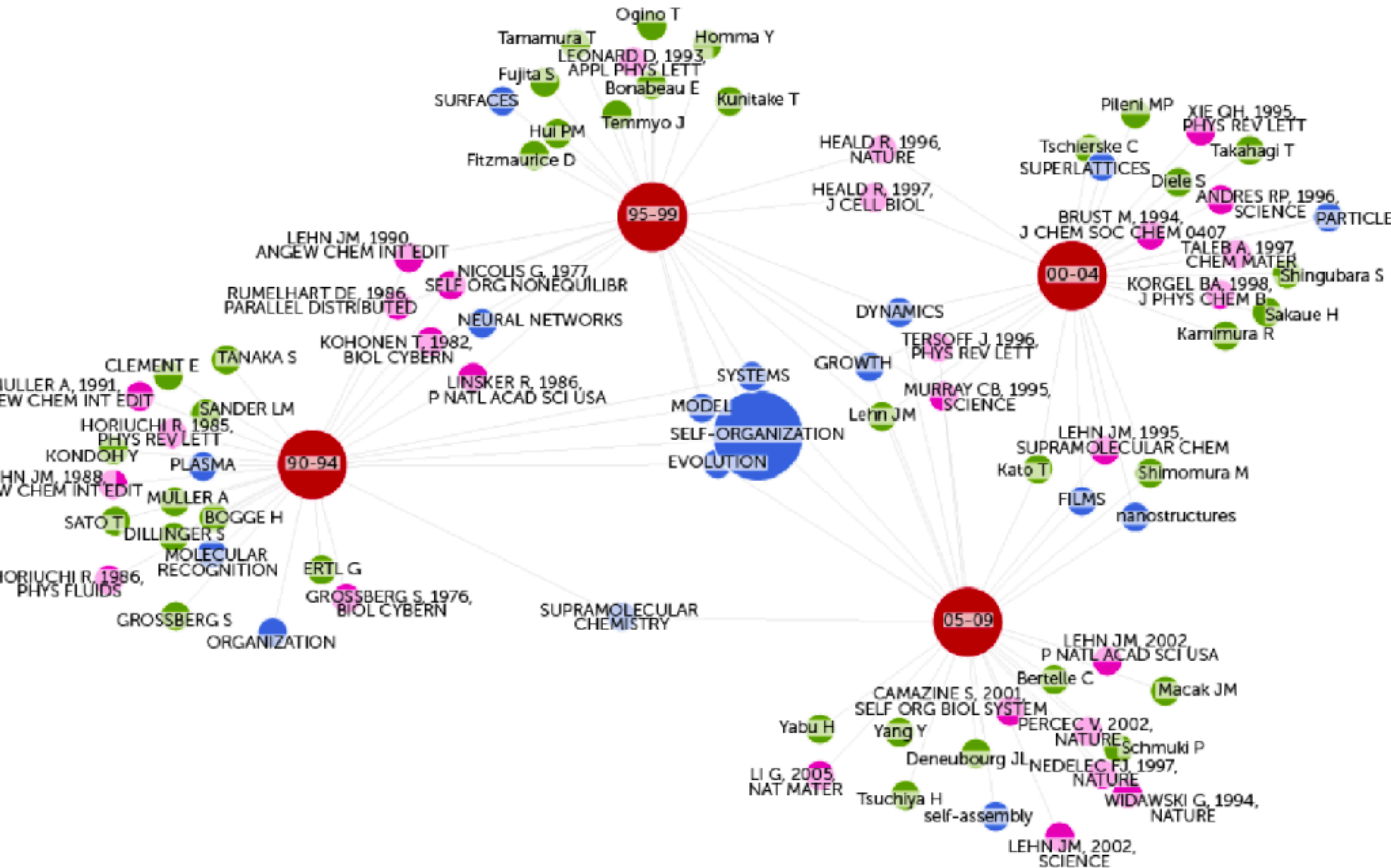
- 1LS - Un monde plat
- Hypothèse non atomiste mais monadique



The Whole is Always Smaller Than Its Parts A Digital Test of Gabriel Tarde's Monads, British Journal of Sociology, 2015

4a

PERIODS AUTHORS KEYWORDS ARTICLES



Understanding Participation

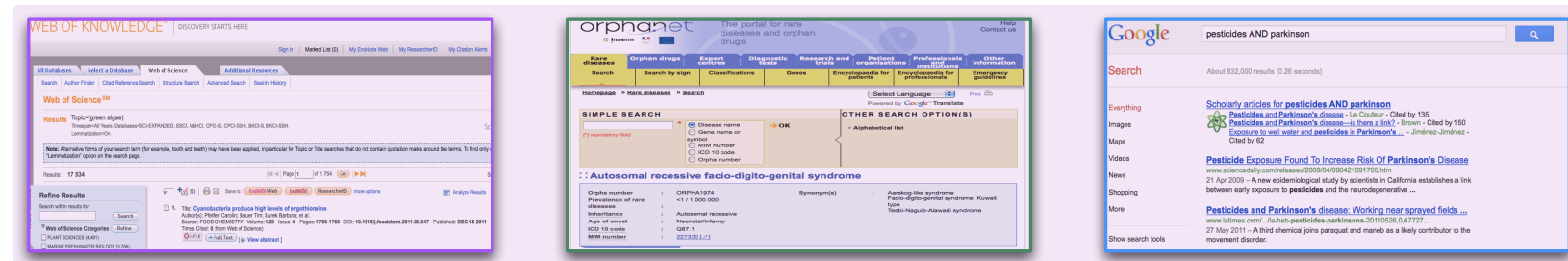


Bonus Track

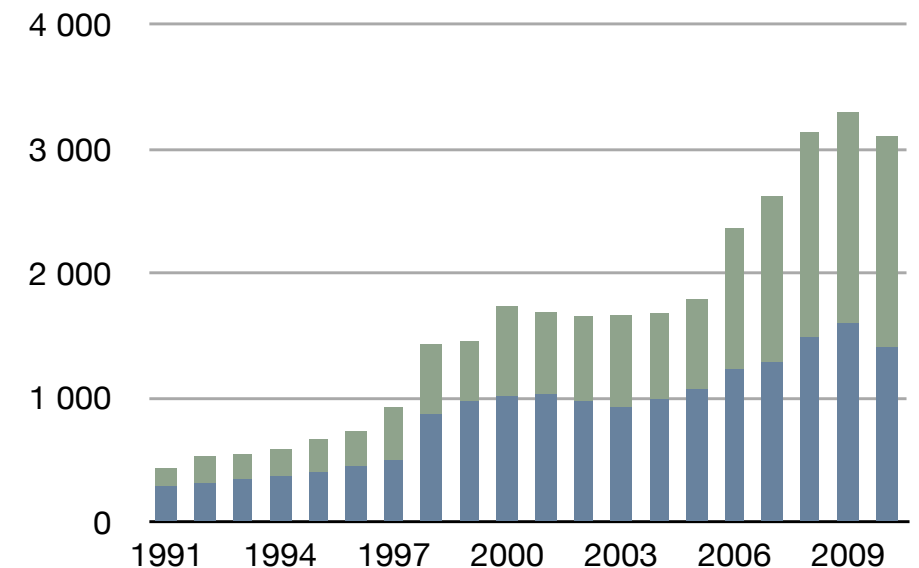
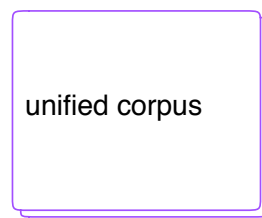
Various sources for different arenas

A project may demand to build an heterogeneous corpus: ISI Web of Science, Orphanet, Doctissimo.

- Capacity to build heterogeneous corpora coming from different data sources,



- Any dataset, whatever its structure, can be turned into a shared normalized database format,



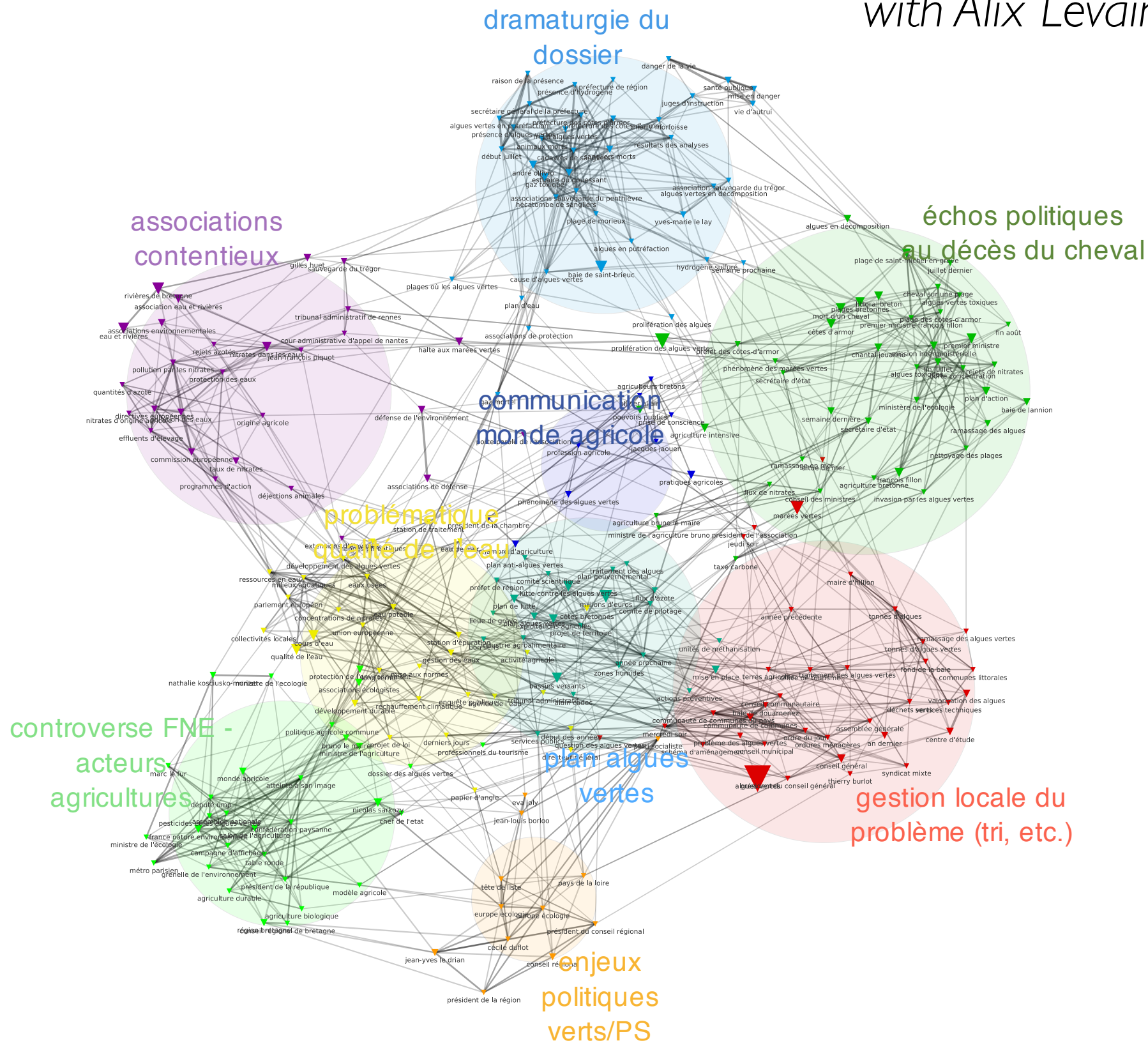
- versatile tools to run standardized analysis on potentially any kind of dataset.

Comparative analysis of knowledge production dynamics in different arenas

Green Algae in the media

with Alix Levain

- Co-existing **frames** in green algae press coverage

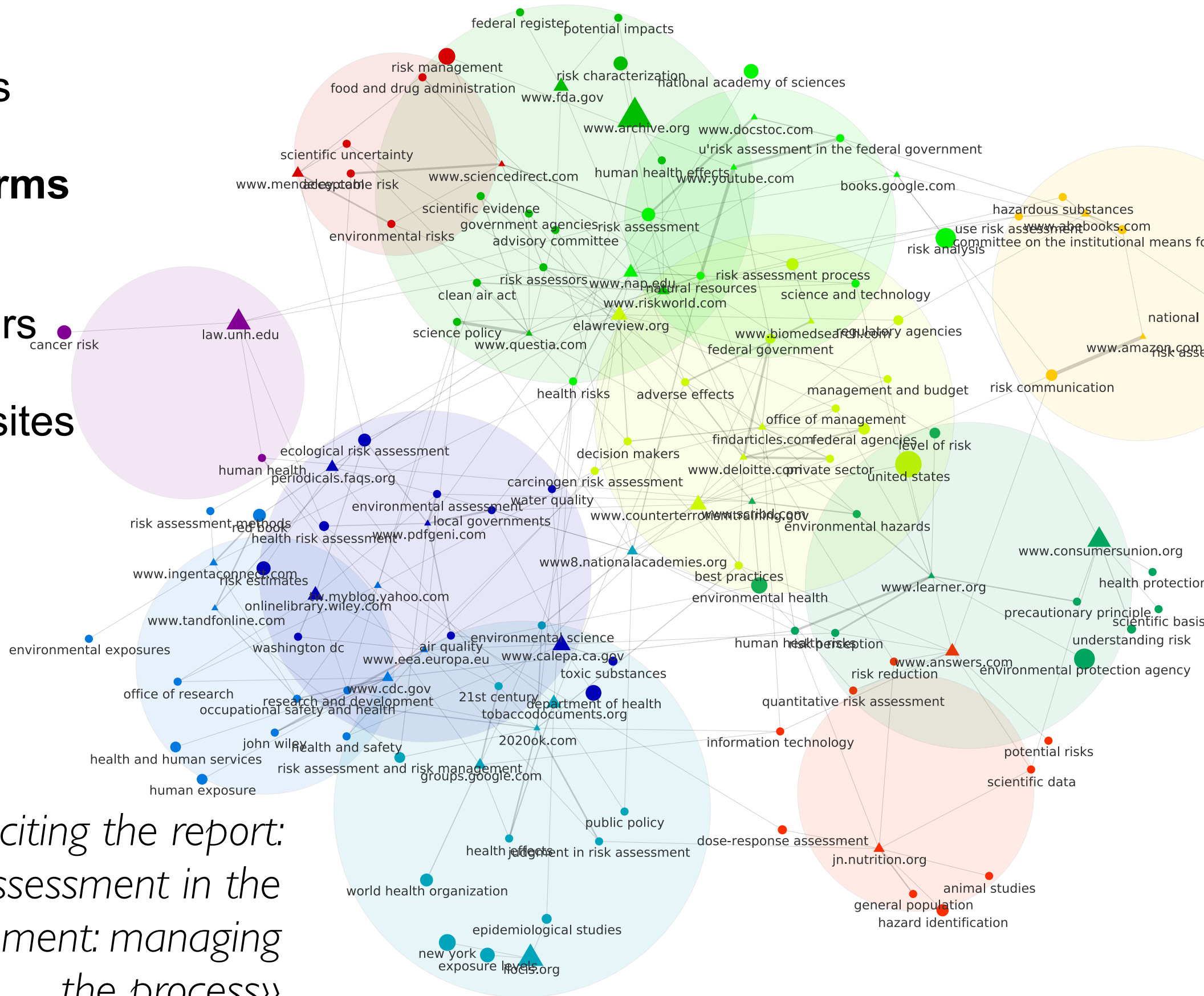


Main Topics around Green Algae Controversies

Heterogeneous Analysis

- Heterogeneous networks:
websites & terms

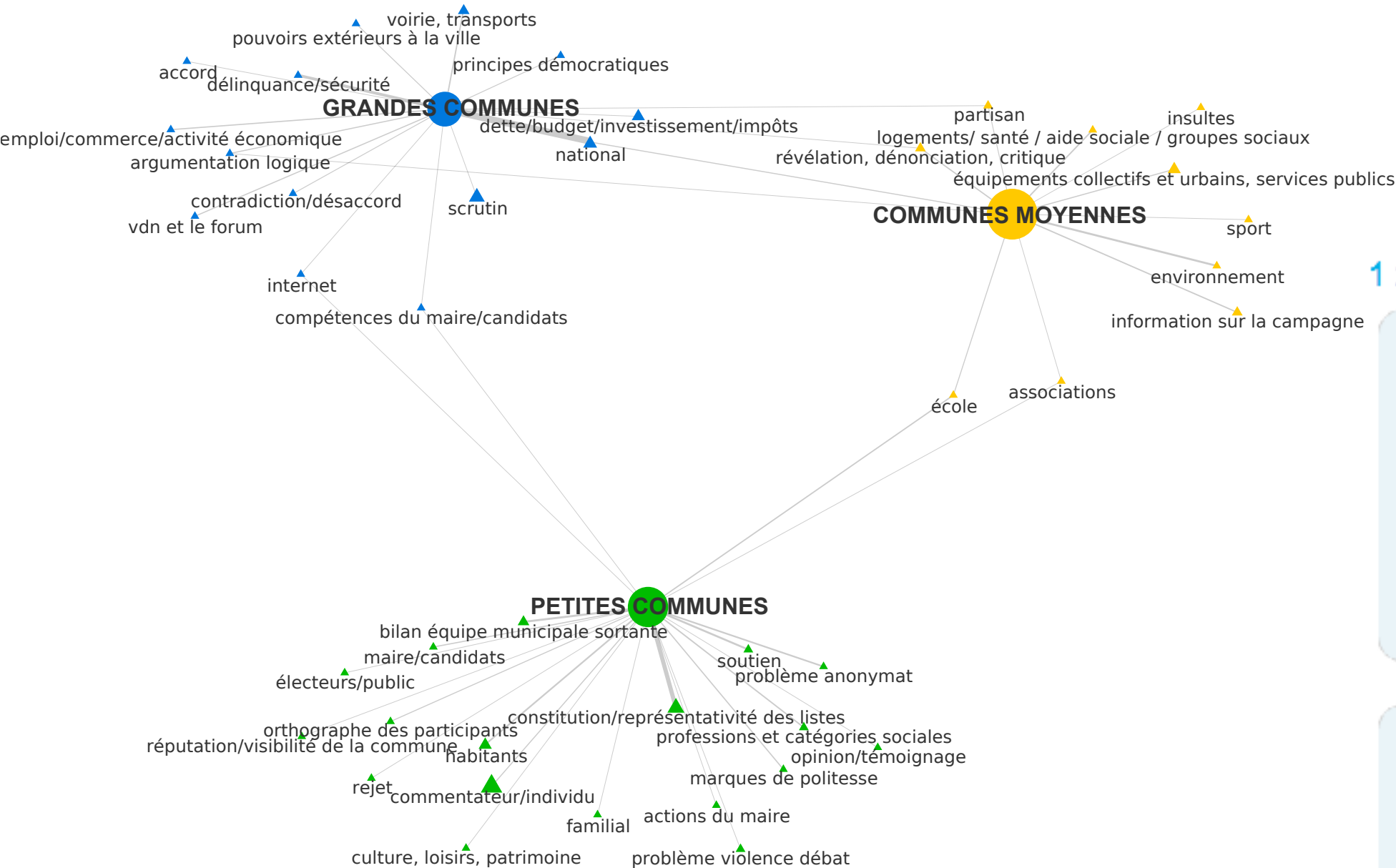
- Colored Clusters
gathers hybrid
groups of websites
and terms



web corpora citing the report:
«Risk assessment in the
federal government: managing
the process»

Heterogeneous Analysis

with Sylvain Parasié



1 268 COMMENTAIRES

1er commentaire

Monsieur Delattre est et continuera à être un excellent maire pour la ville de Coudkerque-Branche. Il n'y a qu'à voir son bilan sur 20 ans. D'une ville sinistrée industriellement il en a fait une ville active où il fait bon vivre et où toutes les friches qui étaient immenses sur l'ensemble de la commune disparaissent pour laisser place à des logements. Ce Monsieur s'est impliqué corps et âme pour sa commune. La majorité des instructions judiciaires dont il est victimes sont dues à des personnes à qui un moment ou un autre de leur vie ou carrière il a rendu service.

à Auxerre ils ont eu Guy Roux, nous on a André Delattre

Merci et Chapeau au plus ancien des élus socialistes du littoral.

D.Deuillat 12 septembre 2007 à 12:40

2ème commentaire

Où tout le monde est méchant envers lui : la justice, les journaux, la CUD (sauf pendant les élections), les services de l'État, etc.

Comme Guy Roux, il pense pouvoir tout faire, et au bout de 2 mois il laisse tomber Lens.

Il peut claquer la porte, mais cela fait trente ans que les revendications ont eu lieu, il serait temps d'en sortir !

Pas de livres pour les bibliothèques, moins de subventions aux assos, moins de transports, moins de culture, mais tout va bien, on a fait 80 logements on vous dit ! et Grande-Synthe et Saint-pol combien dans le même temps ?

vivement que ça change, et comme Guy Roux, qu'il raccroche ses baskets, il

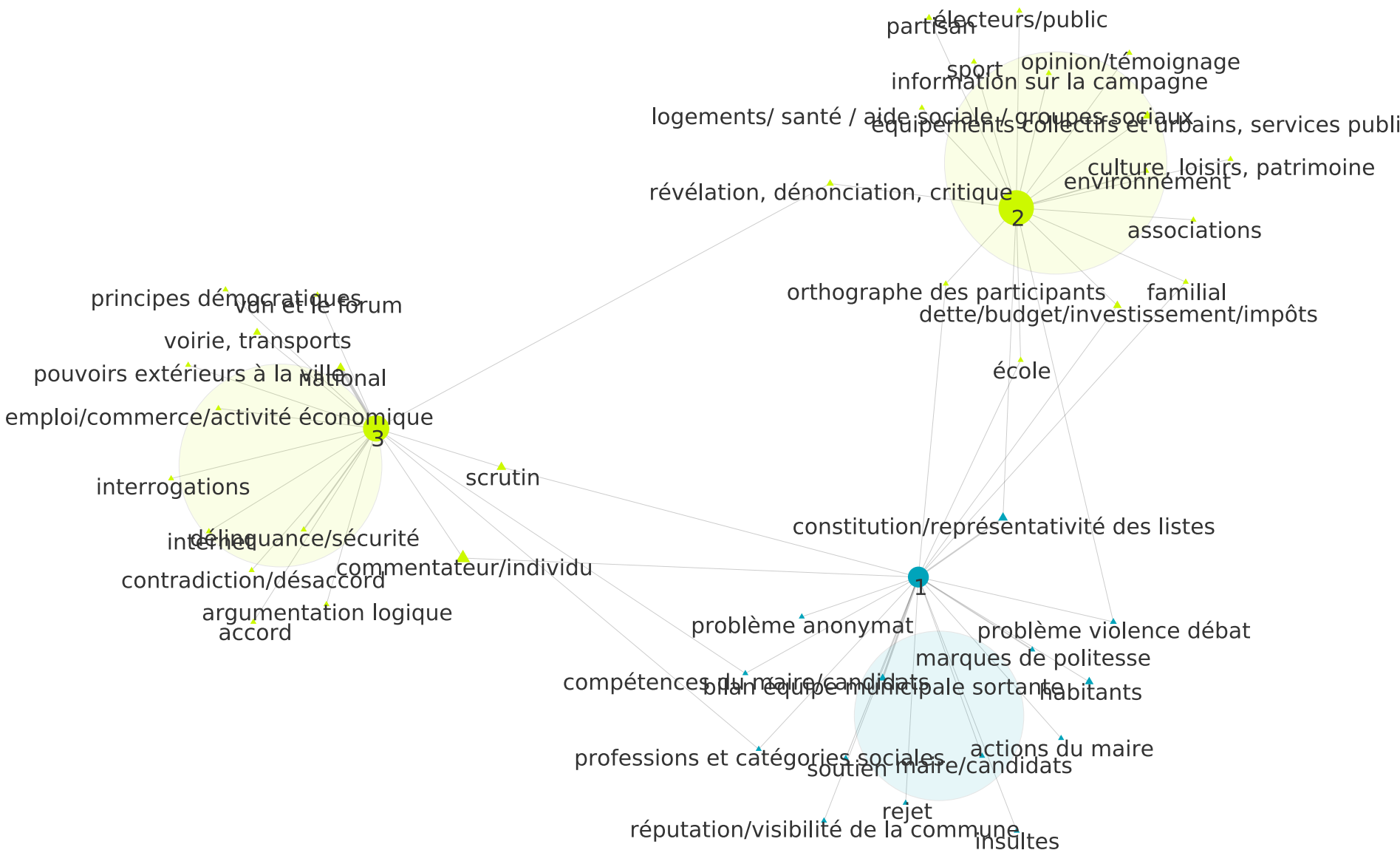
A.Delattre 12 septembre 2007 à 15:03

How socio-political spaces affect public expression in online discussion forums ?

city size and expression modes

Heterogeneous Analysis

5-203



1 268 COMMENTAIRES

1er commentaire

Monsieur Delattre est et continuera à être un excellent maire pour la ville de Coudkerque-Branché. Il n'y a qu'à voir son bilan sur 30 ans. D'une ville sinistrée industriellement il en a fait une ville active où il fait bon vivre et où toutes les friches qui étaient immenses sur l'ensemble de la commune disparaissent pour laisser place à des logements. Ce Monsieur s'est impliqué corps et âme pour sa commune. La majorité des instructions judiciaires dont il est victimes sont dues à des personnes à qui un moment ou un autre de leur vie ou carrière il a rendu service. à Auxerre ils ont eu Guy Roux, nous on a André Delattre. Merci et Chapeau au plus ancien des élus socialistes du littoral.

D.Deuillat 12 septembre 2007 à 12:48

2ème commentaire

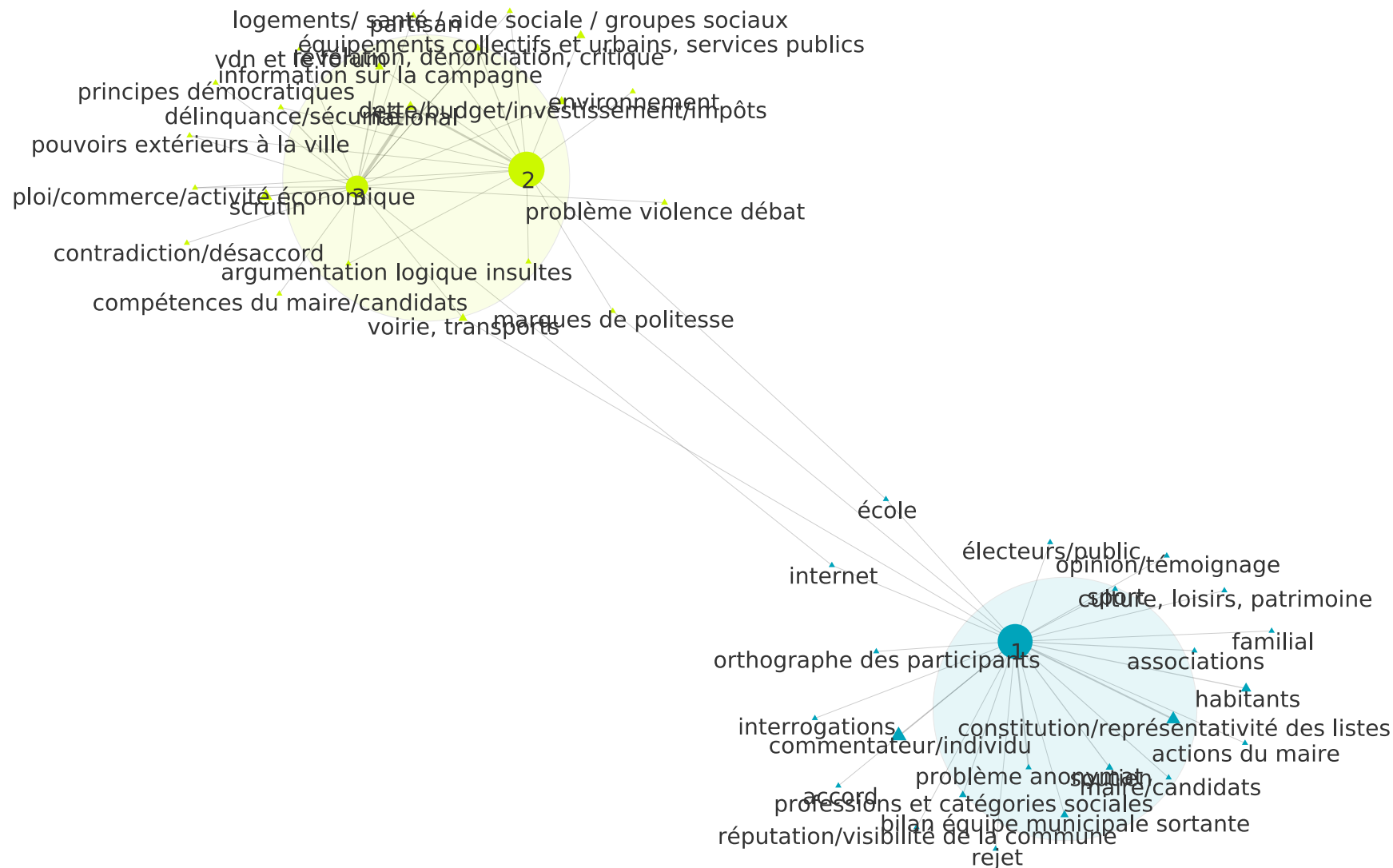
Où tout le monde est méchant envers lui : la justice, les journaux, la CUD (sauf pendant les élections), les services de l'État, etc. Comme Guy Roux, il pense pouvoir tout faire, et au bout de 2 mois il laisse tomber Lens. Il peut claquer le contrat, mais cela fait trente ans que les revendications ont eu lieu, il serait temps d'en sortir ! Pas de livres pour les bibliothèques, moins de subventions aux assos, moins de transports, moins de culture, mais tout va bien, on a fait 80 logements on vous dit ! et Grande-Synthe et Saint-paul combien dans le même temps ? vivement que ça change, et comme Guy Roux, qu'il raccroche ses baskets, il

A.Delattre 12 septembre 2007 à 15:03

1st period map

Heterogeneous Analysis

204-500



2nd period map



1 268 COMMENTAIRES

1er commentaire

Monsieur Delattre est et continuera à être un excellent maire pour la ville de Coudkerque-Branche. Il n'y a qu'à voir son bilan sur 30 ans. D'une ville sinistrée industriellement il en a fait une ville active où il fait bon vivre et où toutes les friches qui étaient immenses sur l'ensemble de la commune disparaissent pour laisser place à des logements. Ce Monsieur s'est impliqué corps et âme pour sa commune. La majorité des instructions judiciaires dont il est victimes sont dues à des personnes à qui un moment ou un autre de leur vie ou carrière il a rendu service. à Auxerre ils ont eu Guy Roux, nous on a André Delattre. Merci et Chapeau au plus ancien des élus socialistes du littoral.

D.Deuillet 12 septembre 2007 à 12:40

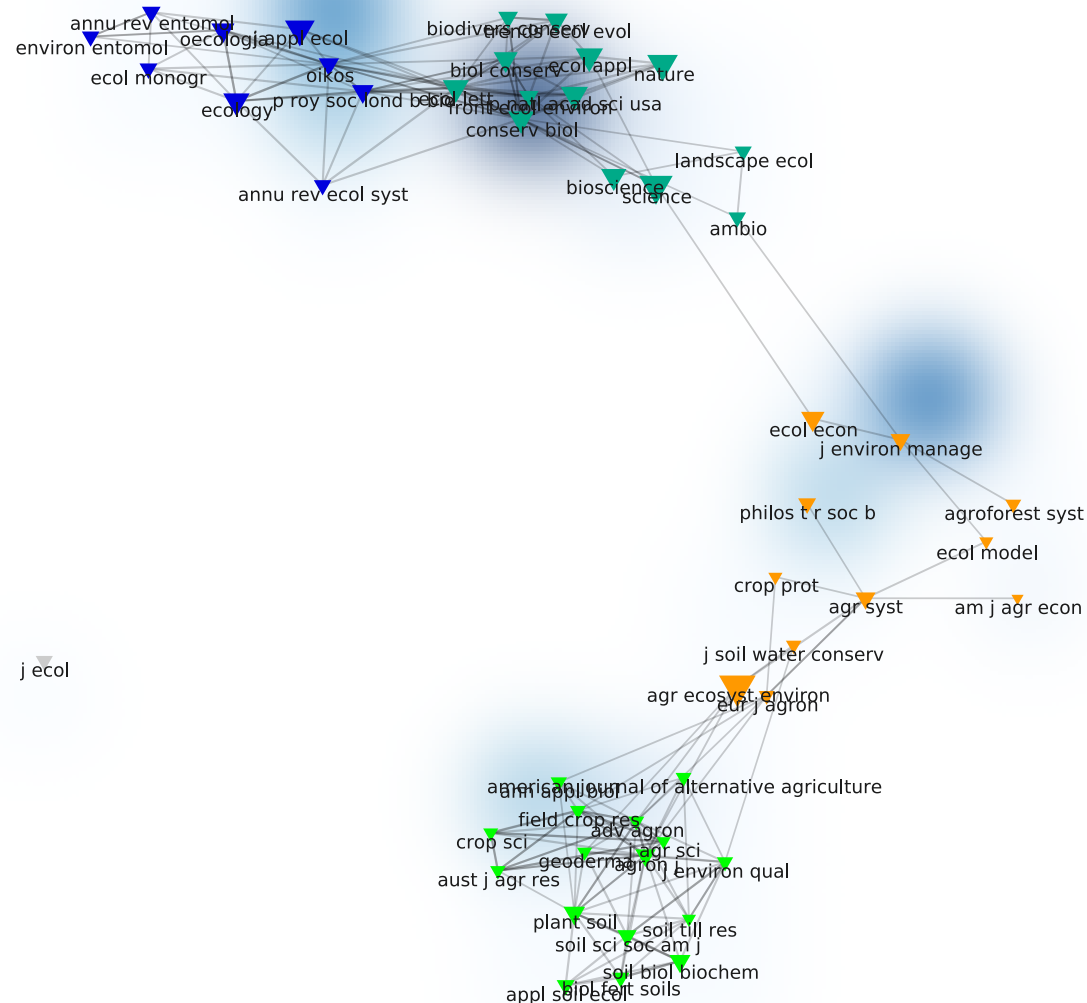
2ème commentaire

Où tout le monde est méchant envers lui : la justice, les journaux, la CUD (sauf pendant les élections), les services de l'État, etc. Comme Guy Roux, il pense pouvoir tout faire, et au bout de 2 mois il laisse tomber Lens. Il peut claquer le contrat, mais cela fait trente ans que les revendications ont eu lieu, il serait temps d'en sortir ! Pas de livres pour les bibliothèques, moins de subventions aux assos, moins de transports, moins de culture, mais tout va bien, on a fait 80 logements on vous dit ! et Grande-Synthe et Saint-paul combien dans le même temps ? vivement que ça change, et comme Guy Roux, qu'il raccroche ses baskets, il

A.Dalettire 12 septembre 2007 à 15:03

Heterogeneous Analysis

uk, 1975-2012

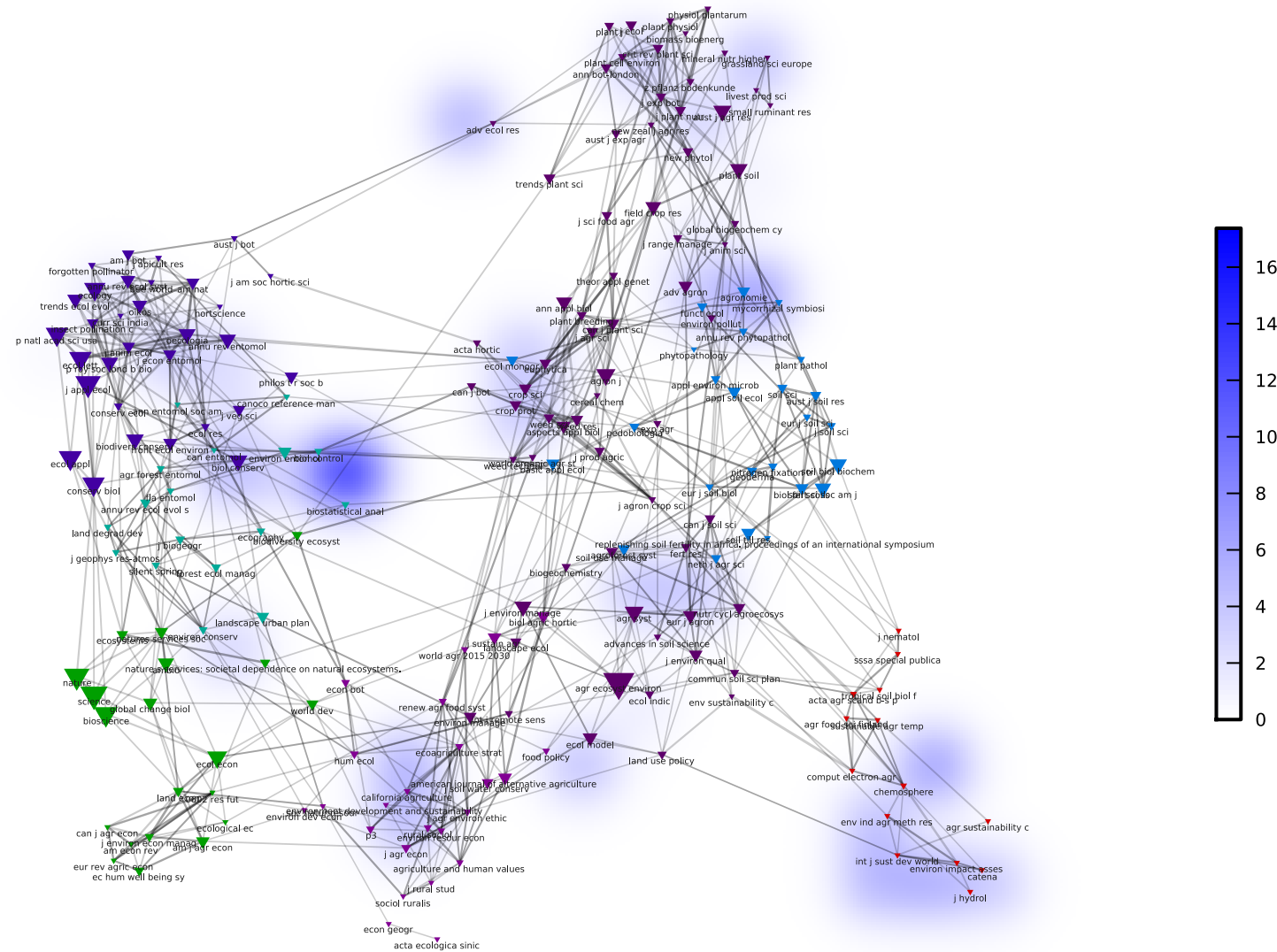


- Journal co-citation map
- Landscape Areas specific to british publications are highlighted in dark (heatmap projection)

Semantic map of synthetic Biology scientific publication (with Benjamin Raimbault)

Heterogeneous Analysis

france, 2005-2007 (2005-2006)

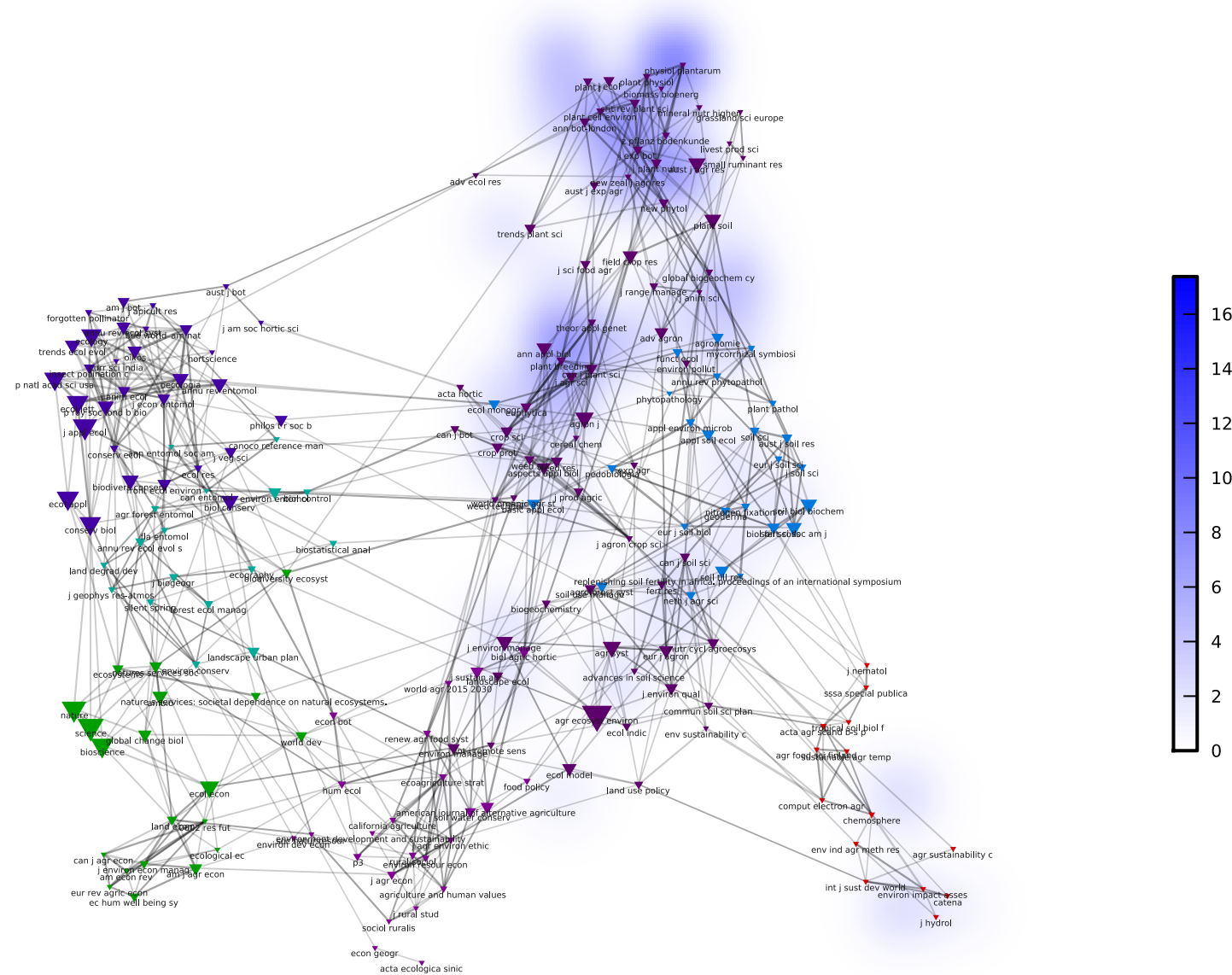


- Journal co-citation map
- Landscape Areas specific to british publications are highlighted in dark (heatmap projection)
- Dynamic profile of france background: semantic network

Semantic map of synthetic Biology scientific publication (with Benjamin Raimbault)

Heterogeneous Analysis

france, 2005-2007 (2007-2007)



- Journal co-citation map
- Landscape Areas specific to british publications are highlighted in dark (**heatmap** projection)
- Dynamic profile of france background: semantic network

Semantic map of synthetic Biology scientific publication (with Benjamin Raimbault)

Espaces Publics Numériques

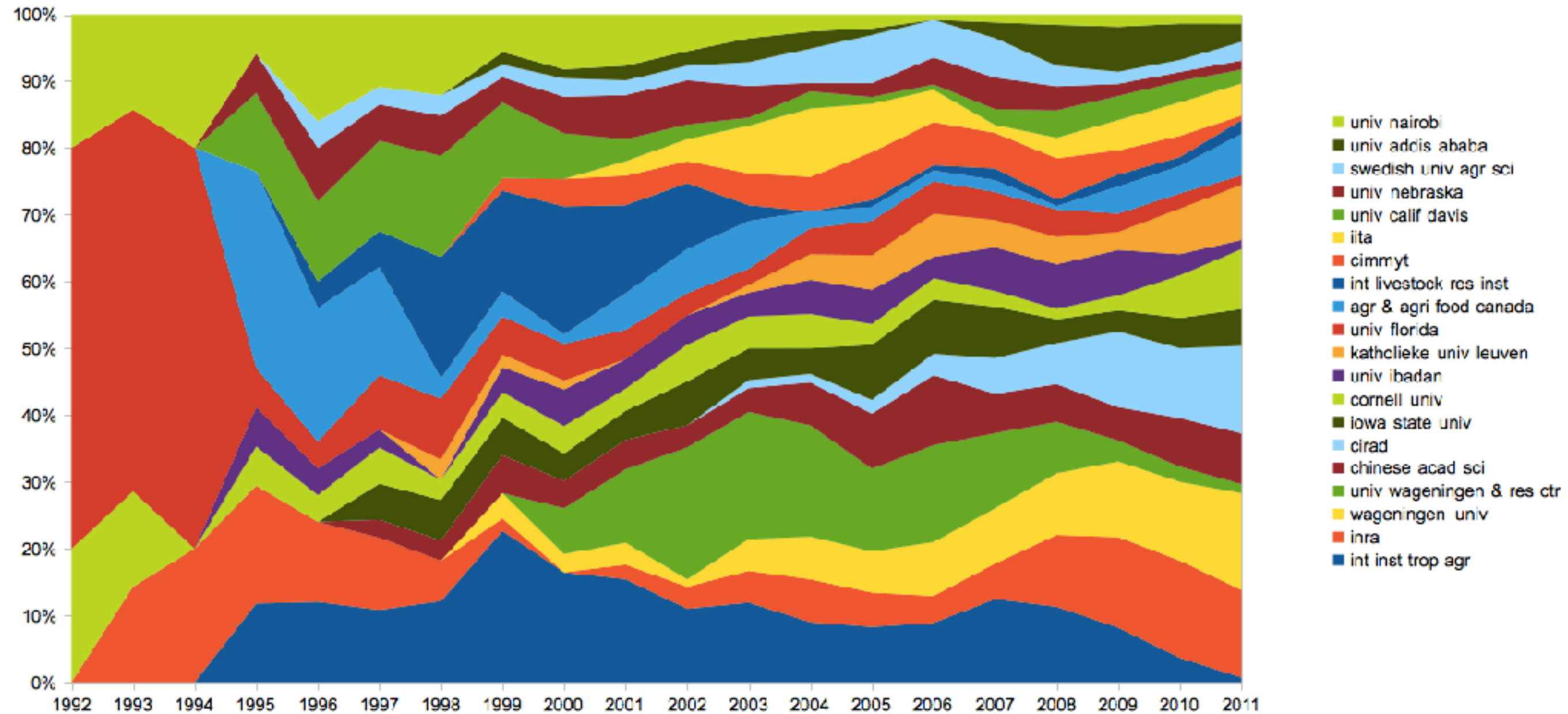


Structure hypertextuelle des liens
partagés sur Facebook et
Dynamique d'un problème public sur

Other Visualizations

Simple Dynamics

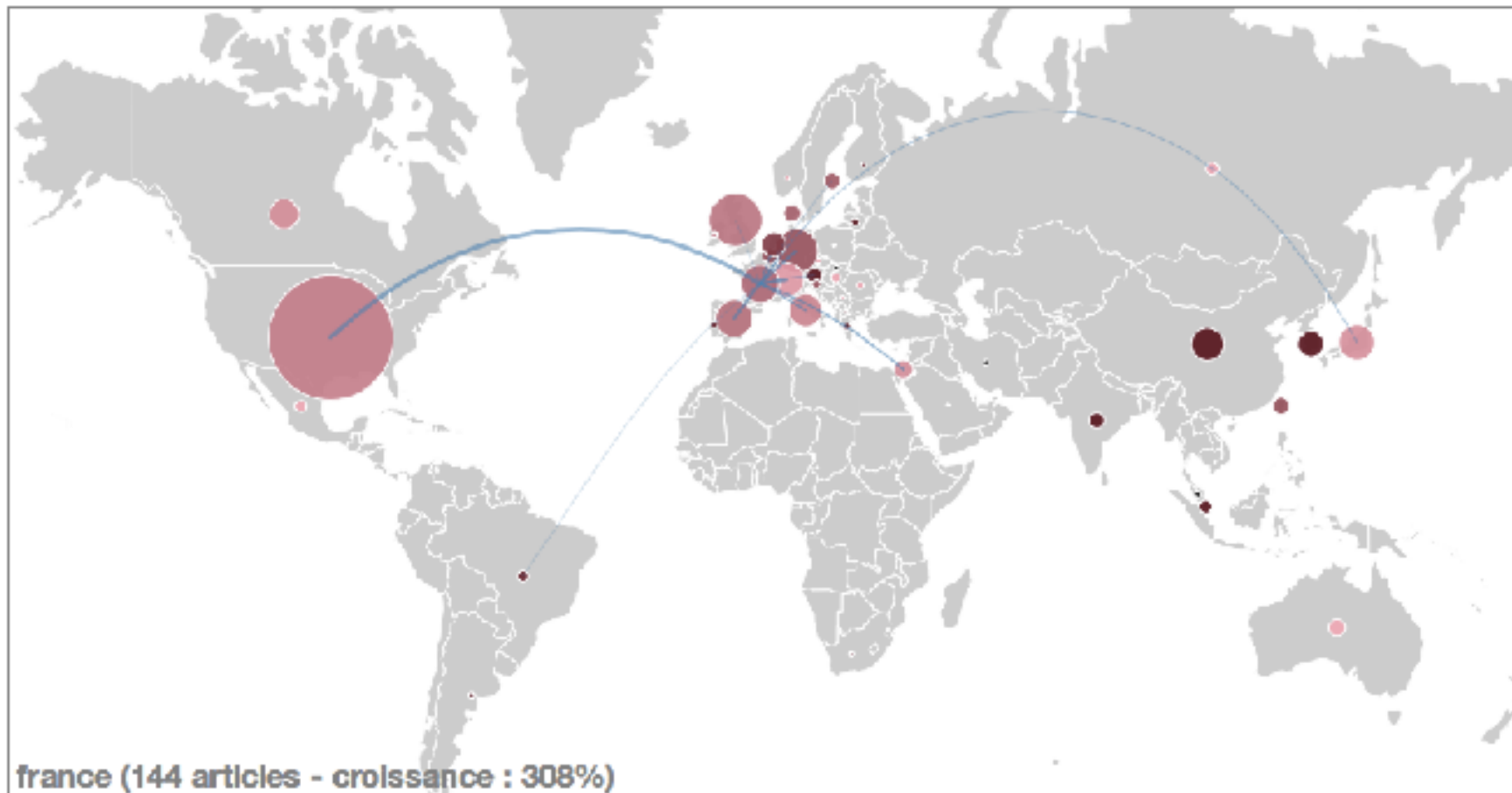
- Time Series Analysis



Research Institutions involved in Agro-Ecology Research

Heterogeneous Analysis

La Biologie de Synthèse dans le Monde



La carte ci-dessus présente une vue d'ensemble de la littérature scientifique en biologie de synthèse à l'échelle mondiale. Les pays dans lesquels sont publiés des articles du domaine (présent dans notre corpus) se sont vus attribuer un cercle dont la surface est proportionnelle au nombre de publications produites. La couleur des cercles correspond au taux de croissance du nombre de publications: du plus clair (moins de 200% de croissance annuelle) au plus foncé (+ de 500 %). En survolant un pays avec le curseur, les collaborations avec d'autres pays apparaissent sous forme de liens dont l'épaisseur est proportionnelle au nombre d'articles co-publiés par des auteurs des deux nationalités.

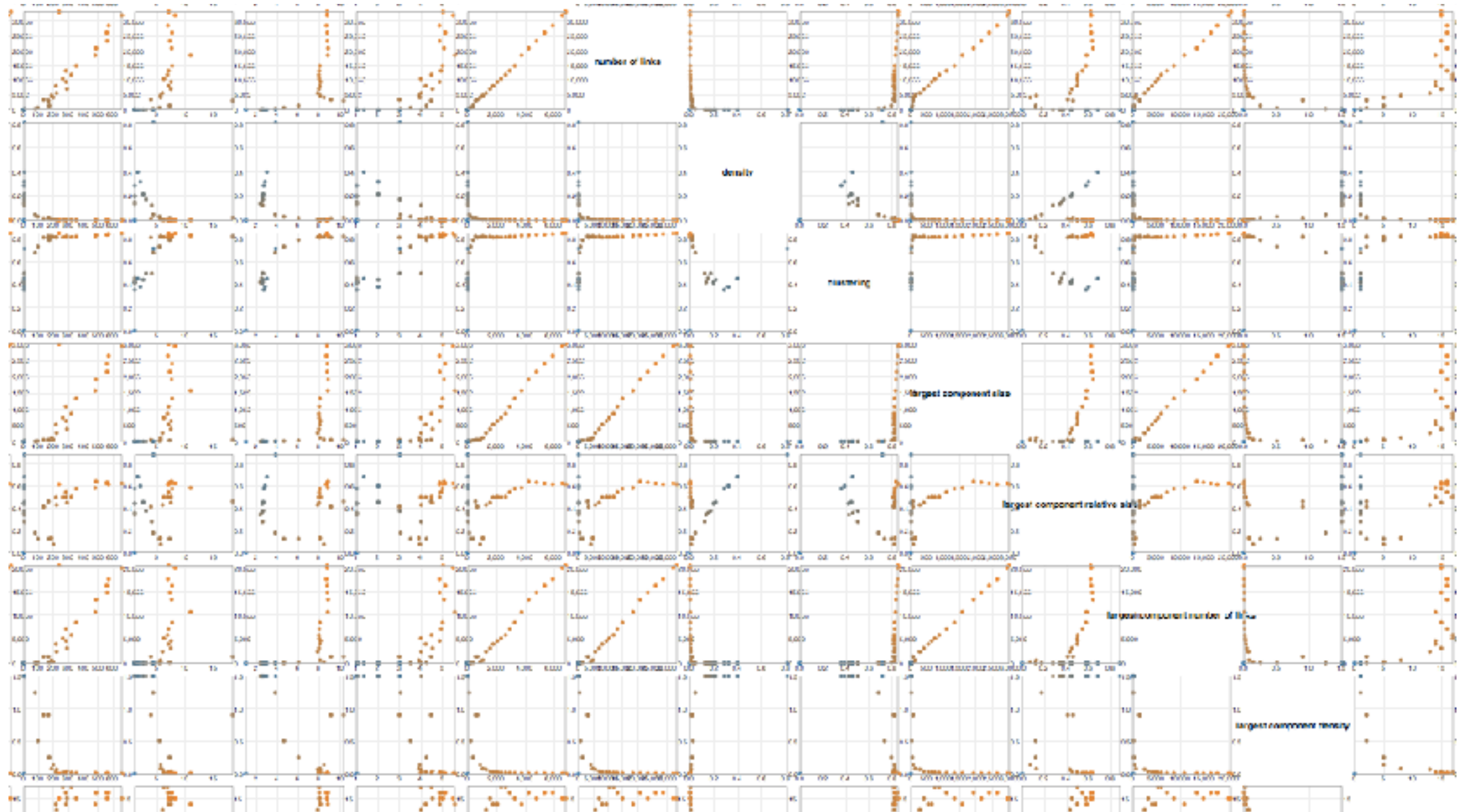
Le corpus de publications considéré comprend l'ensemble des articles de la base de données Thomson Reuters Web of Science mentionnant le terme "synthetic biology" augmenté par une extension lexicale ("synthetic gene networks", "Standard Biological Parts", "artificial genetic system", "synthetic

- Journal co-citation map
- Landscape Areas specific to british publications are highlighted in dark (**heatmap** projection)
- Dynamic profile of france background: semantic network

Semantic map of synthetic Biology scientific publication (with Benjamin Rimbault)

Evolving Networks

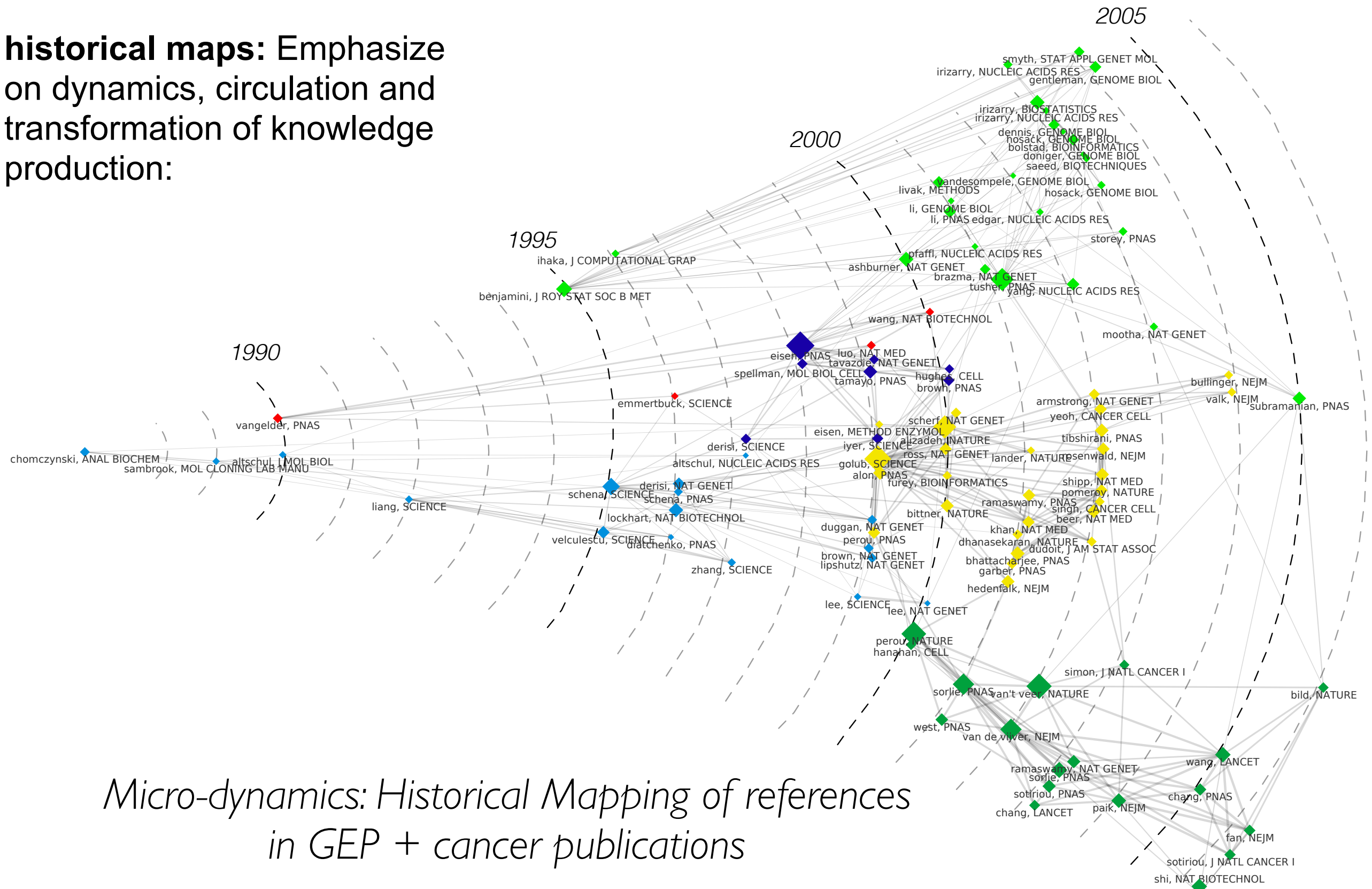
- Detection of **Phase Transition** events in dynamical network



dynamical collaboration network

Historical Mapping

- **historical maps:** Emphasize on dynamics, circulation and transformation of knowledge production:



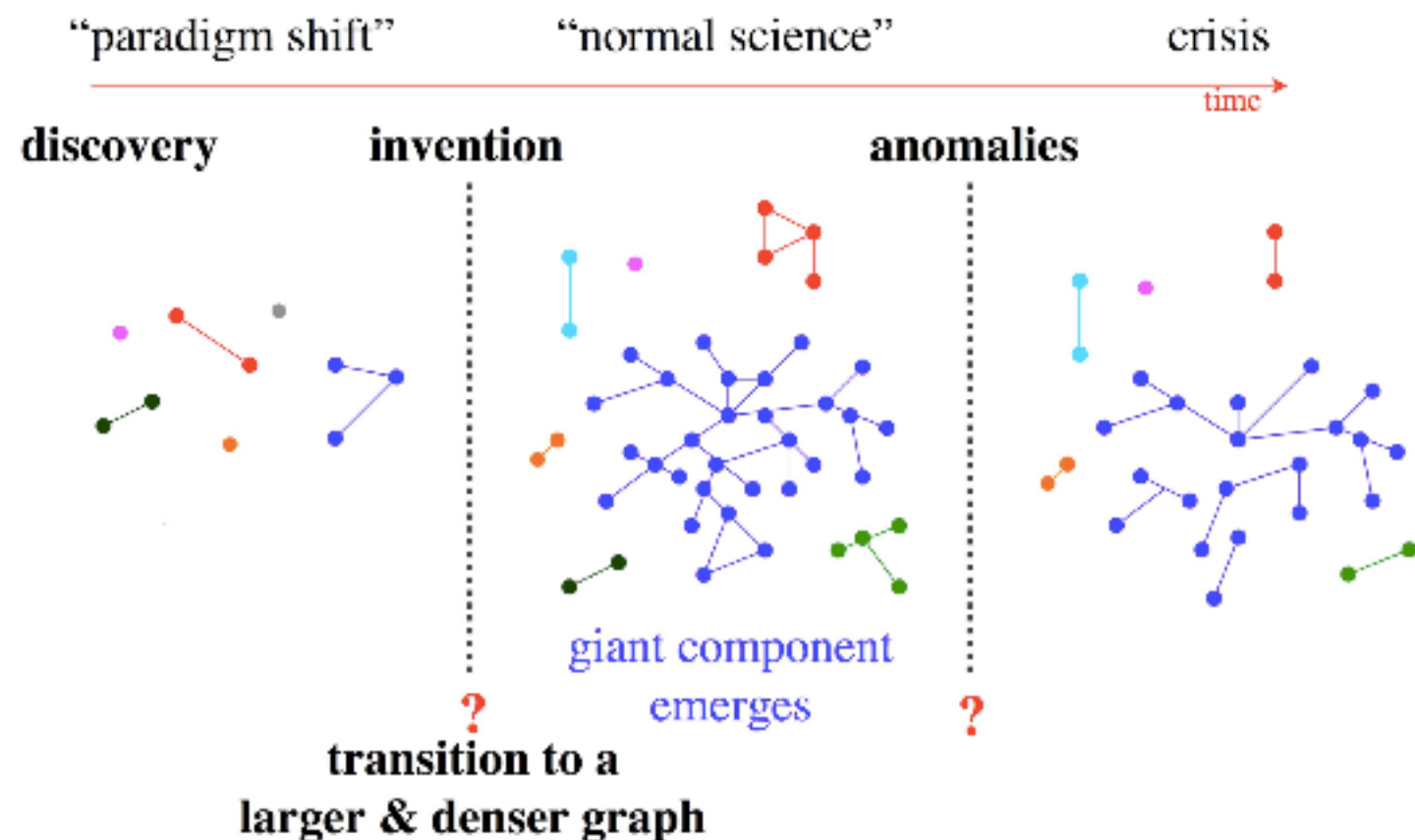
*Micro-dynamics: Historical Mapping of references
in GEP + cancer publications*

Textual Analysis

heterogeneous multi-level networks

Bettencourt, Luís, David I. Kaiser, and Jasleen Kaur. "Scientific discovery and topological transitions in collaboration networks." *Journal of Informetrics* 3.3 (2009): 210-221.

- Generalized co-occurrences analysis framework mixing people, terms, countries, etc...
- Clustering techniques are being used to circulate from micro to macro levels - clusters are made of possibly heterogeneous nodes associated in a singular manner.
- Less pertinent links should be filtered for clusters to be revealed but global structure should be preserved

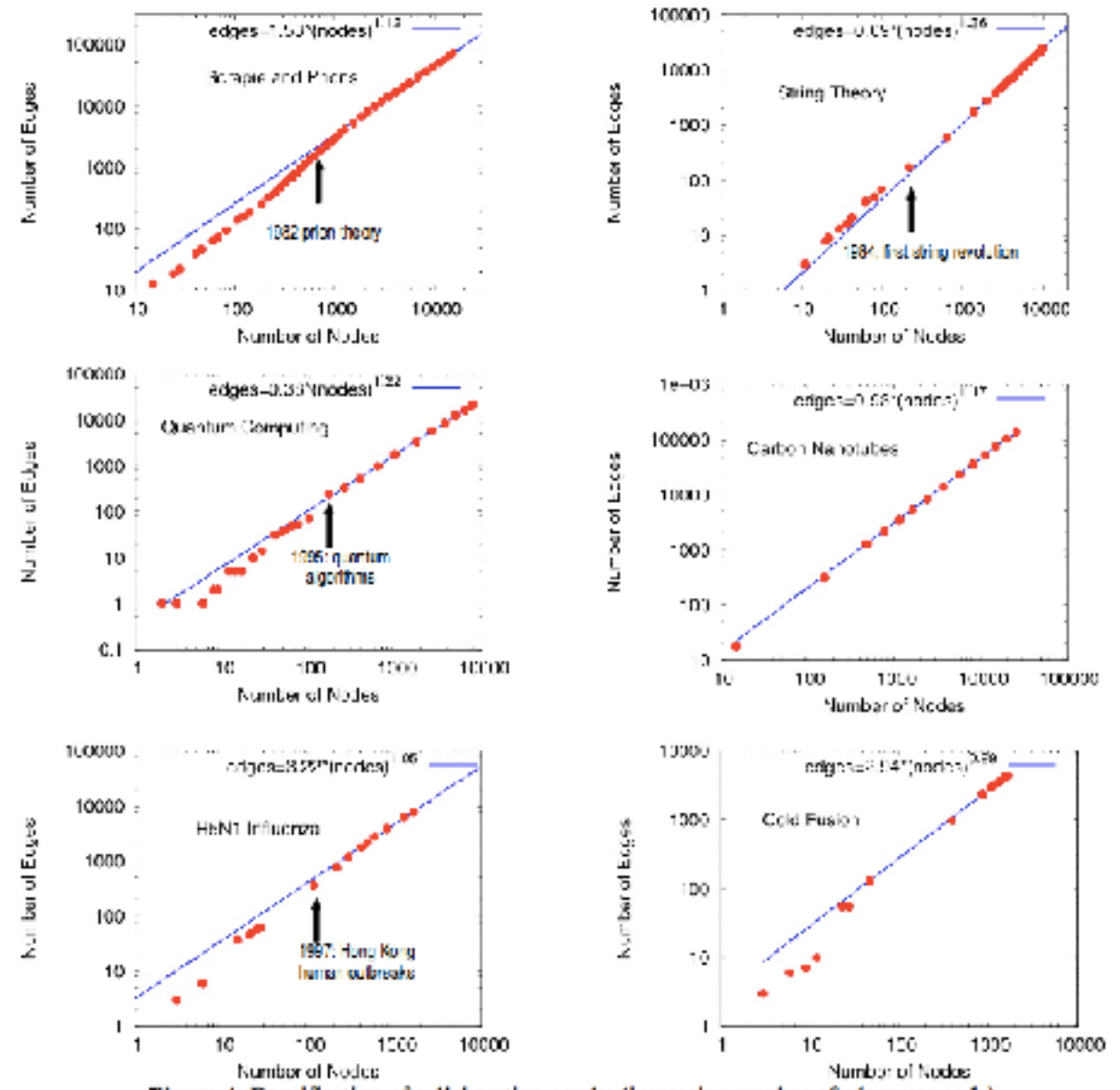


Phase Transitions in collaboration network reveal the different regimes

heterogeneous multi-level networks

Bettencourt, Luís, David I. Kaiser, and Jasleen Kaur. "Scientific discovery and topological transitions in collaboration networks." *Journal of Informetrics* 3.3 (2009): 210-221.

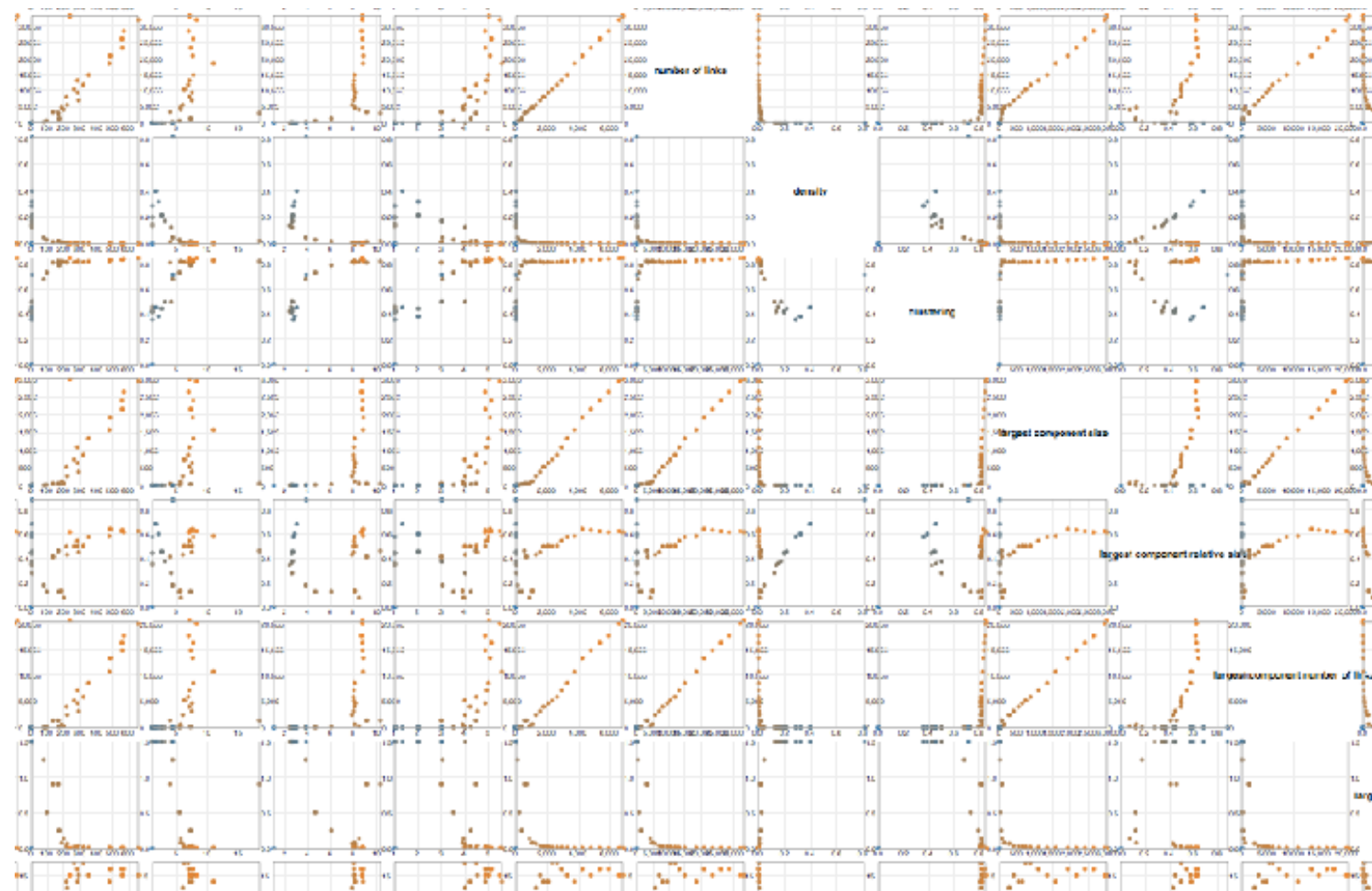
- Generalized co-occurrences analysis framework mixing people, terms, countries, etc...
- Clustering techniques are being used to circulate from micro to macro levels - clusters are made of possibly heterogeneous nodes associated in a singular manner.
- Less pertinent links should be filtered for clusters to be revealed but global structure should be preserved



collaboration networks densify at a faster pace in active fields

heterogeneous multi-level networks

- Generalized co-occurrences analysis framework mixing people, terms, countries, etc...
- Clustering techniques are being used to circulate from micro to macro levels - clusters are made of possibly heterogeneous nodes associated in a singular manner.
- Less pertinent links should be filtered for clusters to be revealed but global structure should be preserved

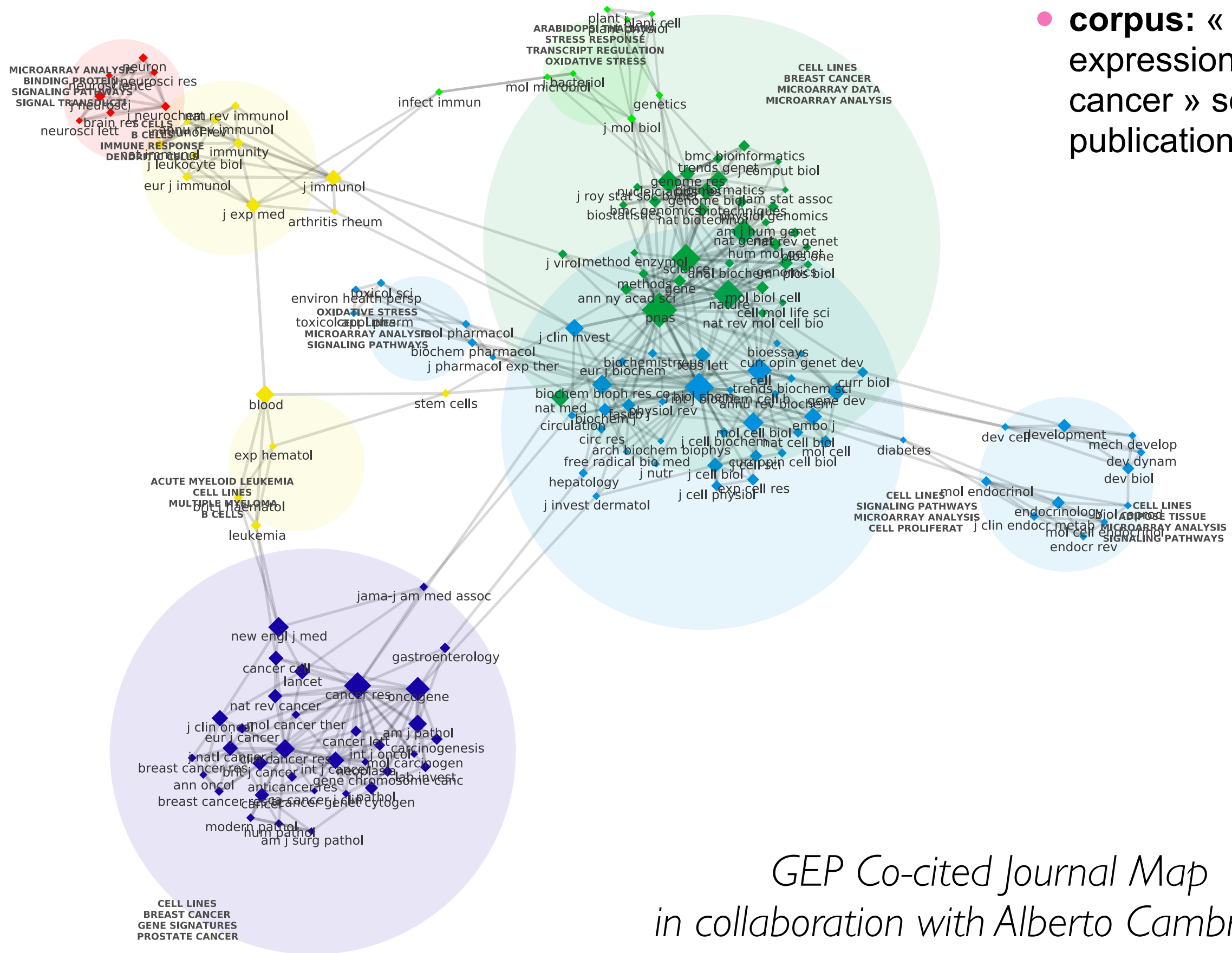


CorText Manager - structural analysis

From heterogeneous structure to dynamics

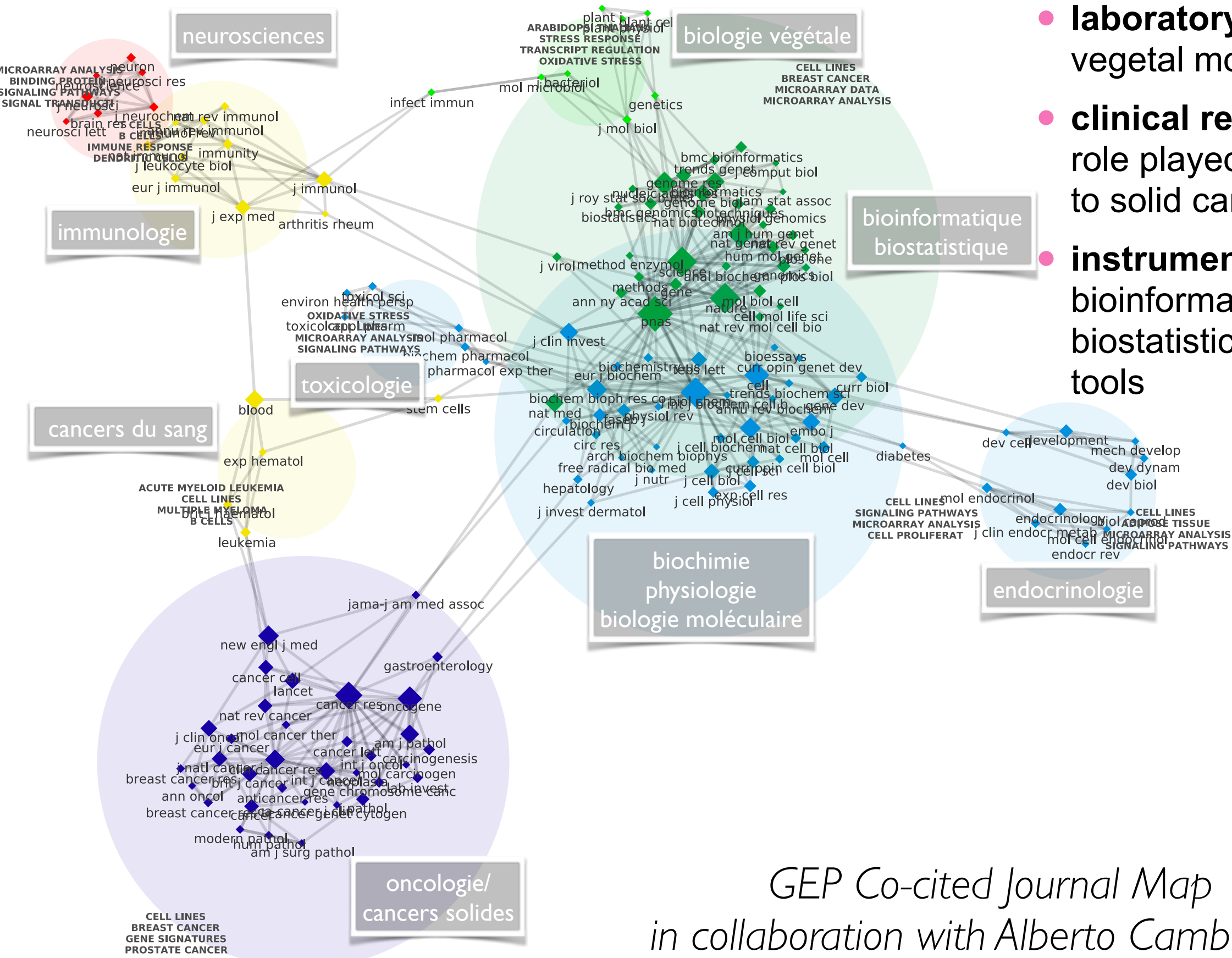
Translational Research Triangle

- **corpus:** « Gene expression profile & cancer » scientific publications



*GEP Co-cited Journal Map
in collaboration with Alberto Cambrosio*

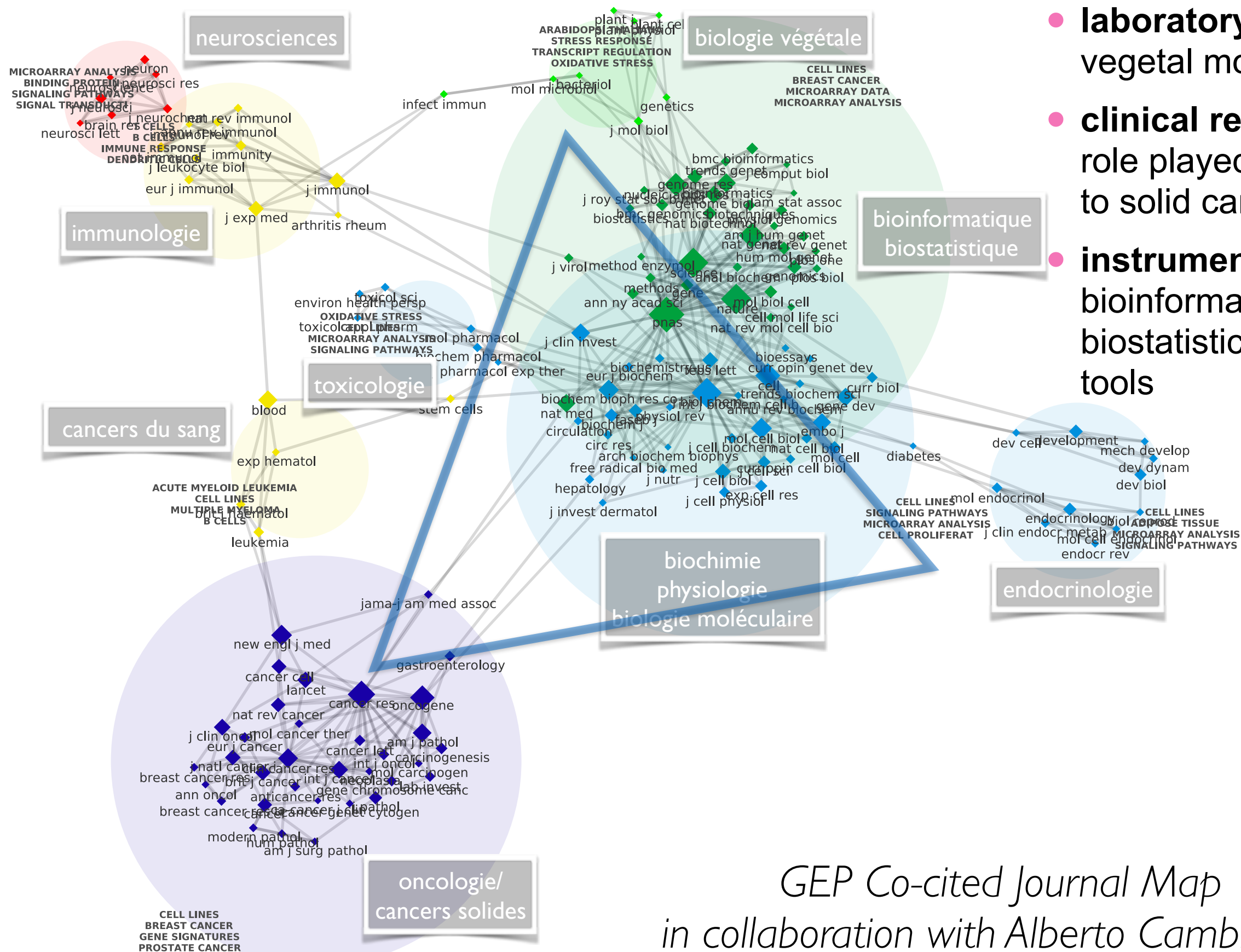
Translational Research Triangle



- **laboratory research** - vegetal model organism
- **clinical research** - central role played by application to solid cancer
- **instrumental research** - bioinformatics and biostatistics as necessary tools

*GEP Co-cited Journal Map
in collaboration with Alberto Cambrosio*

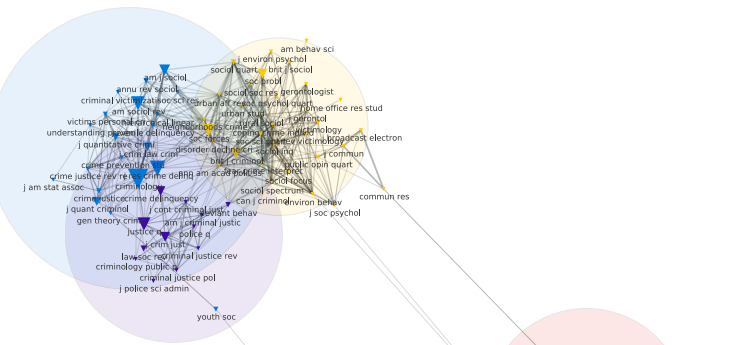
Translational Research Triangle



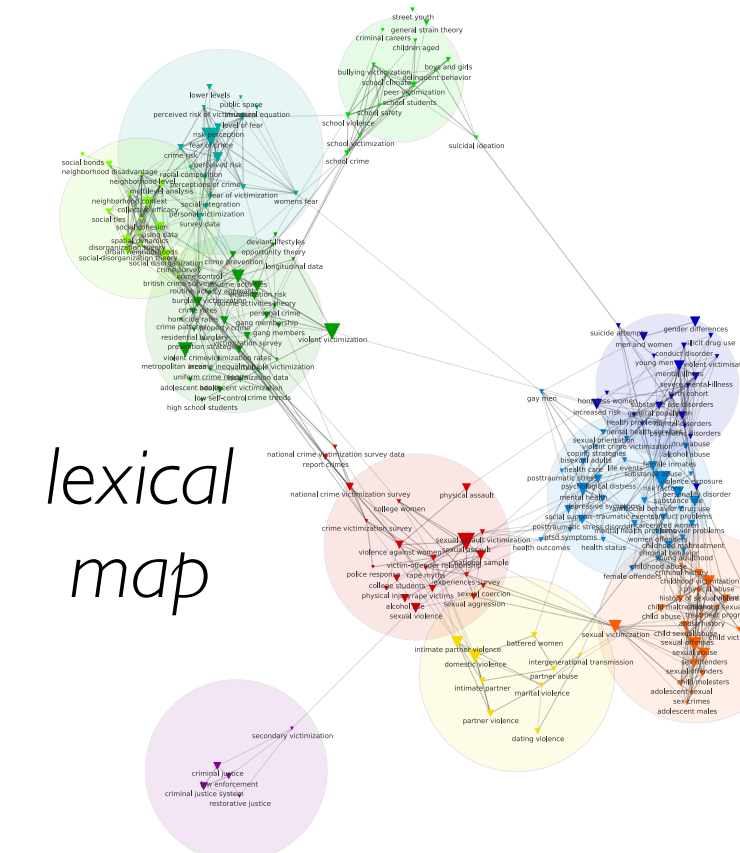
- **laboratory research** - vegetal model organism
- **clinical research** - central role played by application to solid cancer
- **instrumental research** - bioinformatics and biostatistics as necessary tools

*GEP Co-cited Journal Map
in collaboration with Alberto Cambrosio*

Mixing dimensions of analysis



1975-2013



victim-offender relationship & sexual assault

criminal justice system & law enforcement

social disorganization & disorganization theory

fear of crime & perceived risk

routine activities & property crime

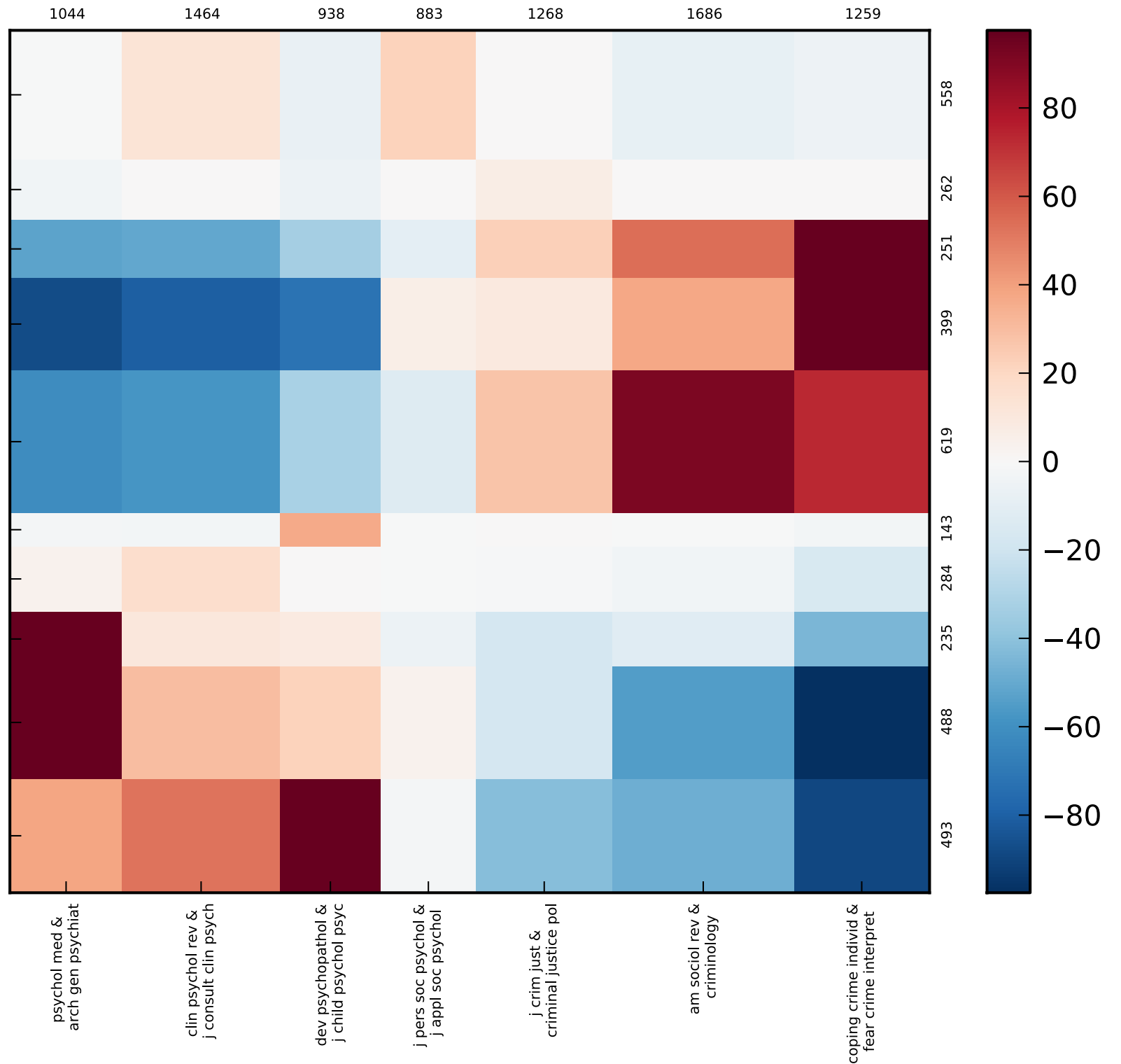
school students & peer victimization

intimate partner violence & domestic violence

mental illness & birth cohort

mental health & substance abuse

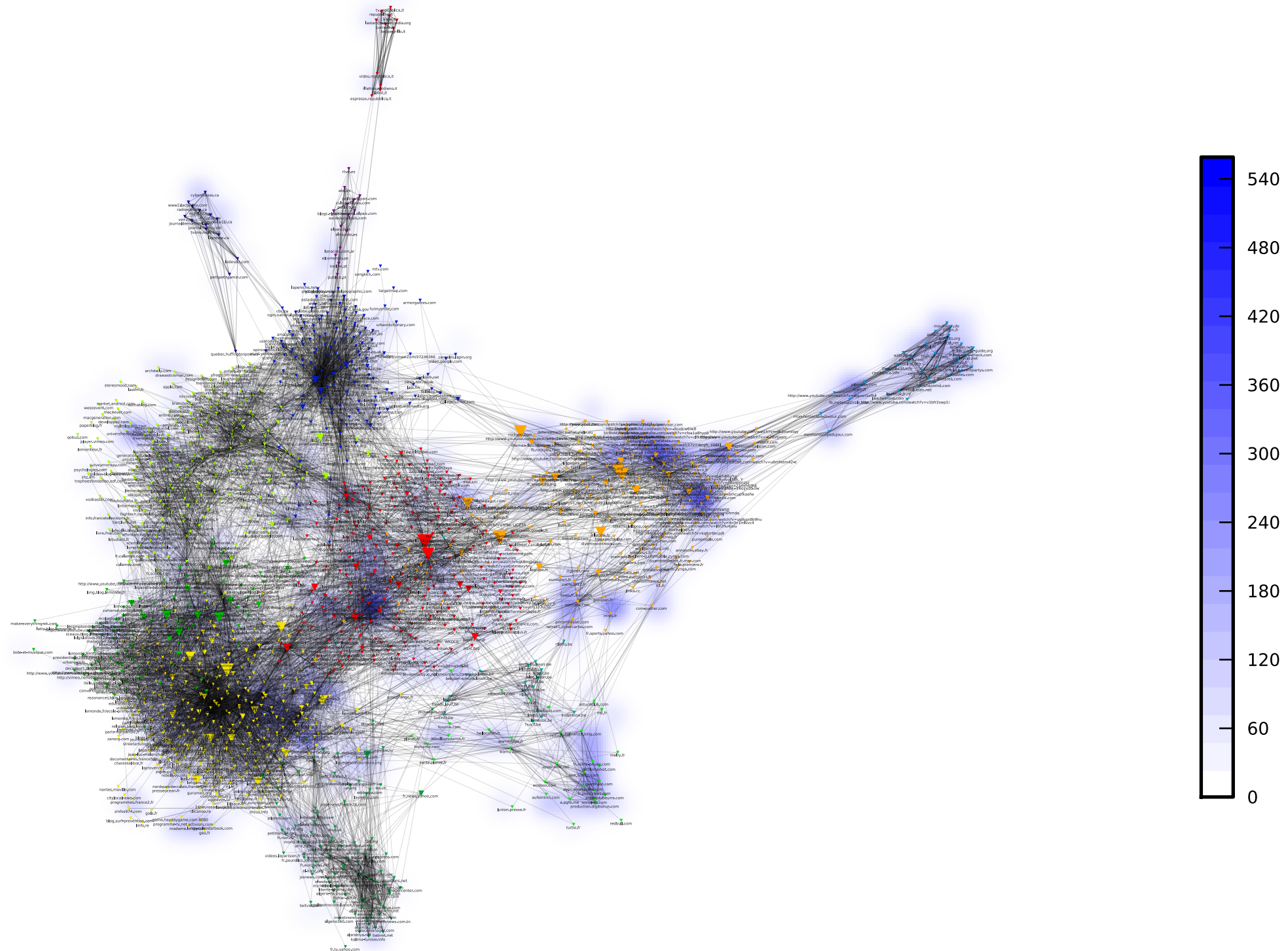
sexual abuse & sex offenders



crime science corpus

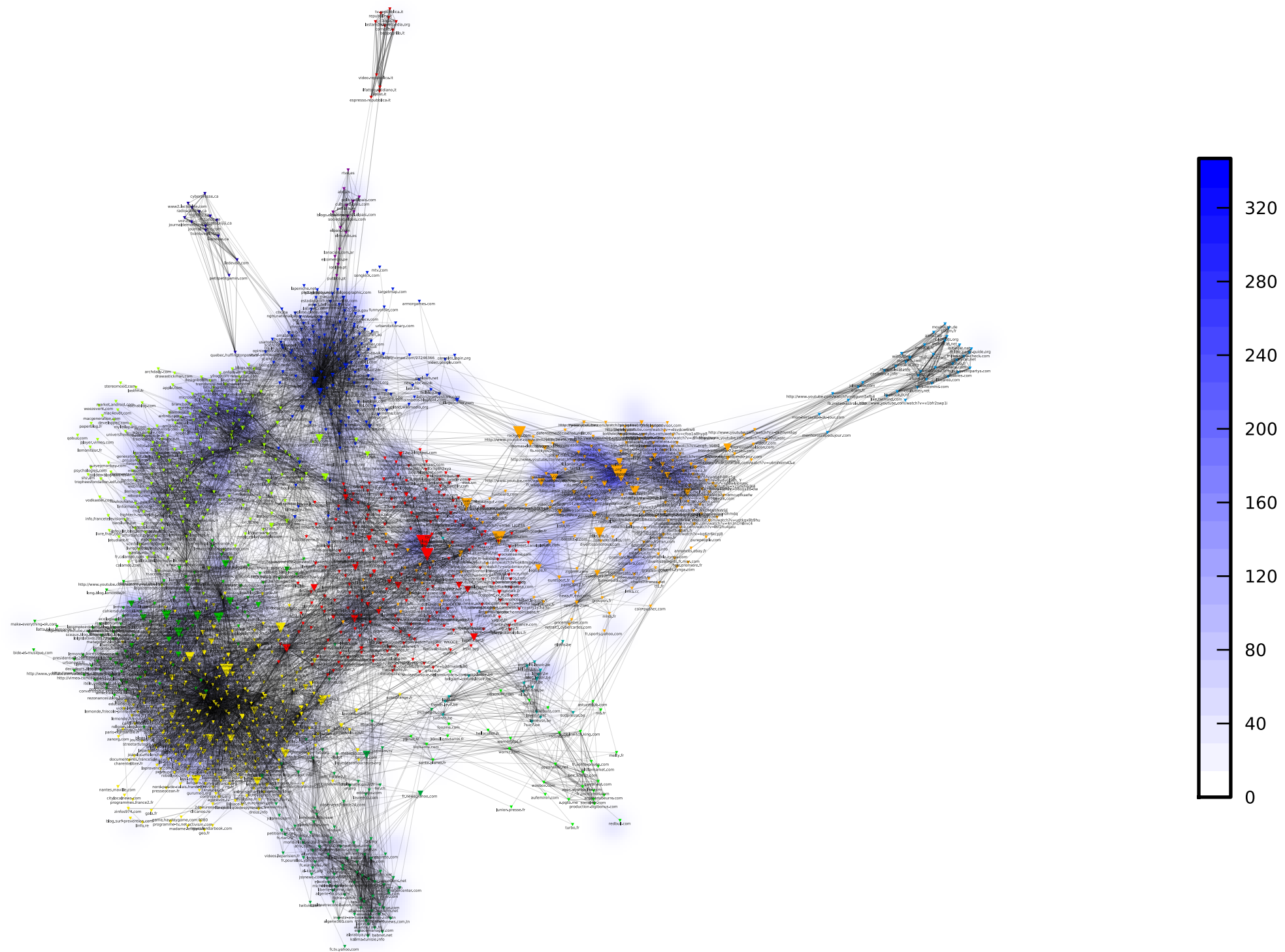
Mixing dimensions of analysis

femme, 8-30



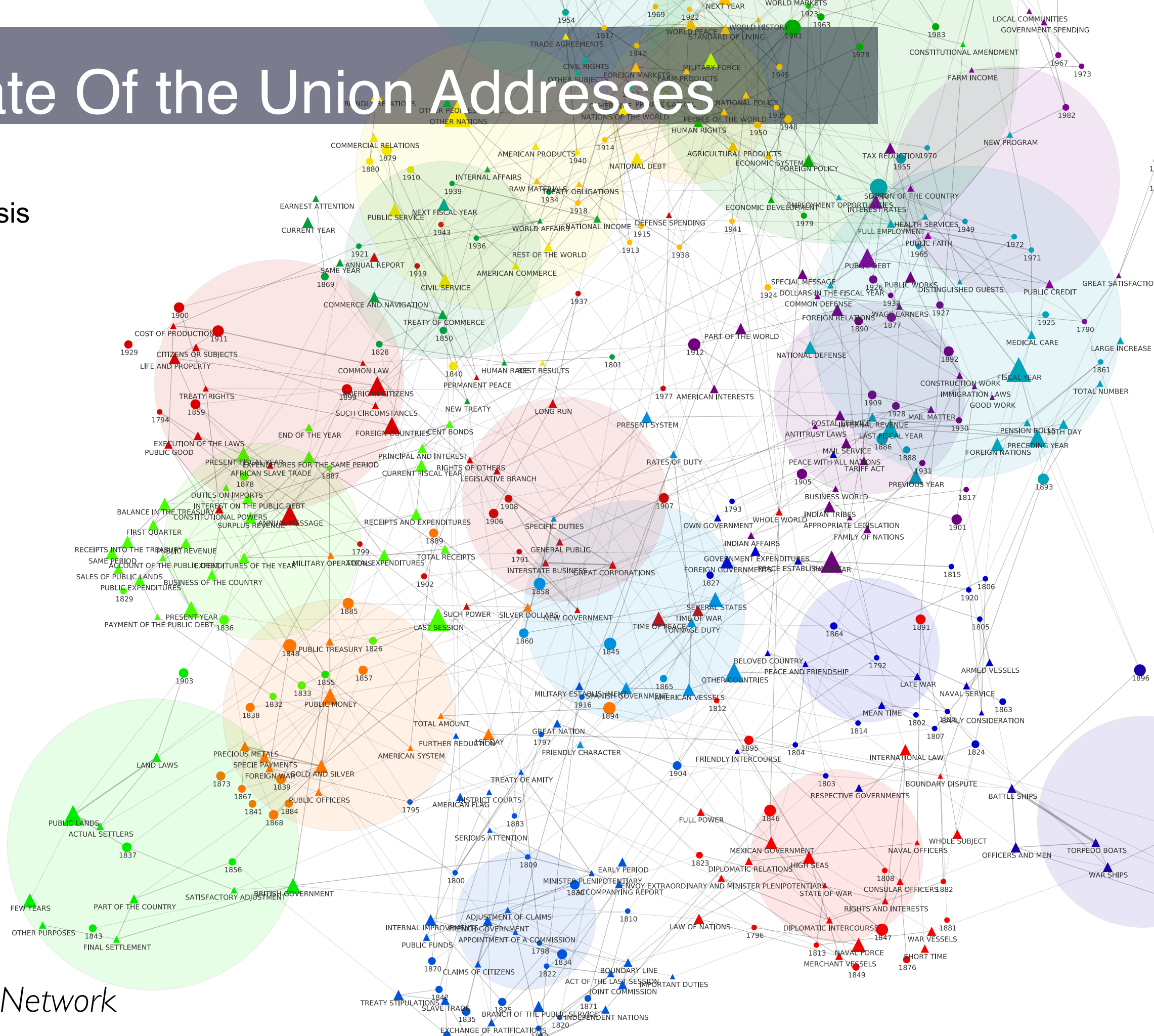
Mixing dimensions of analysis

homme, 8-30



State Of the Union Addresses

Global Analysis

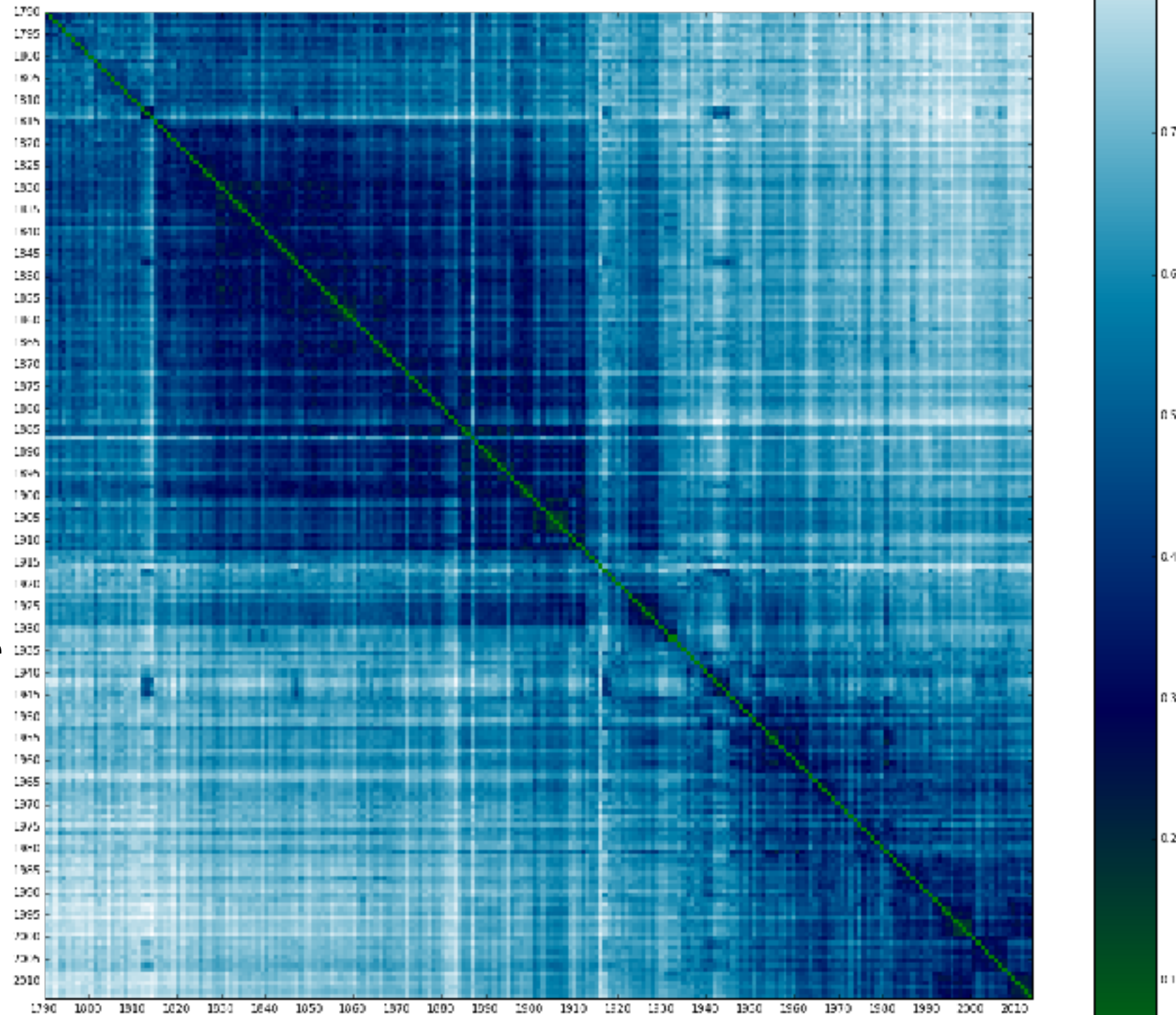


Term & Year Network

State Of the Union Addresses

- Global Analysis

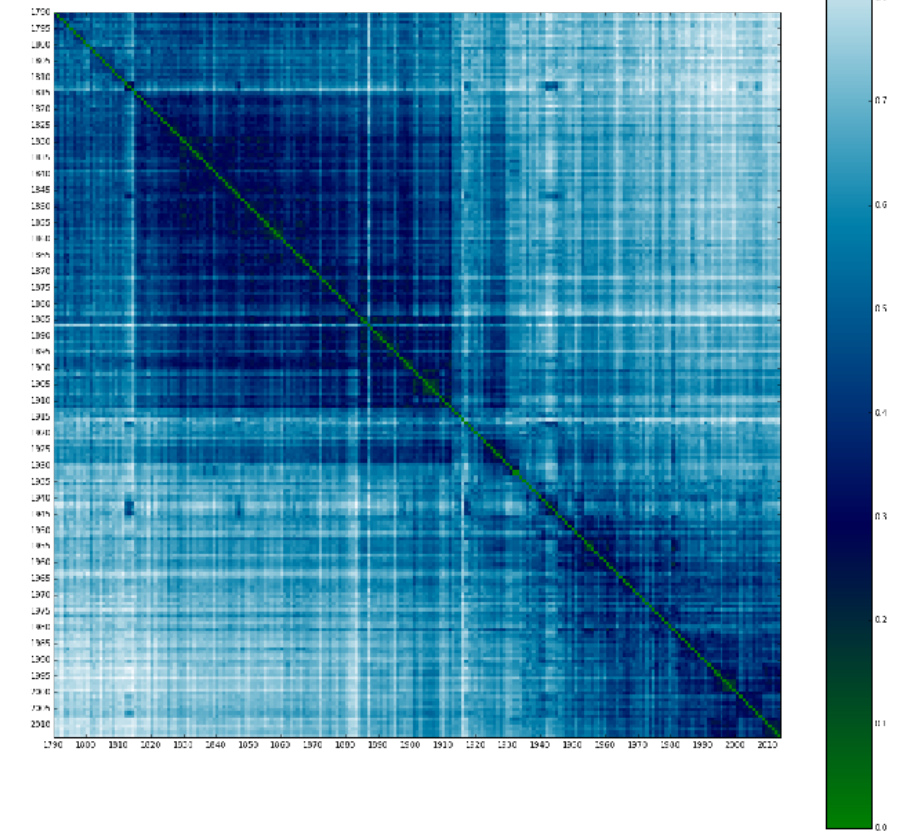
- *Vector space model: each address is reduced to a vector summarizing the term use (tf.idf)*
- *Every pair of years are then compared considering the cosine distance between their respective vectors*



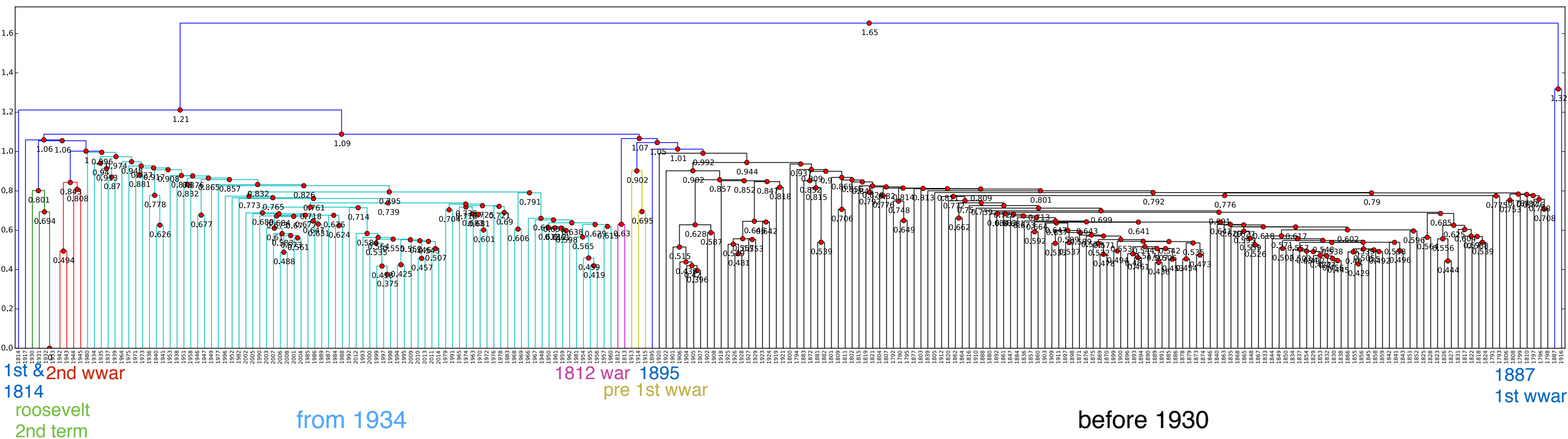
Transformation Matrix

State Of the Union Addresses

Global Analysis



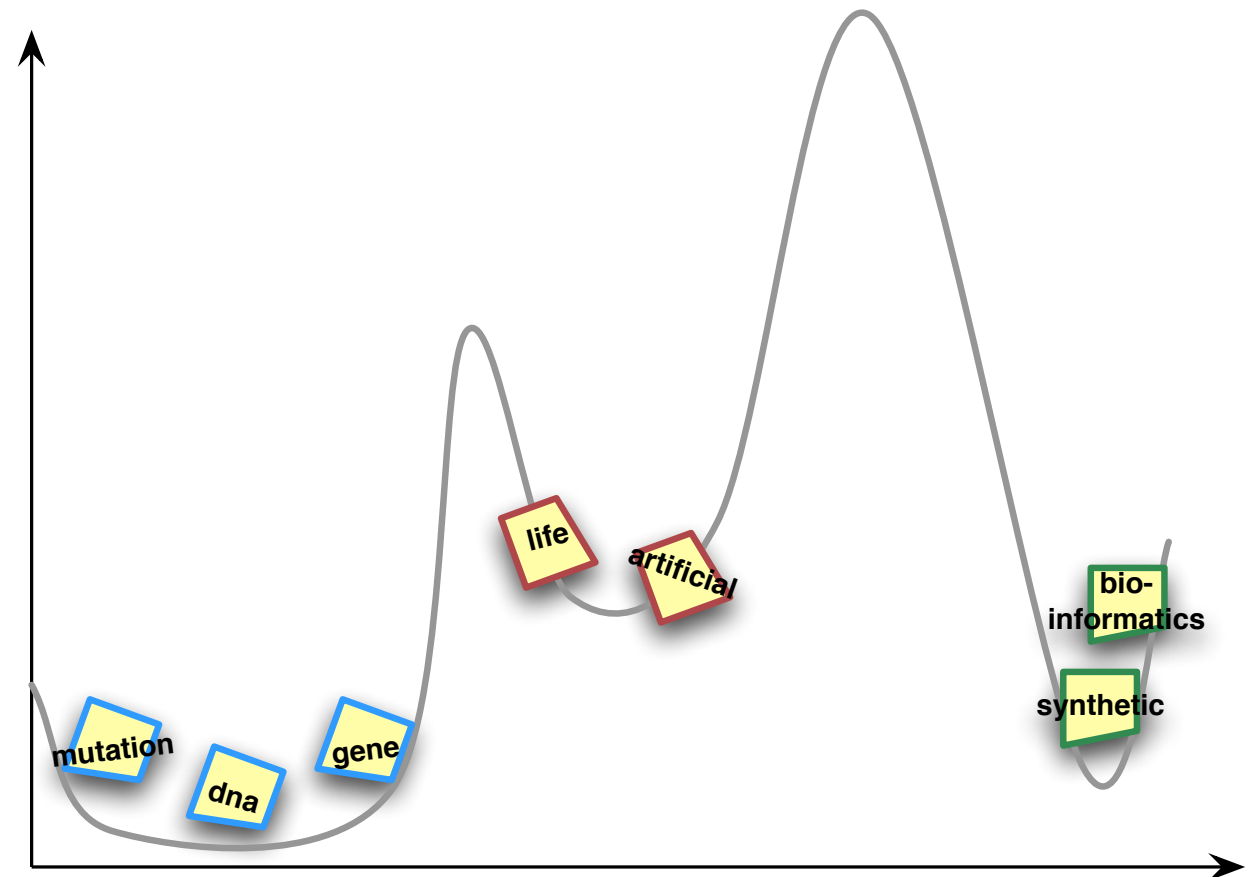
Ascending hierarchical clustering



Textual Analysis

what it is about a text that is interesting?

- **grammatical** criterion, candidate terms are usually limited *noun phrases*,
- **unithood**, phrases should represent a proper semantic unit,
- **termhood**, terms should be domain specific to carry substantial information



Textual Analysis

The phylogenetic position of the elephant shark (*Callorhynchus milii*) is particularly relevant to study the evolution of genes and gene regulation in vertebrates.

Textual Analysis

i. Part-Of-Speech Tagging

The phylogenetic position of the elephant shark (*Callorhynchus milii*) is particularly

DT JJ NN IN DT NN NN (NNS NN)VBZ RB

relevant to study the evolution of genes and gene regulation in vertebrates.

JJ TO VB DT NN IN NNS CC NN NN IN NNS

Textual Analysis

i. Part-Of-Speech Tagging

ii. Tag Chunking - Noun Phrases extraction

ex: Regexp={((Adj|Noun)+|((Adj|Noun)*NounPrep?)(Adj|Noun)*Noun}

The phylogenetic position of the elephant shark (*Callorhynchus milii*) is particularly

DT JJ NN IN DT NN NN (NNS NN)VBZ RB

relevant to study the evolution of genes and gene regulation in vertebrates.

JJ TO VB DT NN IN NNS CC NN NN IN NNS

Textual Analysis

i. Part-Of-Speech Tagging

ii. Tag Chunking - Noun Phrases extraction

ex: Regexp={((Adj|Noun)+|((Adj|Noun)*NounPrep?)(Adj|Noun)*Noun}

iii. Stemming and filtering of empty words

gene regulation in vertebrate -> {gene regul vertebr}

phylogenetic position of the elephant shark : {eleph phylogenet posit shark}

phylogenetic position -> {phylogenet posit}

Textual Analysis

i. Part-Of-Speech Tagging

ii. Tag Chunking - Noun Phrases extraction

ex: Regexp={((Adj|Noun)+|(Adj|Noun)*NounPrep?)(Adj|Noun)*}Noun}

iii. Stemming and filtering of empty words

iv. Output: classes of candidate multi-terms:

- cellular isoform prion protein = {isoform of cellular prion protein ; cellular isoform of the prion protein ; cellular prion protein isoform ; isoform of the cellular prion protein ; cellular isoform of prion protein}
- conform: {conformers ; conformational ; conformation ; conformer ; conformations}
- resist scrapi: {resistance against scrapie ; scrapie resistance ; scrapie resistant ; Scrapie resistance}
- associ genotyp prp = {association of PrP genotype ; associations between PrP genotypes ; association between PrP genotype ; associations of the PrP genotype ; associations between PrP genotypes}

Textual Analysis

Unithood: extracting semantic units with C-value

- Simple frequency-based approach : «Real» Terms tend to appear more frequently than non-terms
- C-value approach (Frantzi K. & Ananiadou S., 2000):
 - Longer phrases are more likely to be relevant,
 - Nested terms may induce false positive, ex: « self organizing maps » vs « new promoter ».

Textual Analysis

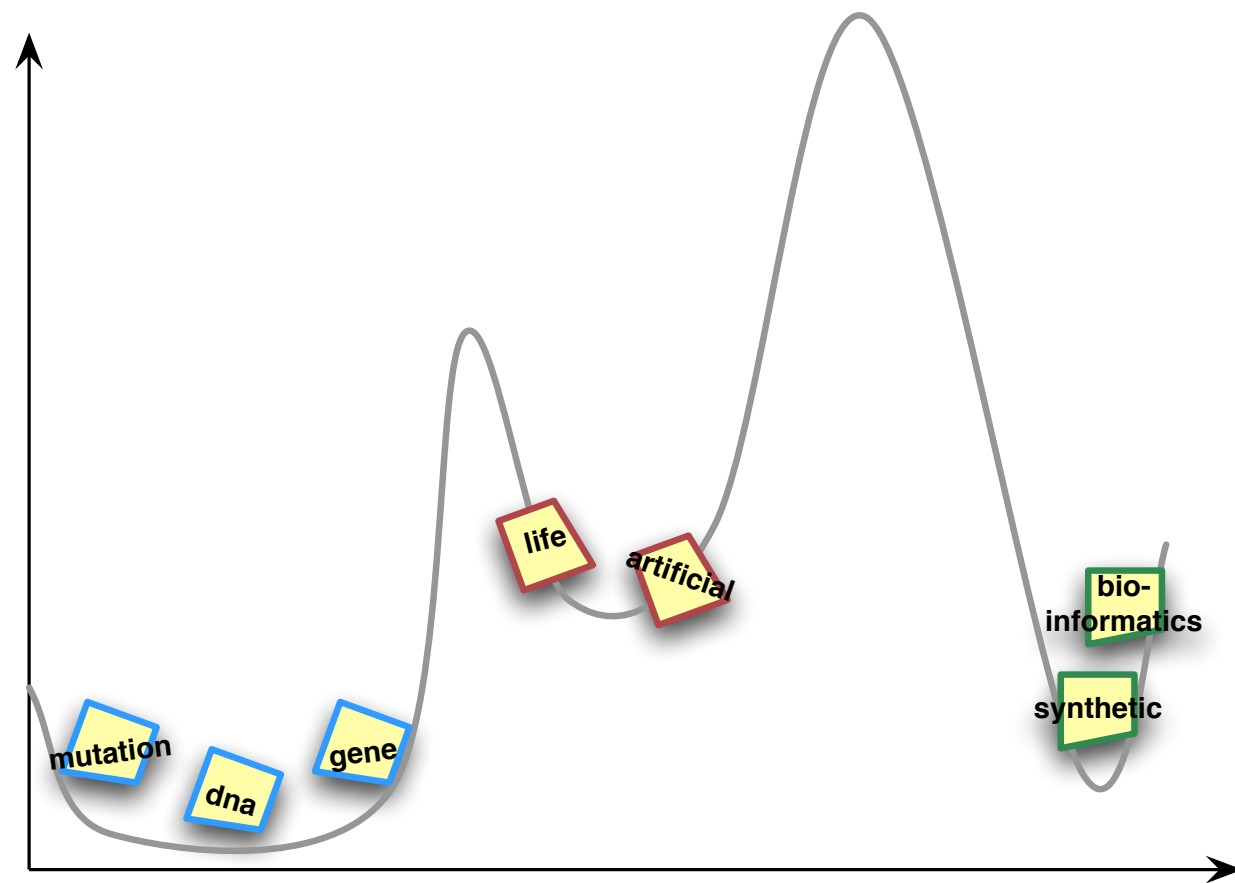
Termhood

- Candidate terms should be thematically specific ; terms not specific to a specific thematic subfield have neutral meaning given the whole domain and should be excluded
- On the contrary, terms which distribution is biased toward certain topics are more likely to have interesting meaning.
- Co-occurrences between existing candidate terms are extracted to compute the **Xhi2 score** of specificity of each term compared to other terms (Matsuo Y. & Ishizuka M., 2004).

$$\chi^2(w) = \sum_{g \in G} \frac{(\text{freq}(w, g) - n_w p_g)^2}{n_w p_g}$$

Textual Analysis

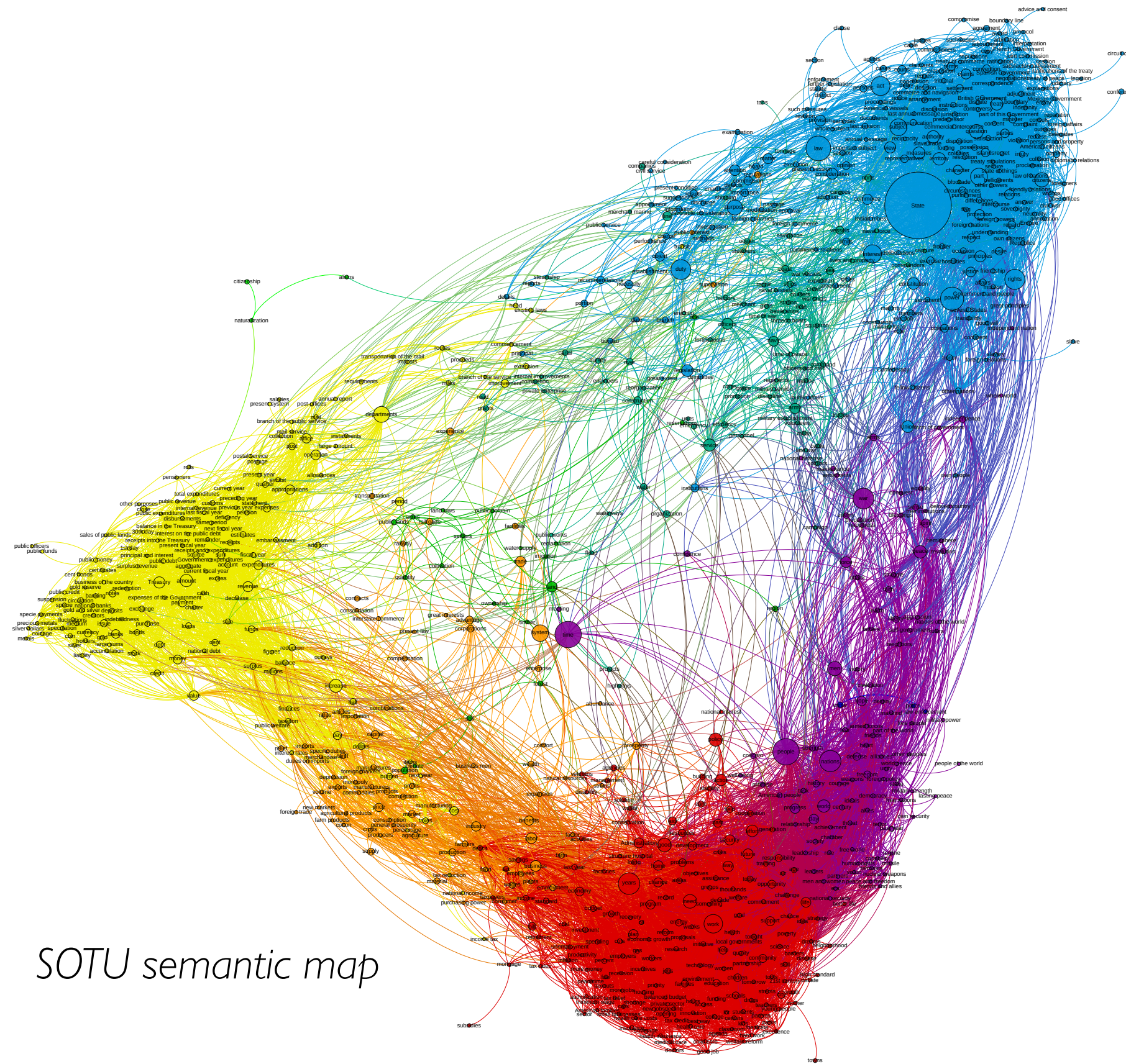
NLP analysis outcome



label	forms
heterologous expression	heterologous expression
complex networks	complex networks; complex network
design principles	design principle; design principles
transgene control	transgene control
gene expression noise	gene expression noise; noise in gene expression; gene-expression noise; noise in gene-expression
biosynthetic pathways	biosynthetic pathways; biosynthetic pathway
mathematical model	mathematical model; mathematical models; Mathematical models
metabolic pathways	metabolic pathway; metabolic pathways
control theory	control theory
reporter gene expression	reporter gene expression; reporter gene-expression
gene expression systems	gene expression systems; gene-expression systems
biological molecules	biological molecules; biological molecule
control circuit	control circuit; control circuits; circuit from a control; circuits control
toggle switch	Toggle switch; toggle switch; toggle switches
protein networks	protein networks; networks of proteins; proteins and networks
restriction enzymes	restriction enzymes; restriction enzyme
transcriptional network	transcriptional networks; transcriptional network; transcription networks
system design	system design; design these systems; systems design
Bacillus subtilis	Bacillus subtilis; bacillus subtilis
multiple DNA fragments	multiple DNA fragments; multiple dna fragments
drug delivery	drug delivery; delivery of drugs
biological clock	biological clock; biological clocks; Biological Clocks
mammalian cells	mammalian cells; Mammalian Cells
fuel production	production of fuels; Production of fuels; fuel production
artificial gene networks	artificial gene networks
control strategy	control strategy; control strategies; strategies that control
gene expression levels	gene expression levels; level of gene expression; gene-expression levels; level of gene-expression
computational approach	computational approach; Computational approaches; computational approaches; computational approach-; computational Approach; Computational approach-; Computational approach; Computational Approach
DNA synthesis	DNA synthesis; dna synthesis
gene networks	gene networks; gene network; Gene networks; networks of genes; Gene Network; genes and networks
biological parts	Biological Parts; biological parts
genetic information system	genetic information system; genetic -information system; genetic Information system
engineered microbes	engineered microbes; Engineering microbes; engineering microbes
Registry of Standard Biological Parts	registry of standard biological parts; Registry of Standard Biological Parts; Registry of Standard-Biological Parts; Registry of STANDARDS Biological Parts; Registry of standards Biological Parts; Registry of Standards Biological Parts

State Of the Union Addresses

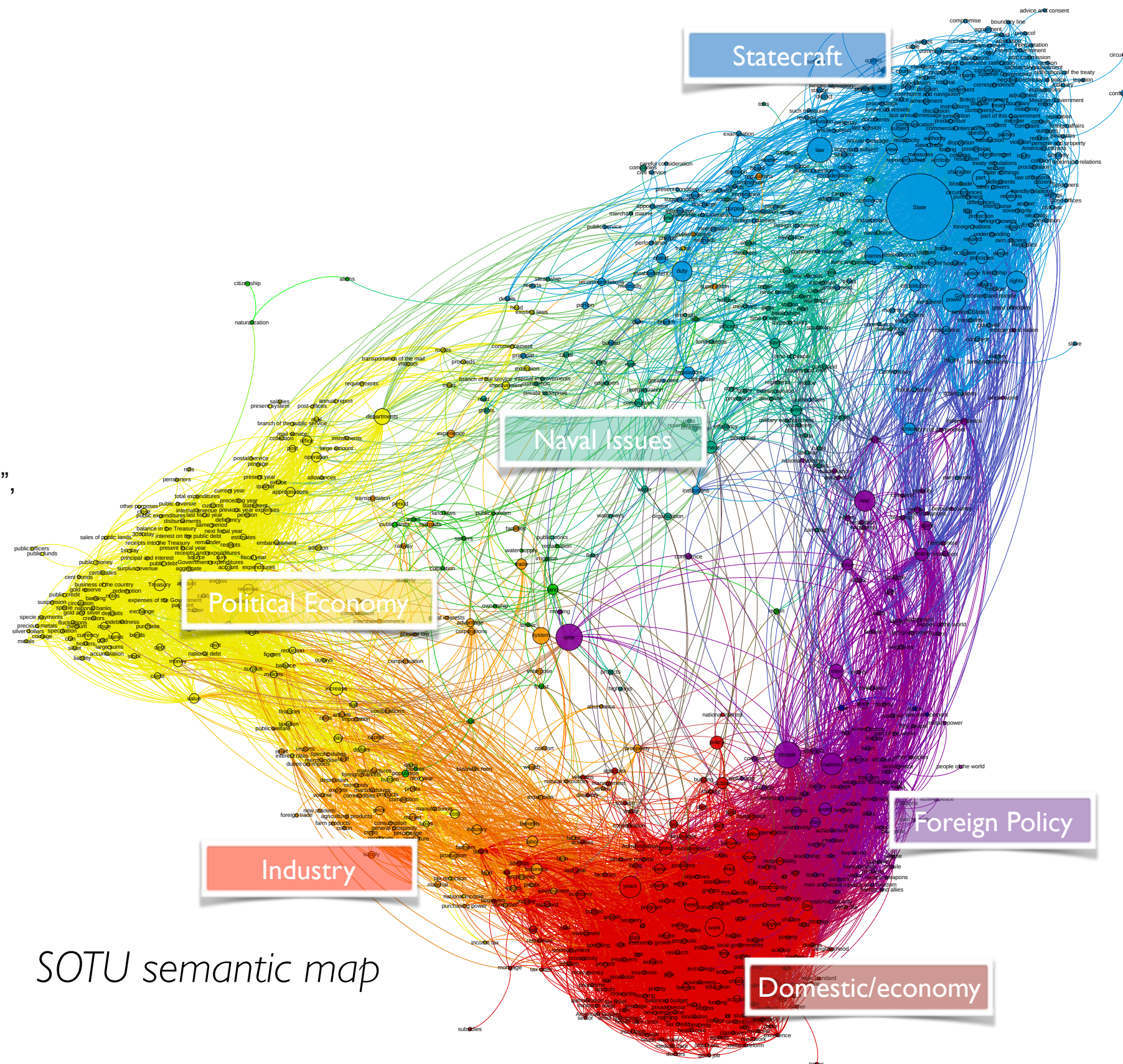
- Global Analysis
- State-of-the-Union
Addresses synthesized in
one semantic network



SOTU semantic map

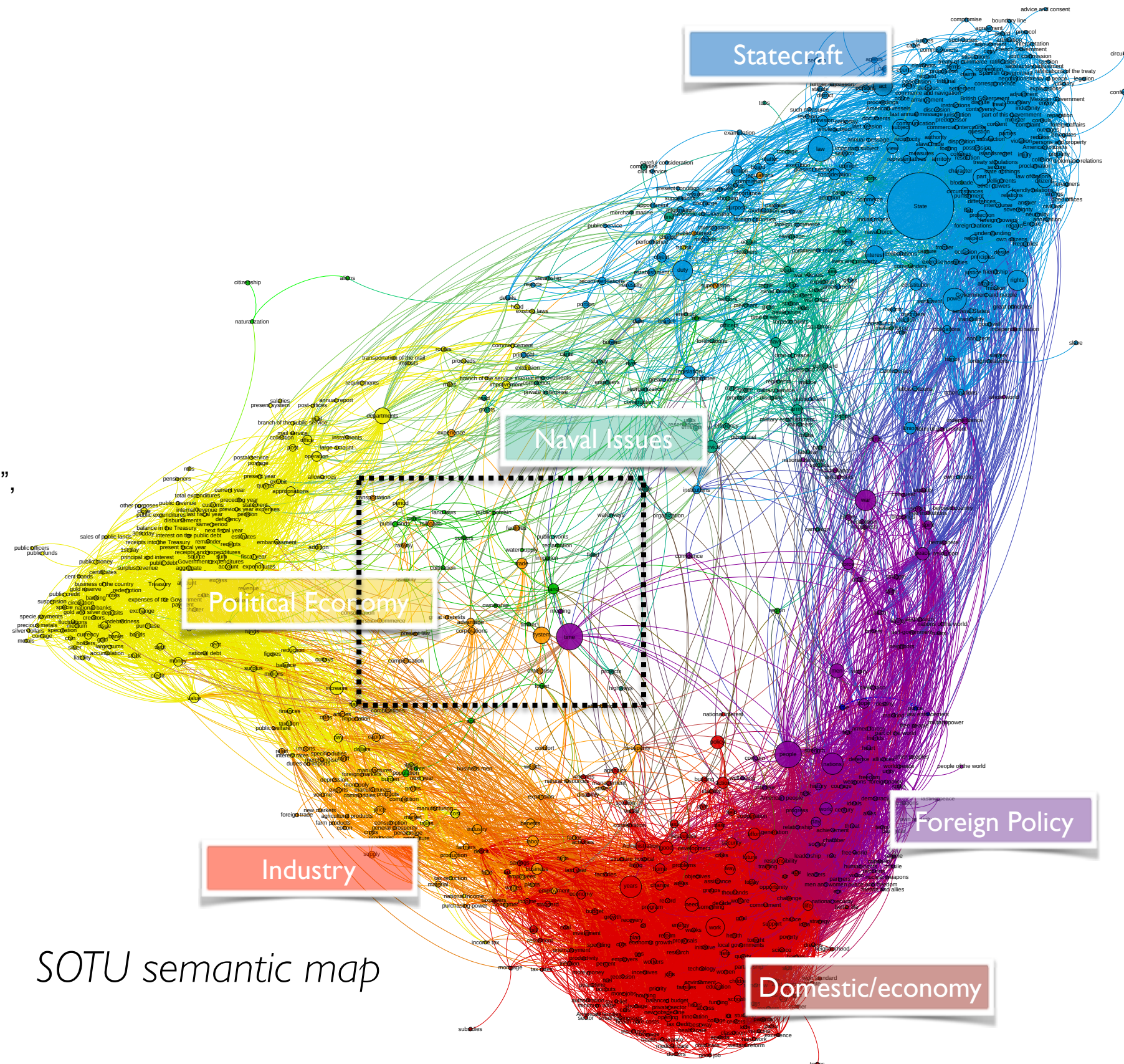
State Of the Union Addresses

- Global Analysis
- State-of-the-Union Addresses synthesized in one semantic network
- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Land** — “land”, “settlers”, “acres”
- **Immigration** — “naturalization”
- **Crime** — “criminal”



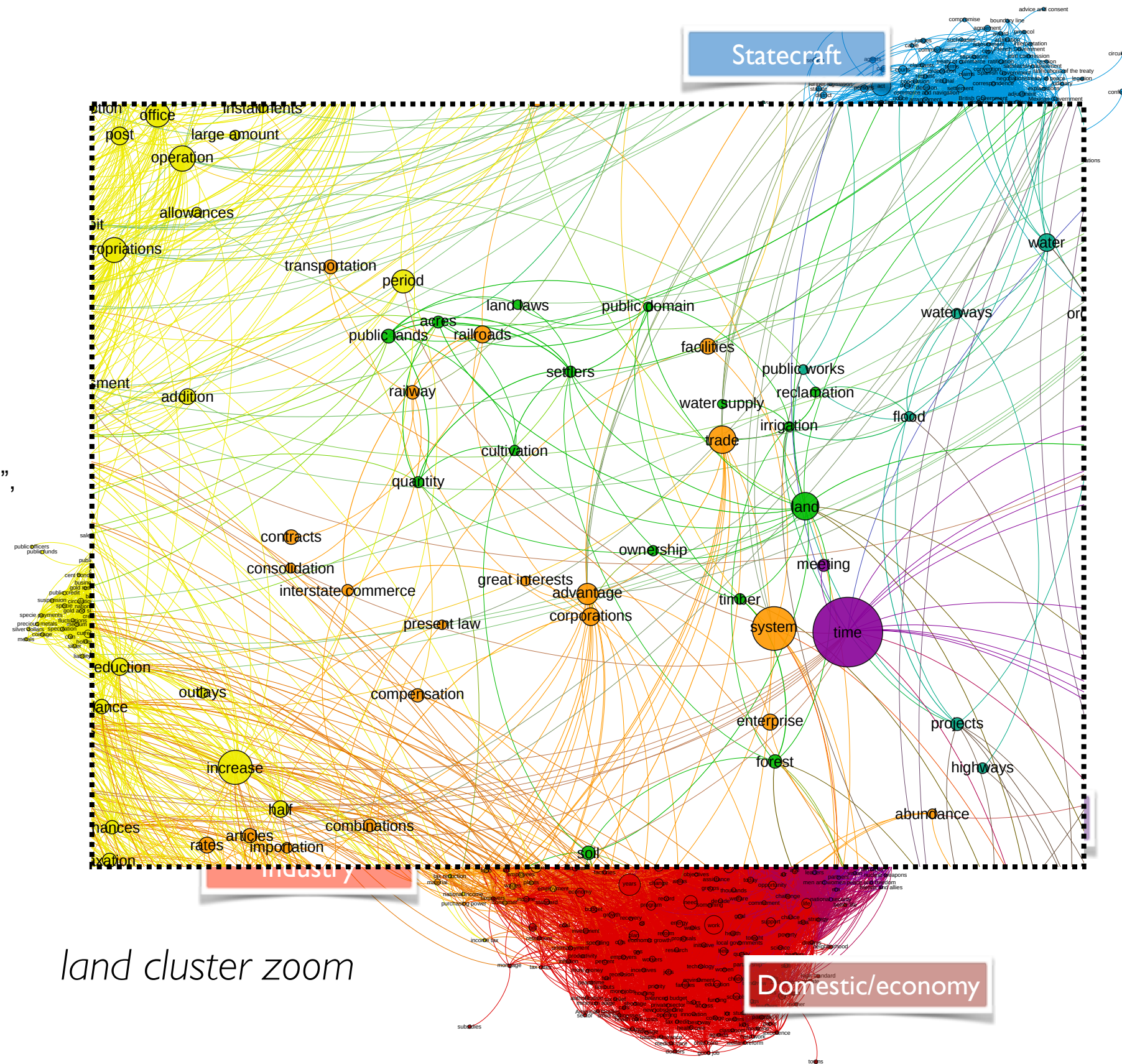
State Of the Union Addresses

- Global Analysis
- State-of-the-Union Addresses synthesized in one semantic network
- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
- **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
- **Land** — “land”, “settlers”, “acres”
- **Immigration** — “naturalization”
- **Crime** — “criminal”



State Of the Union Addresses

- # Global Analysis
- ## State-of-the-Union
- Addresses synthesized in one semantic network
- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
 - **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
 - **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
 - **Domestic/economy** — “Need”, “work”, “economy”, “years”, “policy”
 - **Industry** — “system”, “trade”, “corporations”, “business”, “labor”
 - **Political economy** — “Treasury”, “amount”, “appropriations”, “value”
 - **Land** — “land”, “settlers”, “acres”
 - **Immigration** — “naturalization”
 - **Crime** — “criminal”



State Of the Union Addresses

- Global Analysis
- State-of-the-Union
Addresses synthesized in
one semantic network
- Capture topics with a
delimited time span

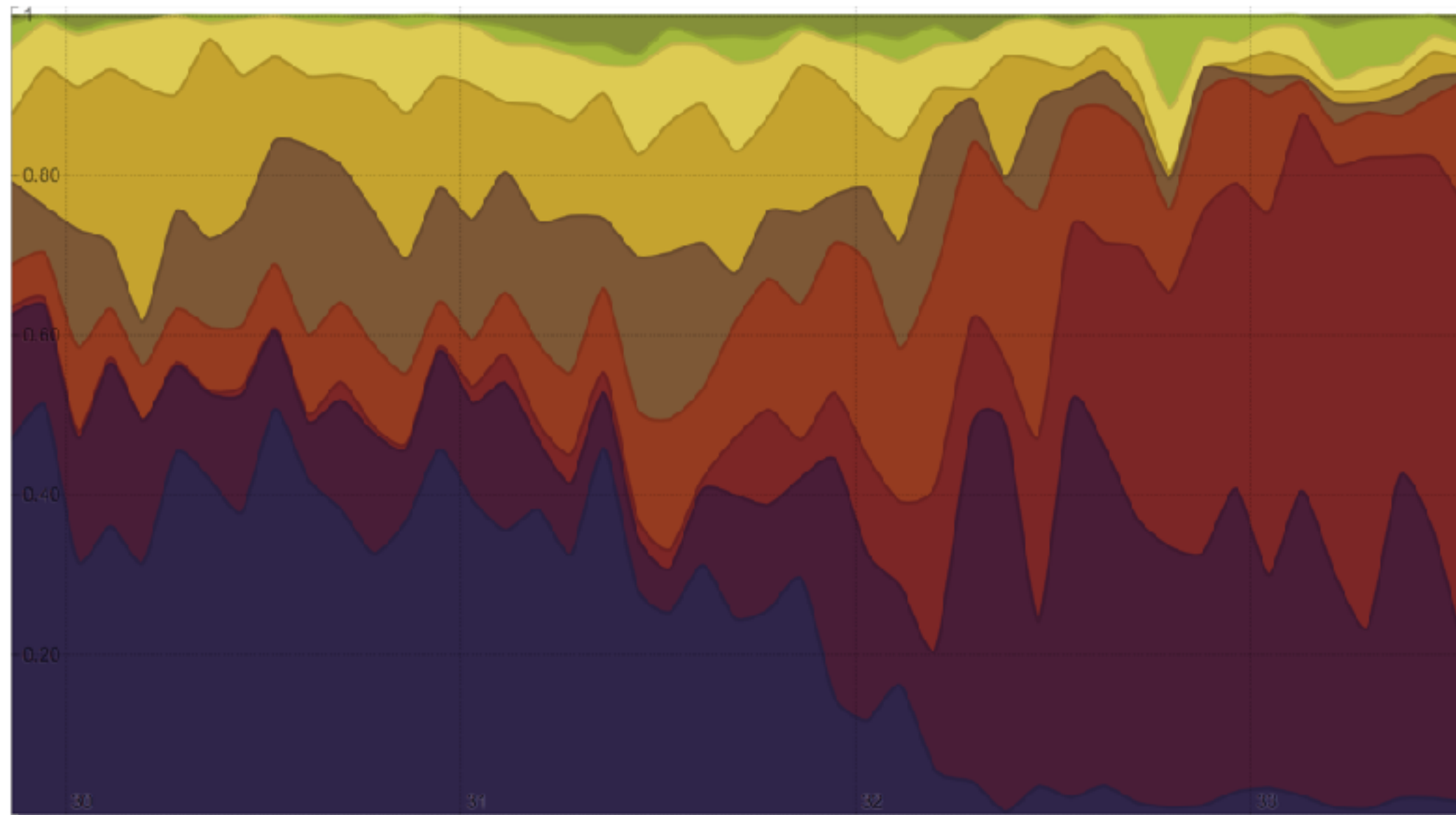
Field Evolution

- ✓ naturalization & citizenship
- ✓ police & crime
- ✓ public lands & settlers
- ✓ navy & sea
- ✓ amount & Treasury
- ✓ industry & production
- ✓ jobs & help
- ✓ freedom & world
- ✓ State & minister



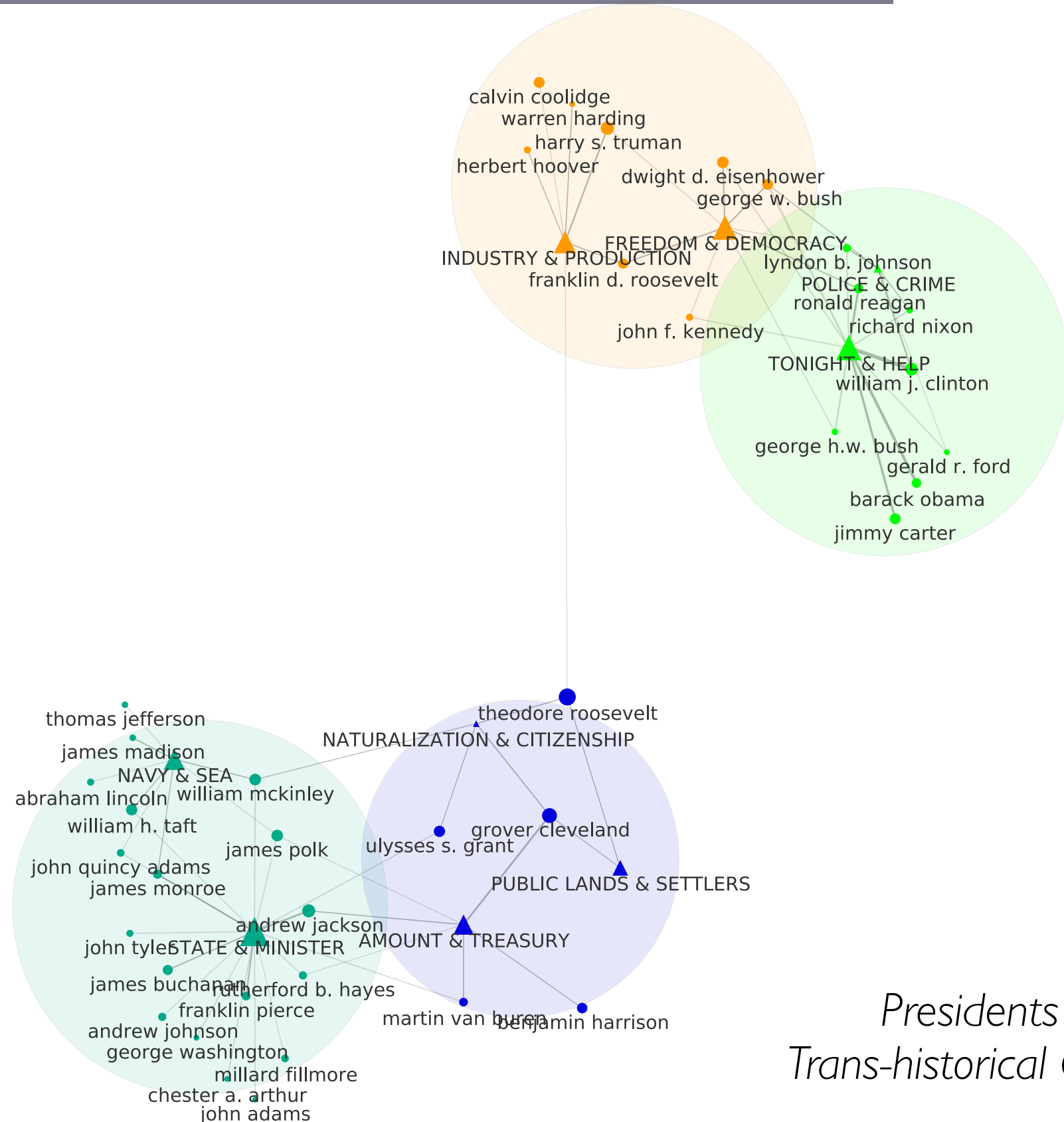
- stack
- stream
- pct
- value
- cardinal
- linear
- step

Smoothing



State Of the Union Addresses

- Global Analysis
- State-of-the-Union
Addresses synthesized in
one semantic network
- Capture topics with a
delimited time span



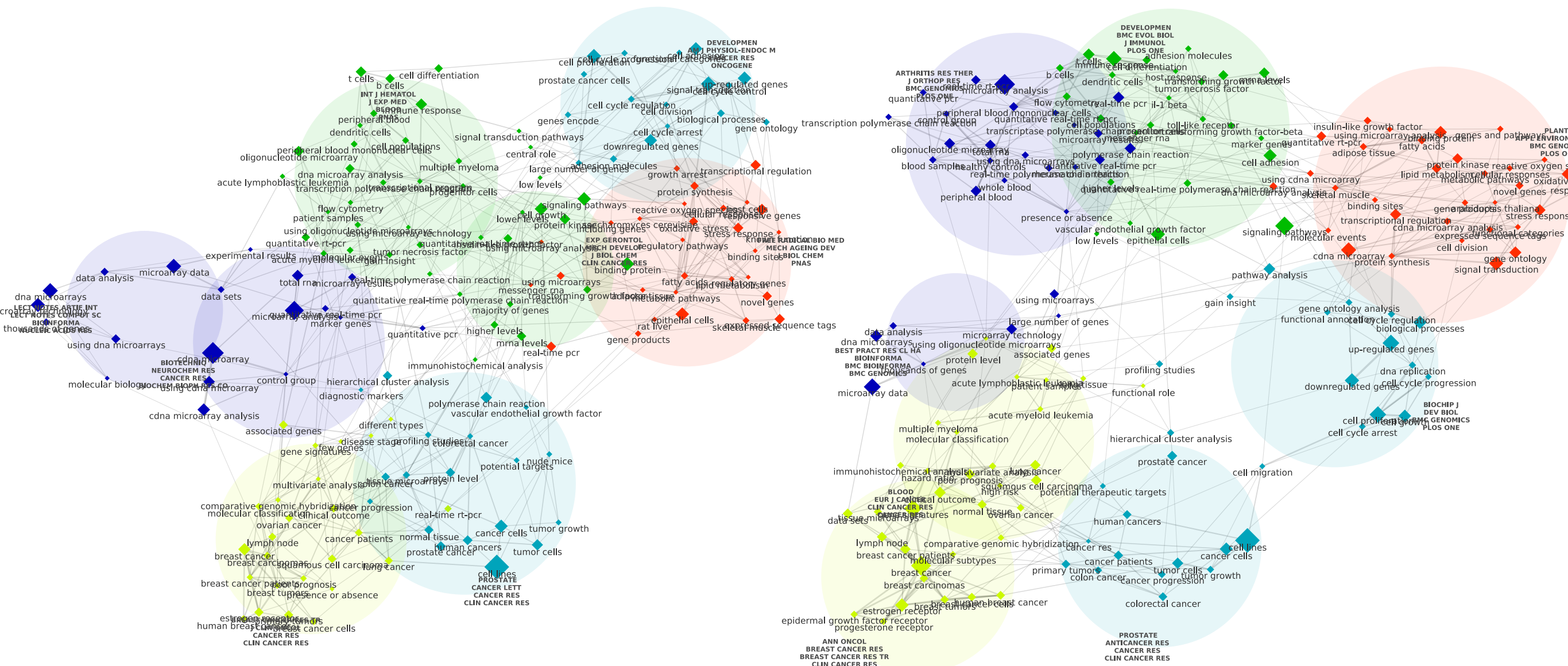
*Presidents &
Trans-historical Clusters*

Dynamical mapping

Inter-temporal matching between two maps

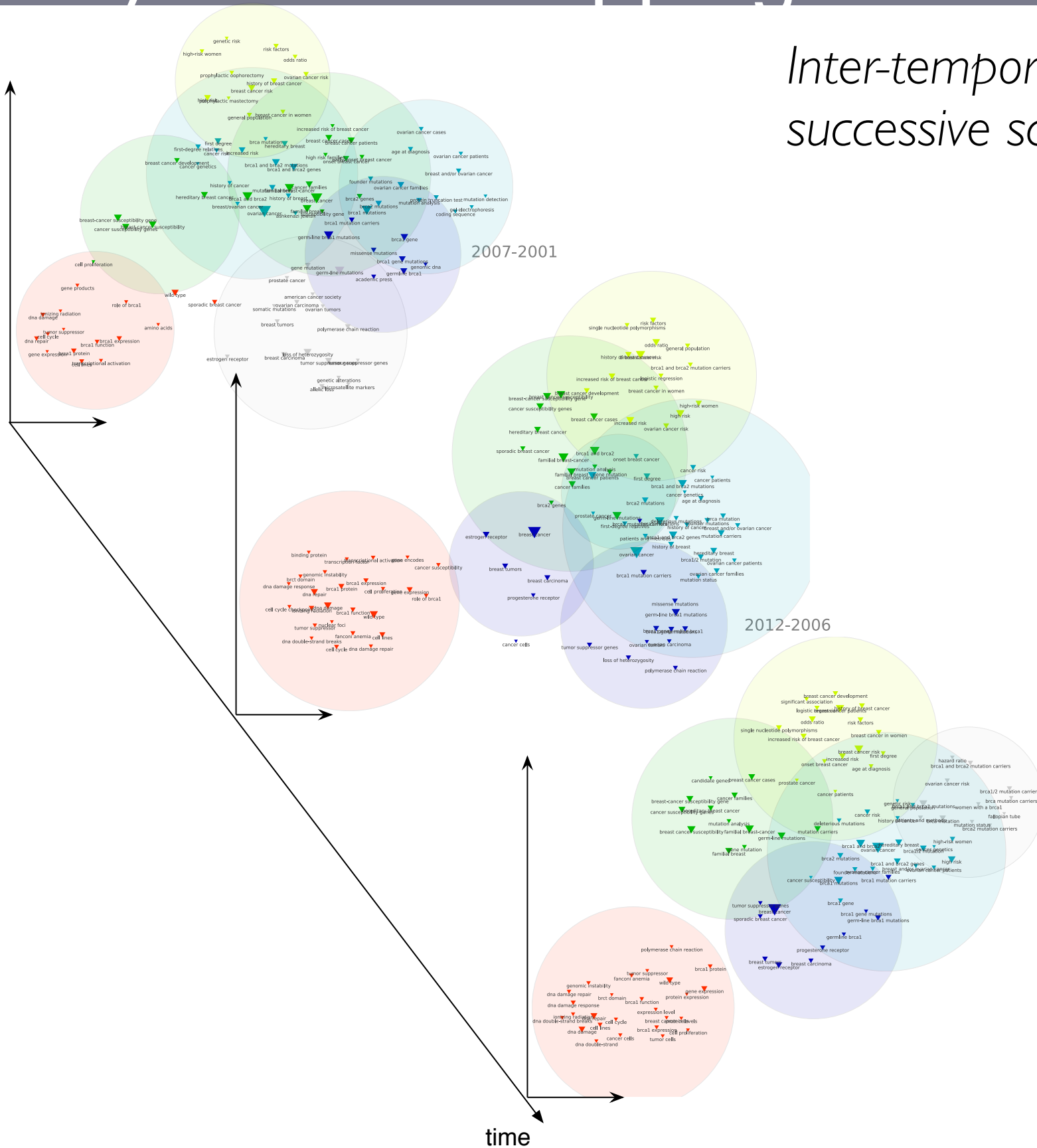
2001-2005

2006-2010



Dynamical mapping

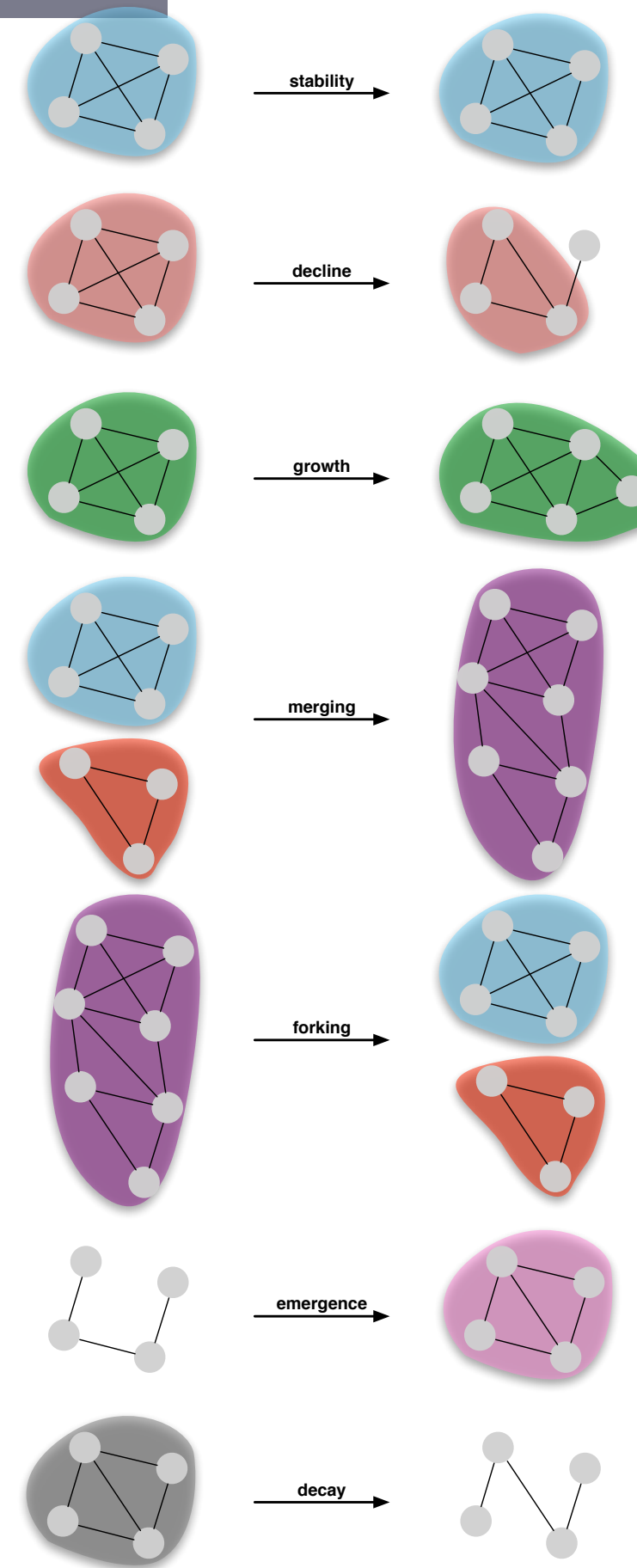
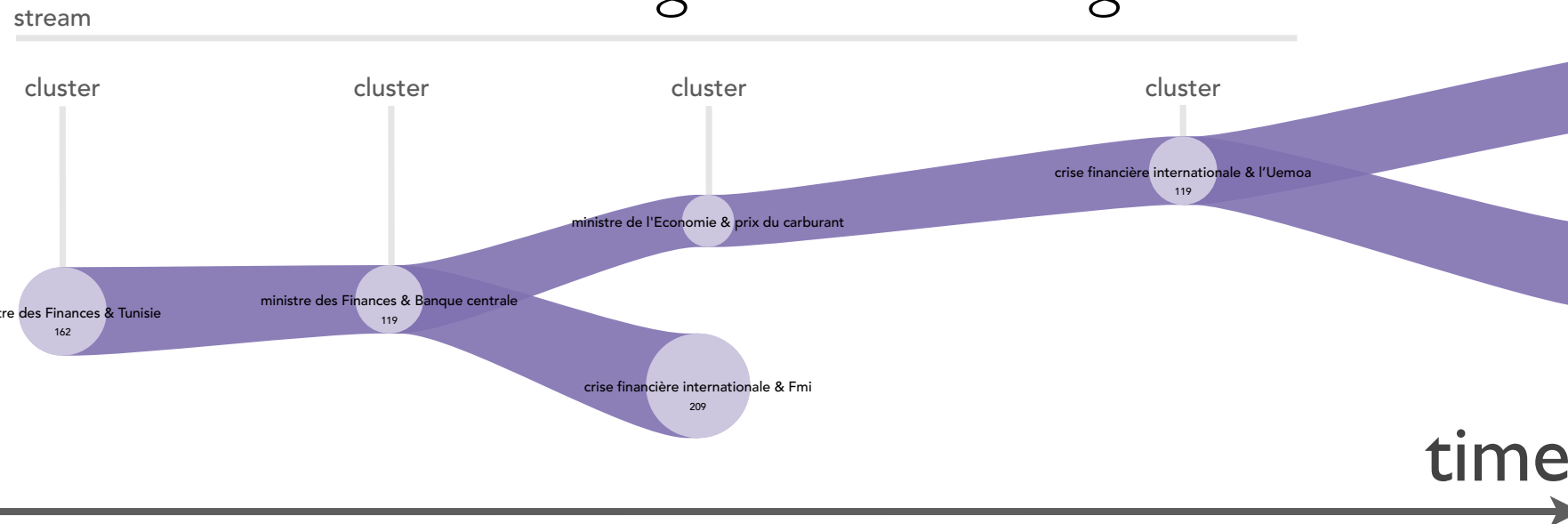
Inter-temporal matching between successive science maps



Stream Reconstruction

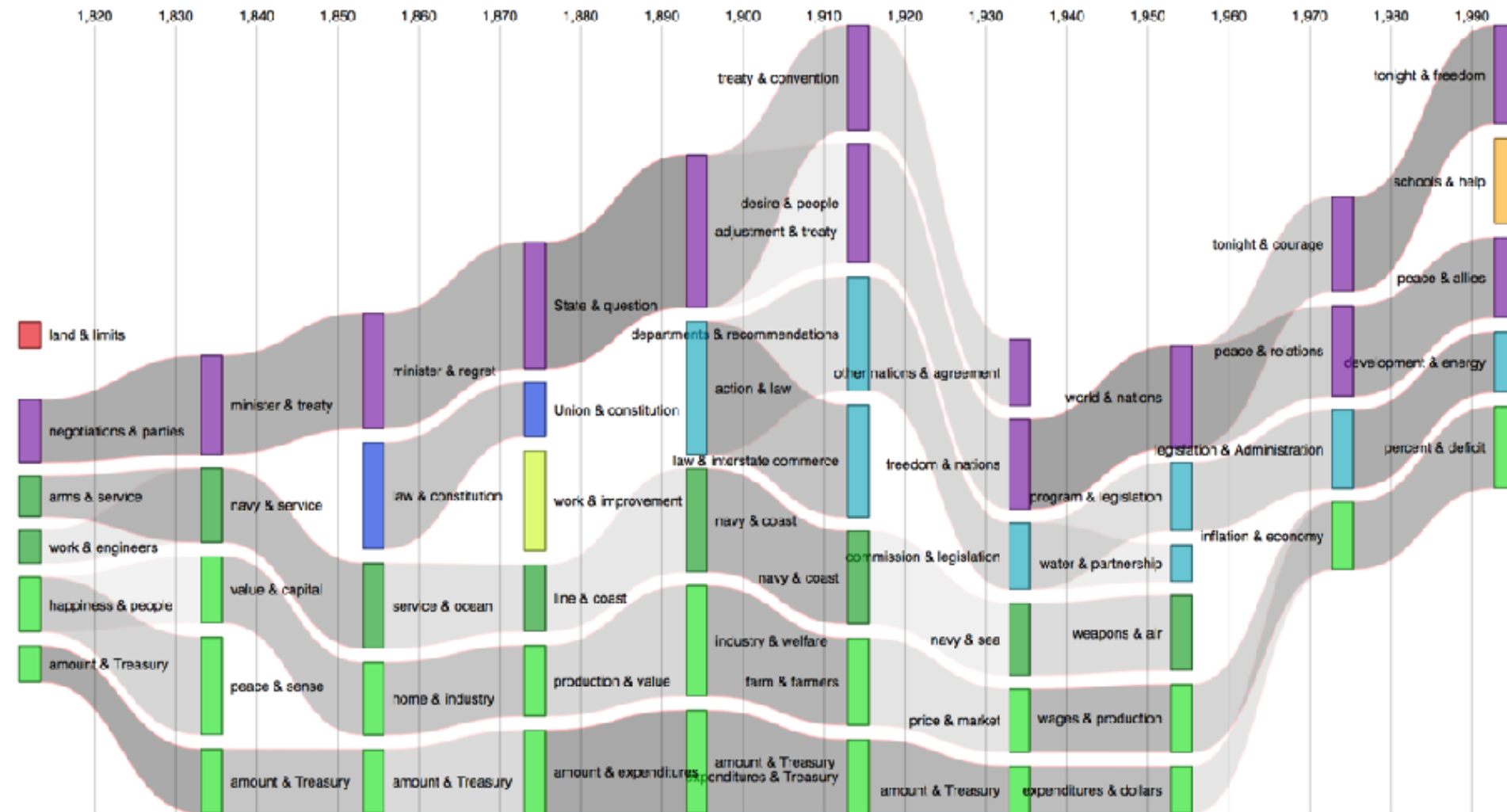
Possible events
at mesoscopic
level

Clusters are intertemporally grouped
into streams according to their lineage



State Of the Union Addresses

- Global Analysis
- State-of-the-Union Addresses synthesized in one semantic network
- Capture topics with a delimited time span
- Conversely, river networks map the semantic structure at successive periods and then reconstruct (dis)continuities



River Network