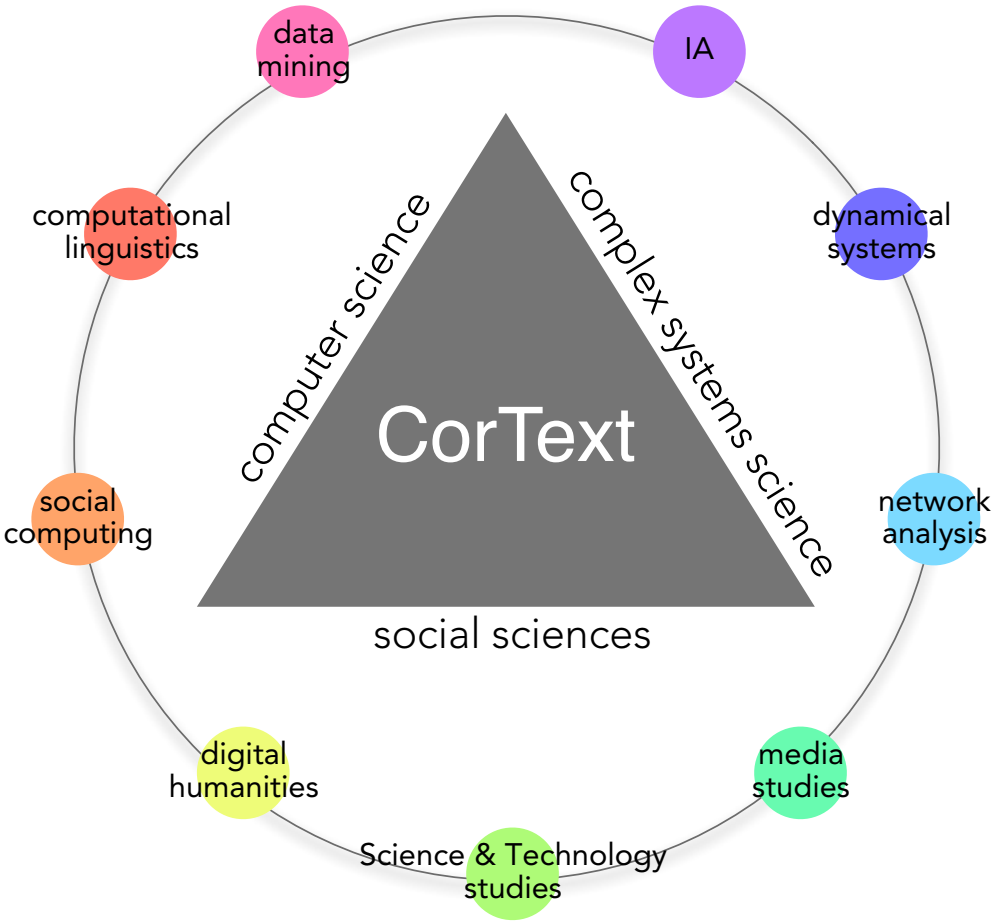
Modeling and mapping knowledge dynamics with CorText platform

Marne, Juin 2017

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



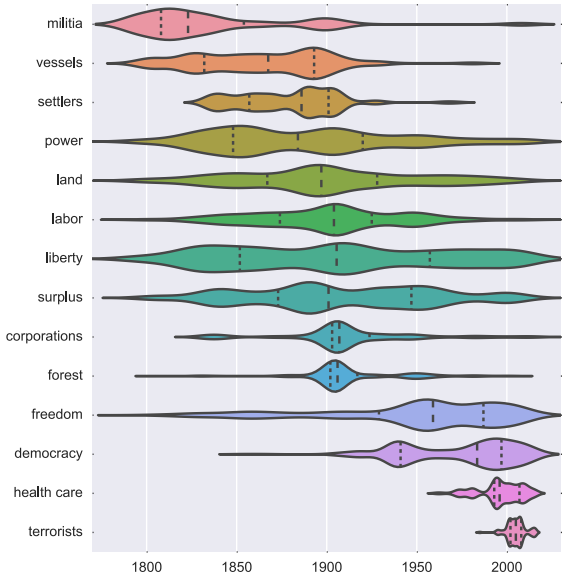
Multi-level perspective

Dynamics

Relationships

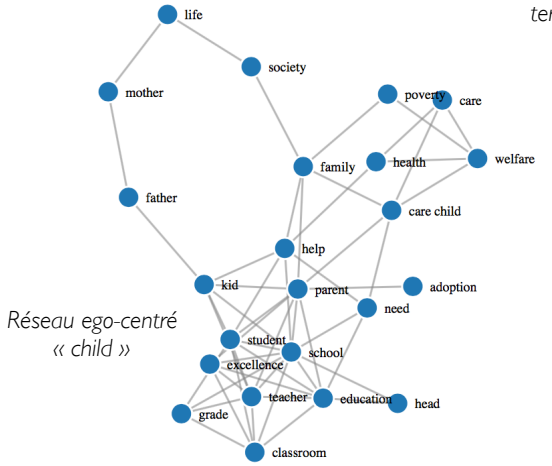
Texts

ECHELLE MICRO



Profils d'évolution des quelques termes

Dynamique des mots



Réseau ego-centré « child »

Structure locale

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|&navigation 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|&lgreatest importance
health care costs	health care costs|&lcost of health care|&lhealth care cost
income tax	income tax|&ltax the income|&ltax on the 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

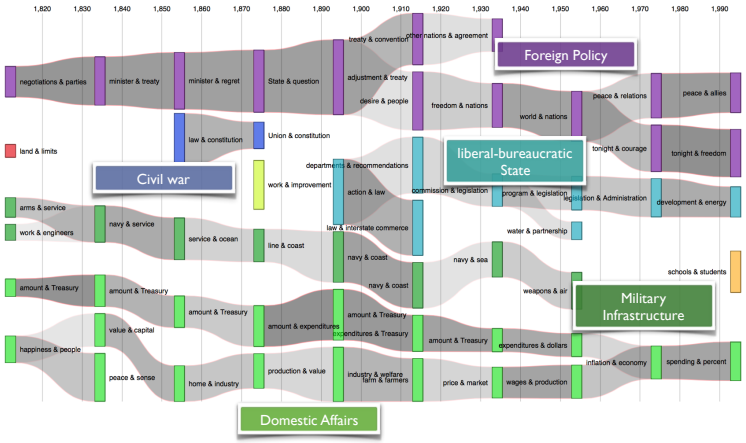
Multi-level perspective

Dynamics

Relationships

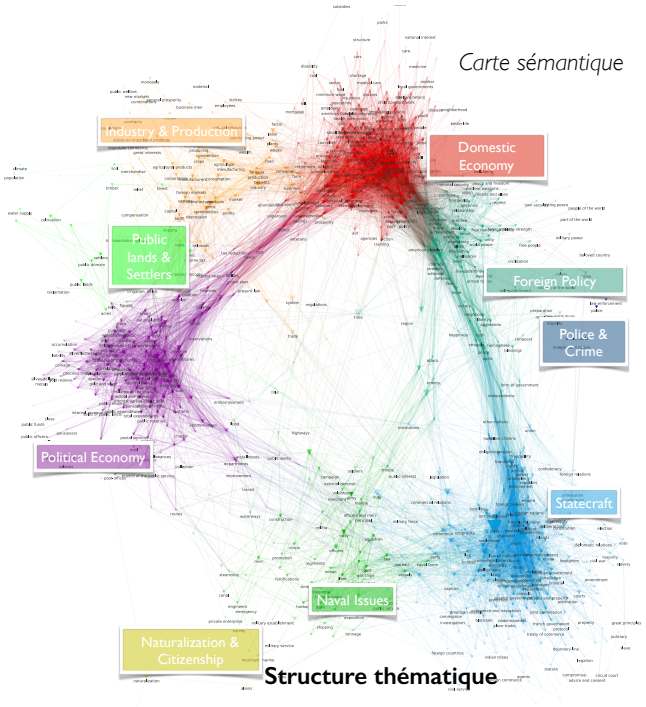
Texts

Diagramme alluvial de l'évolution des thématiques



Dynamique des grandes thématiques

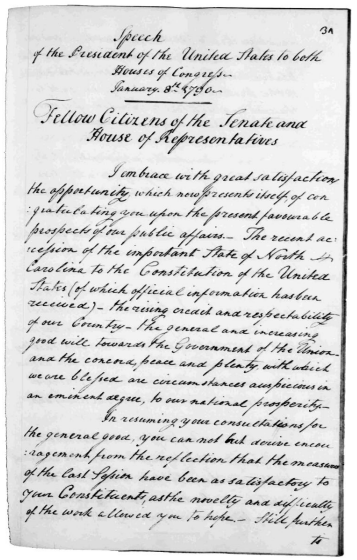
Carte sémantique



Structure thématique

ECHELLE MESO

Discours de Georges Washington



Speech
of the President of the United States to both
Houses of Congress.
January 8th 1790.

Fellow Citizens of the Senate and
House of Representatives

I embrace with great satisfaction
the opportunity which now presents itself of con-
gratulating you upon the present favourable
prospects of our public affairs. The recent ac-
cession of the important State of North
Carolina to the Constitution of the United
States (of which official information has been
received) the rising credit and respectability
of our Country, the general and increasing
good will towards the Government of the Union
and the concord, peace and plenty with which
we are blessed are circumstances auspicious in
an eminent degree, to our national prosperity.

In resuming your consultations for
the general good, you can not but derive encou-
ragement from the reflection that the measures
of the last Session have been as satisfactory to
your Constituents, as the novelty and difficulty
of the work allowed you to hope. It will further

Documents

Multi-level perspective

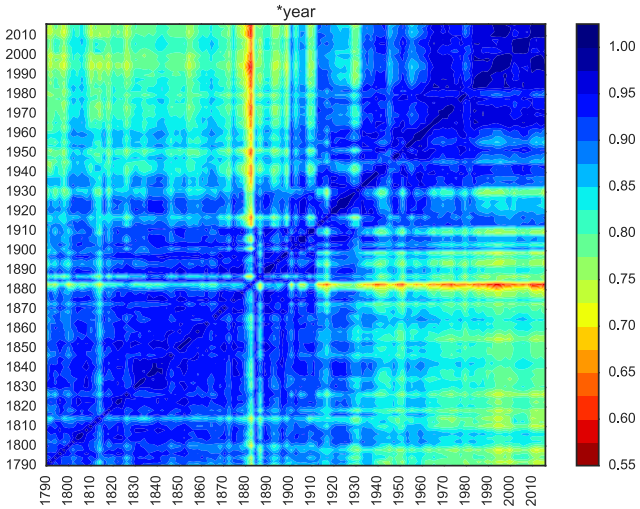
Dynamics

Relationships

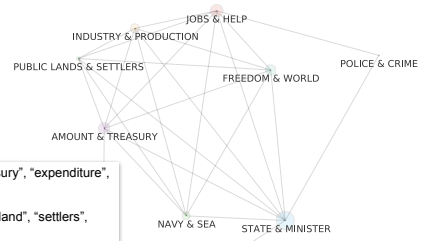
Texts

ECHELLE MACRO

Variations inter-annuelles



Dynamique des discours

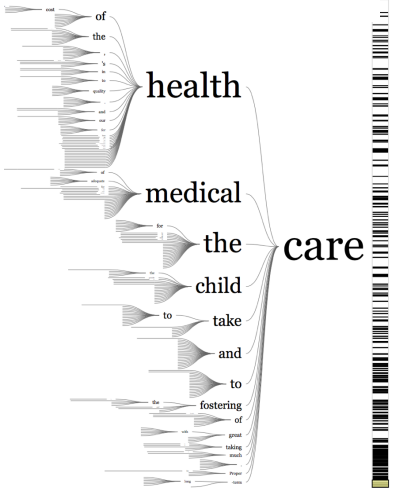


- **Political economy** — "Treasury", "expenditure", "appropriations", "value"
- **Public lands & Settlers** — "land", "settlers", "acres", "forest"
- **Industry & Production** — "factory", "business", "farm", "agriculture"
- **Domestic/economy**— "school", "doctors", "productivity", "economy"
- **Foreign policy** — "war", "free nation", "force", "peace"
- **Police & Crime** — "criminal", "law inforcement"
- **Statecraft** — "state", "power", "frontier", "treaty of commerce"
- **Naval issues** — "navy", "sea", "coast", "officers", "vessels"
- **Naturalization & Citizenship** — "naturalization", "aliens"

Articulation entre les grandes tâches de la gouvernance

Structure macro des grands thèmes

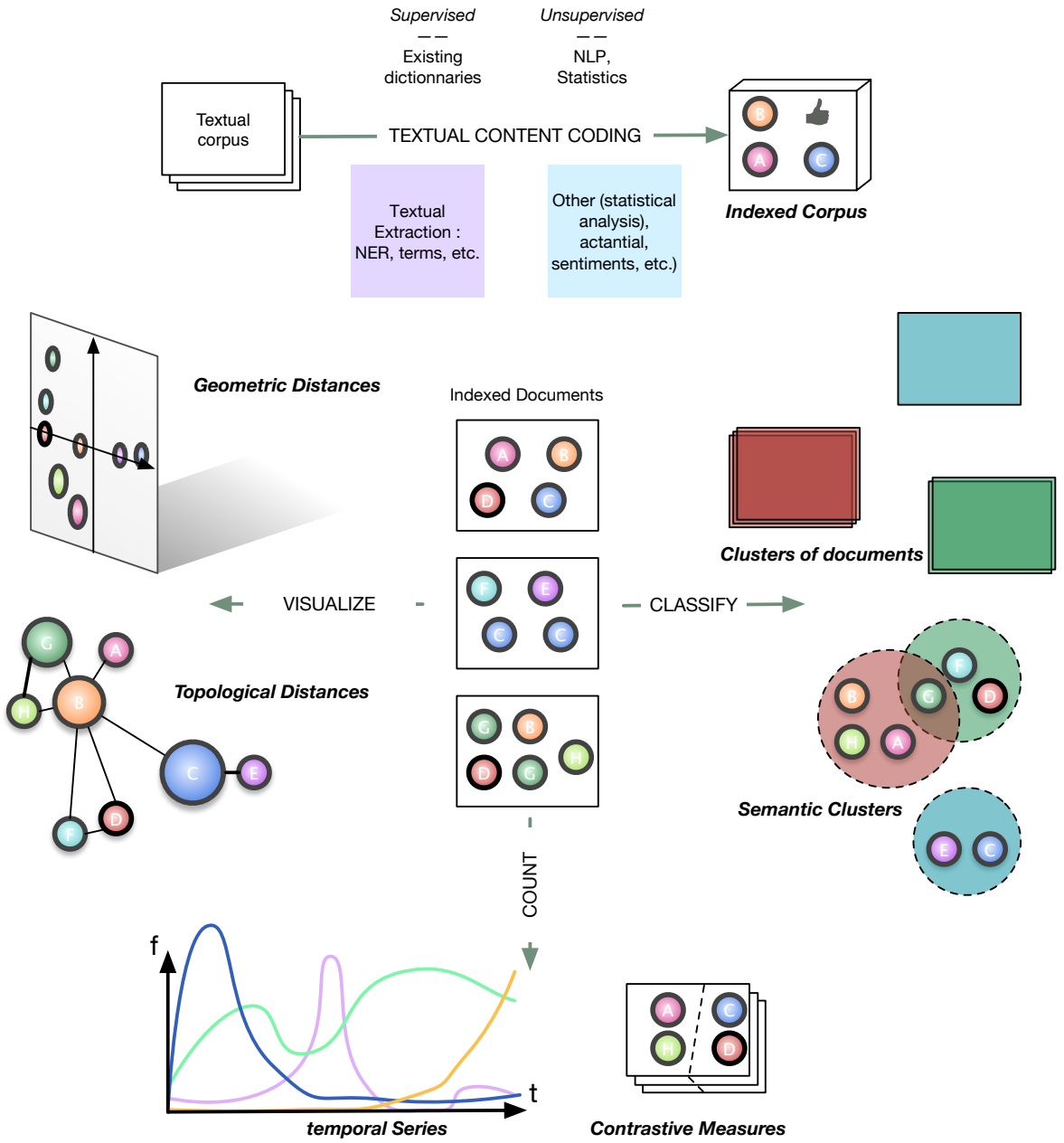
Arbre de texte



children, and jeopardize the promise of a secure retirement for future generations. But we can't ask senior citizens and working families to shoulder the entire burden of deficit reduction while asking nothing more from the wealthiest and most powerful. We won't grow the middle class simply by shifting the cost of health care or college onto families that are already struggling, or by forcing communities to lay off more teachers, cops, and firefighters. Most Americans -- Democrats, Republicans, and Independents -- understand that we can't just cut our way to prosperity. They know that broad-based economic growth requires a balanced approach to deficit reduction, with spending cuts and revenue, and with everybody doing their fair share. And that's the approach I offer tonight. On

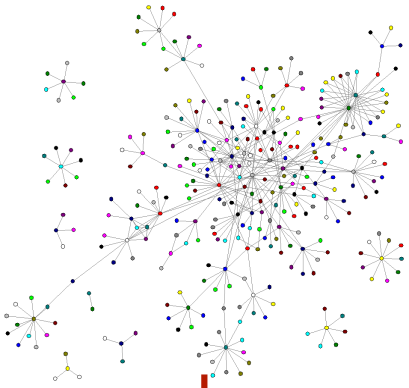
Corpus

Textual Corpus Processing

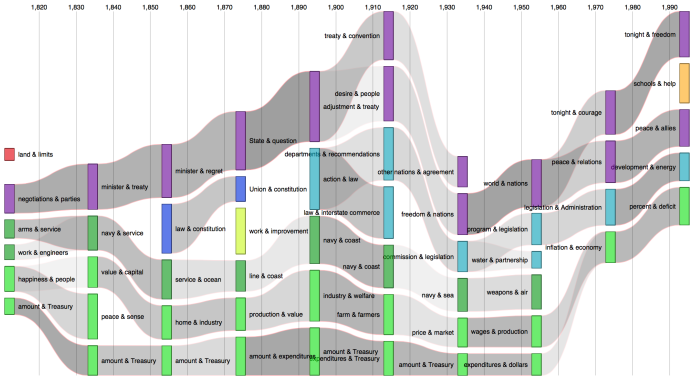


From traces to landscapes

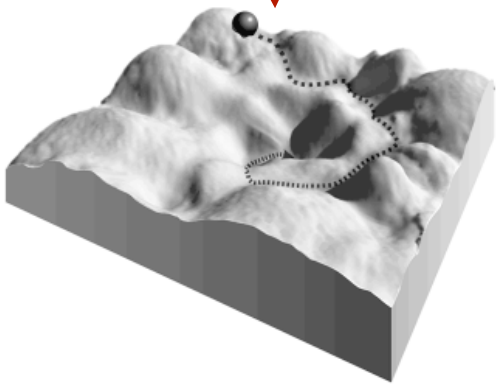
- Network as traces of local heterogeneous dynamics
- Reconstruction of network dynamics
- Landscape modeling of socio-semantic system



Network

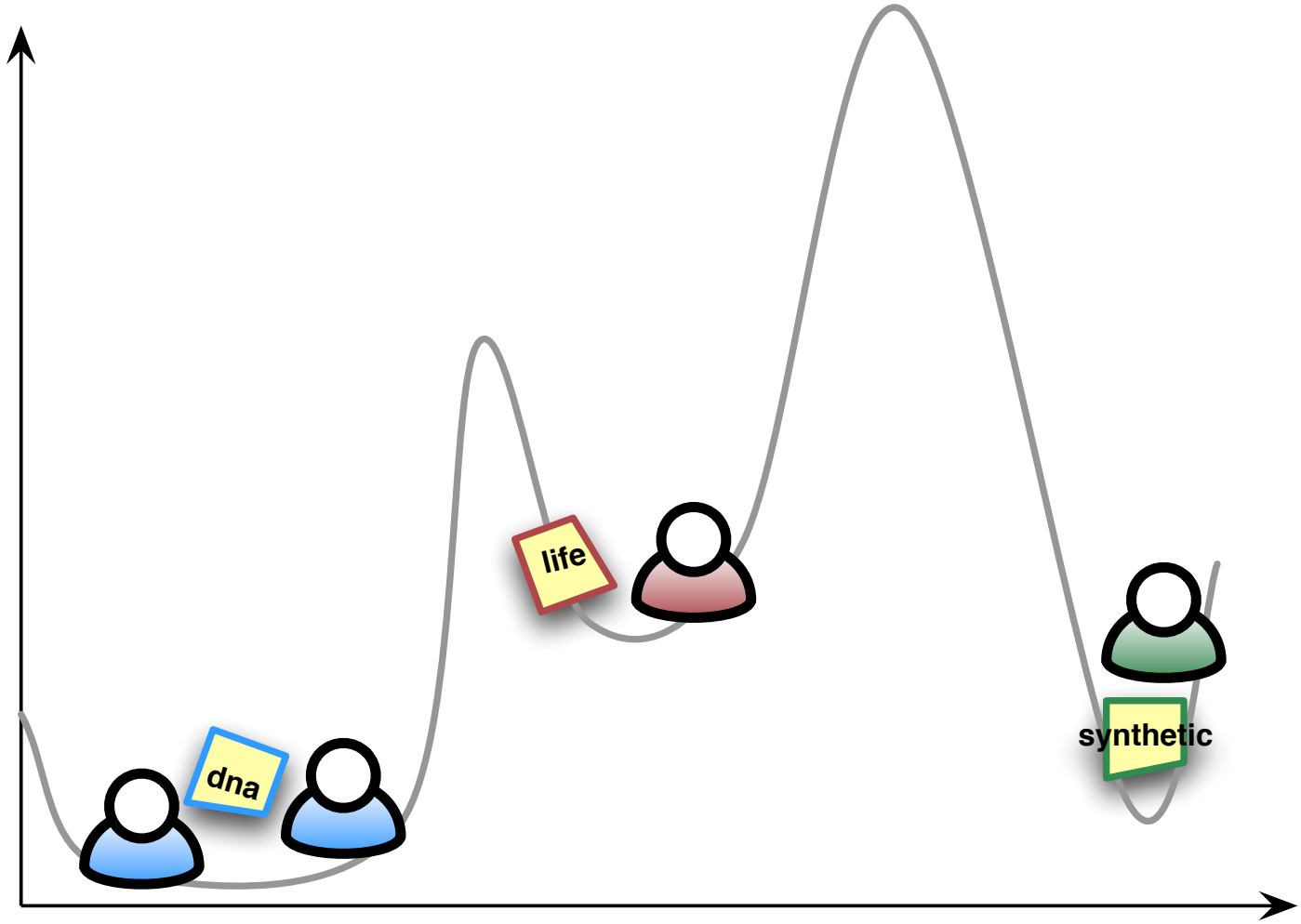


Streams



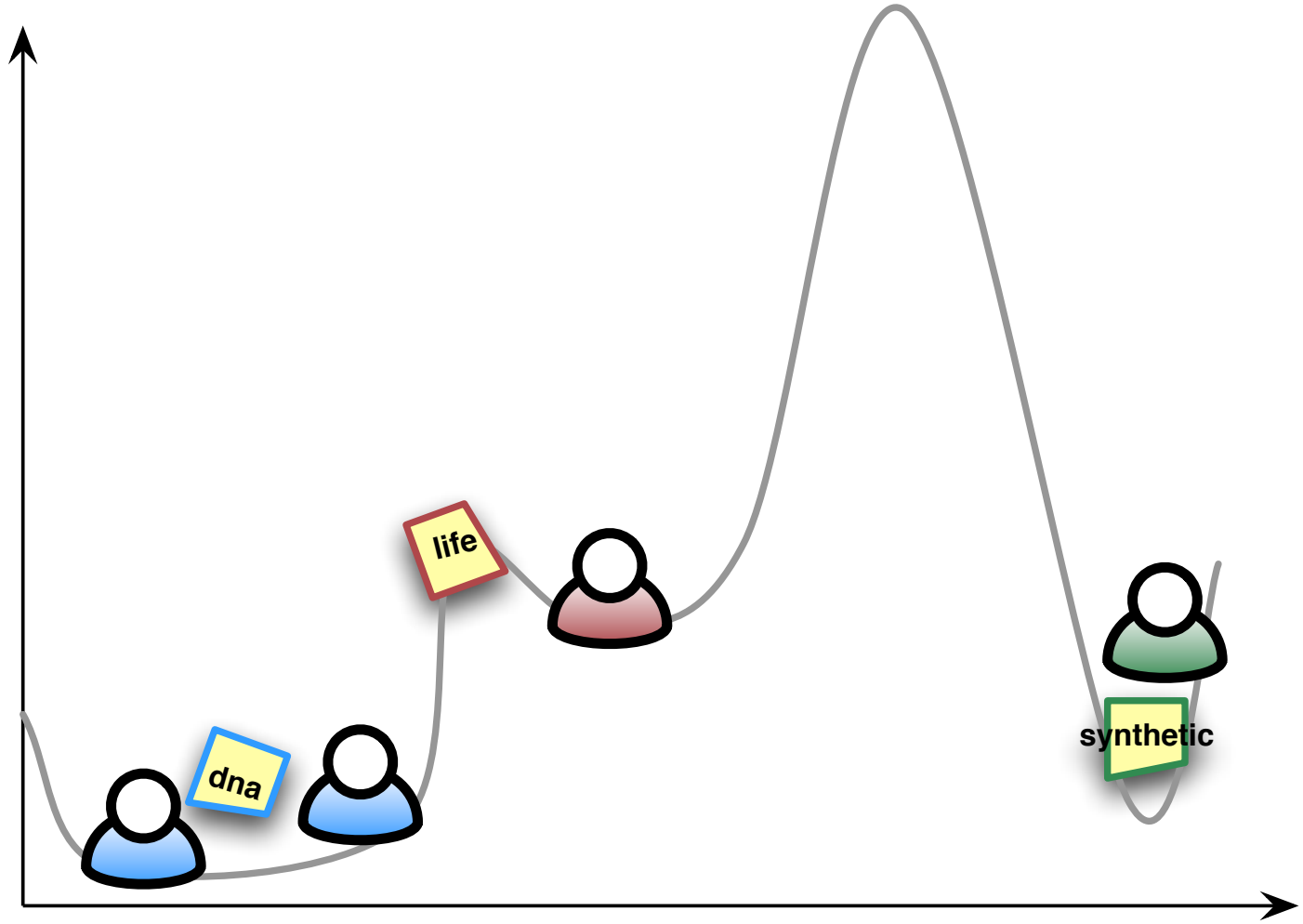
Landscape?

From traces to landscapes



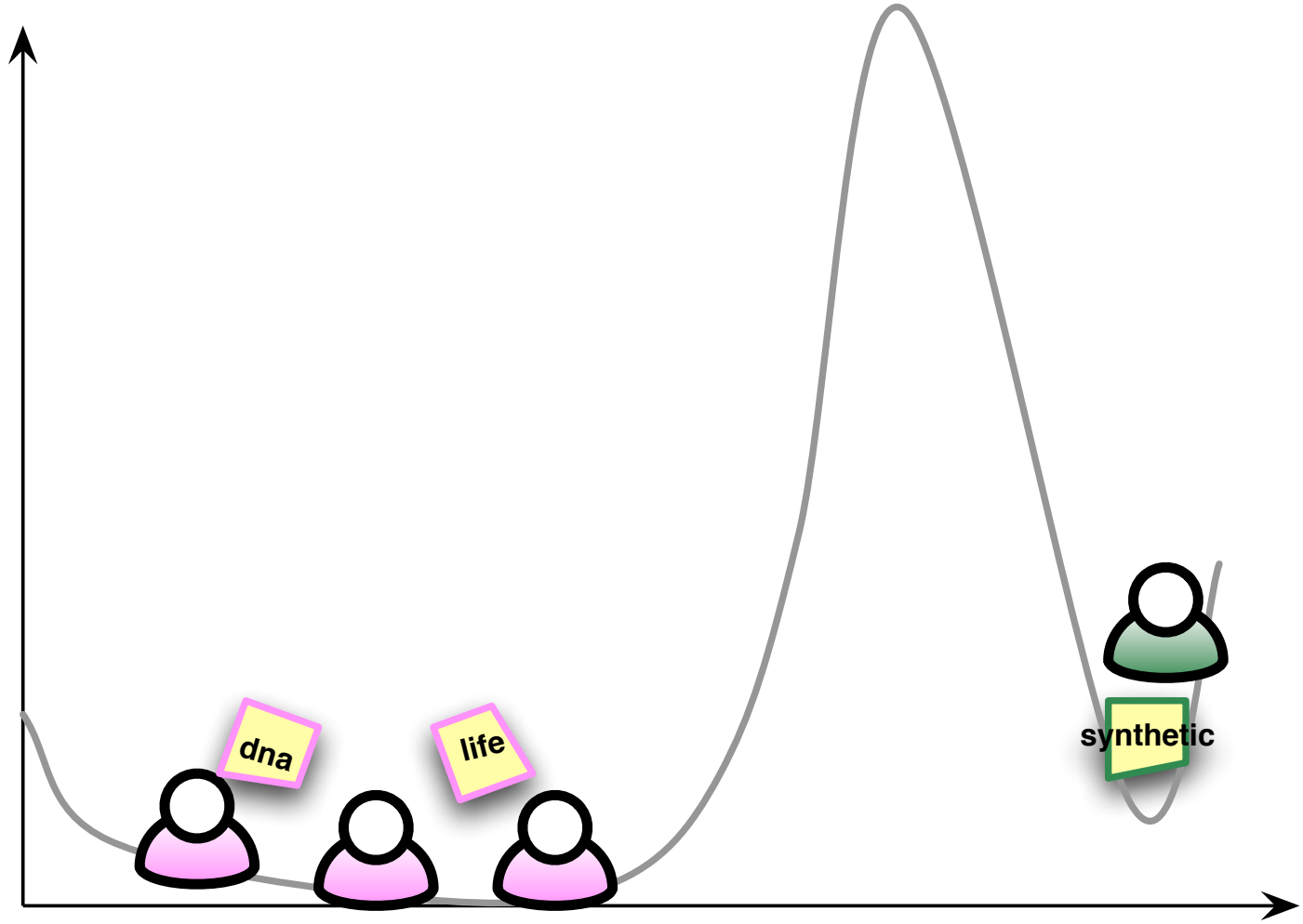
Landscape and its actors (t=1)

From traces to landscapes



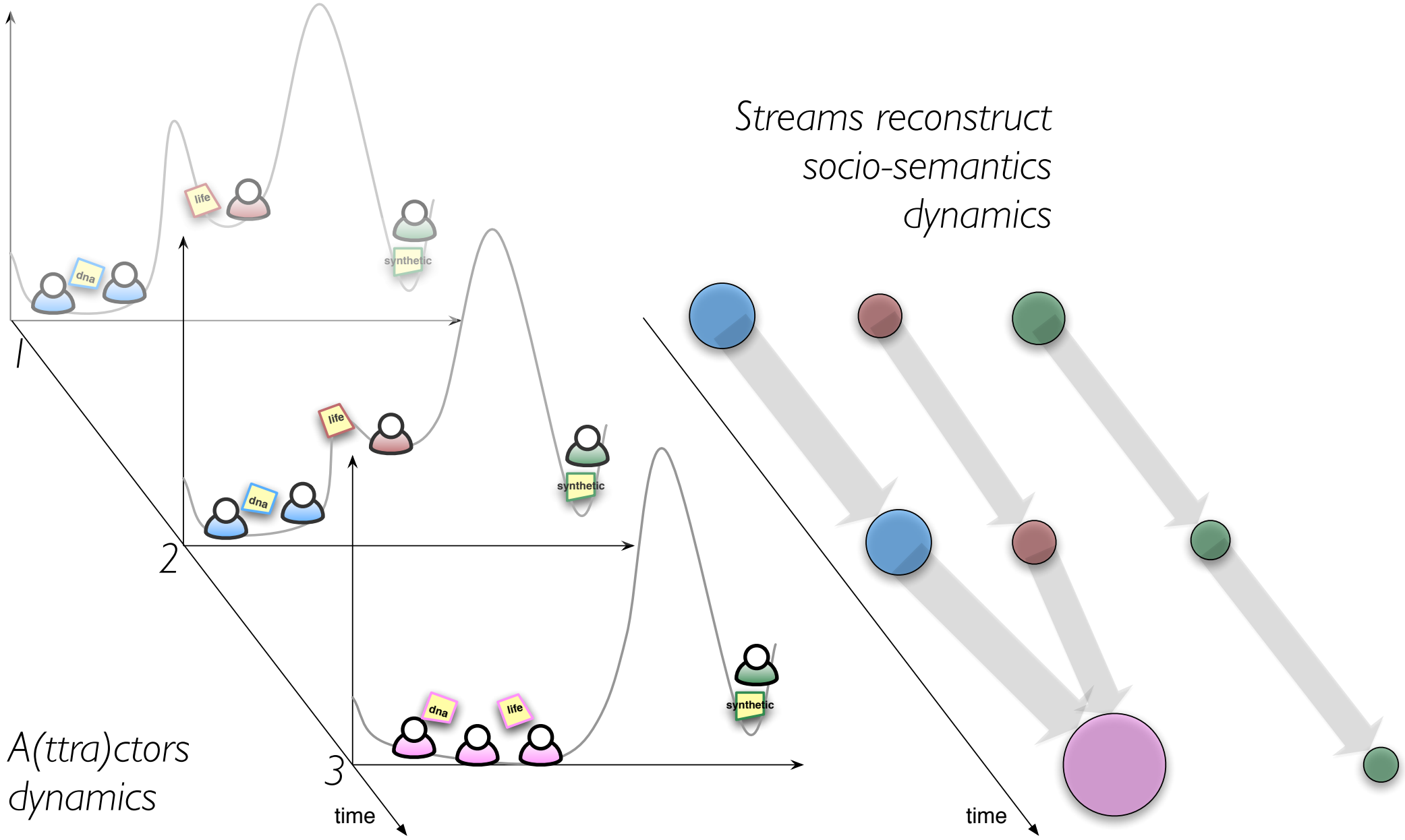
Landscape and its actors (t=2)

From traces to landscapes



Landscape and its actors (t=3)

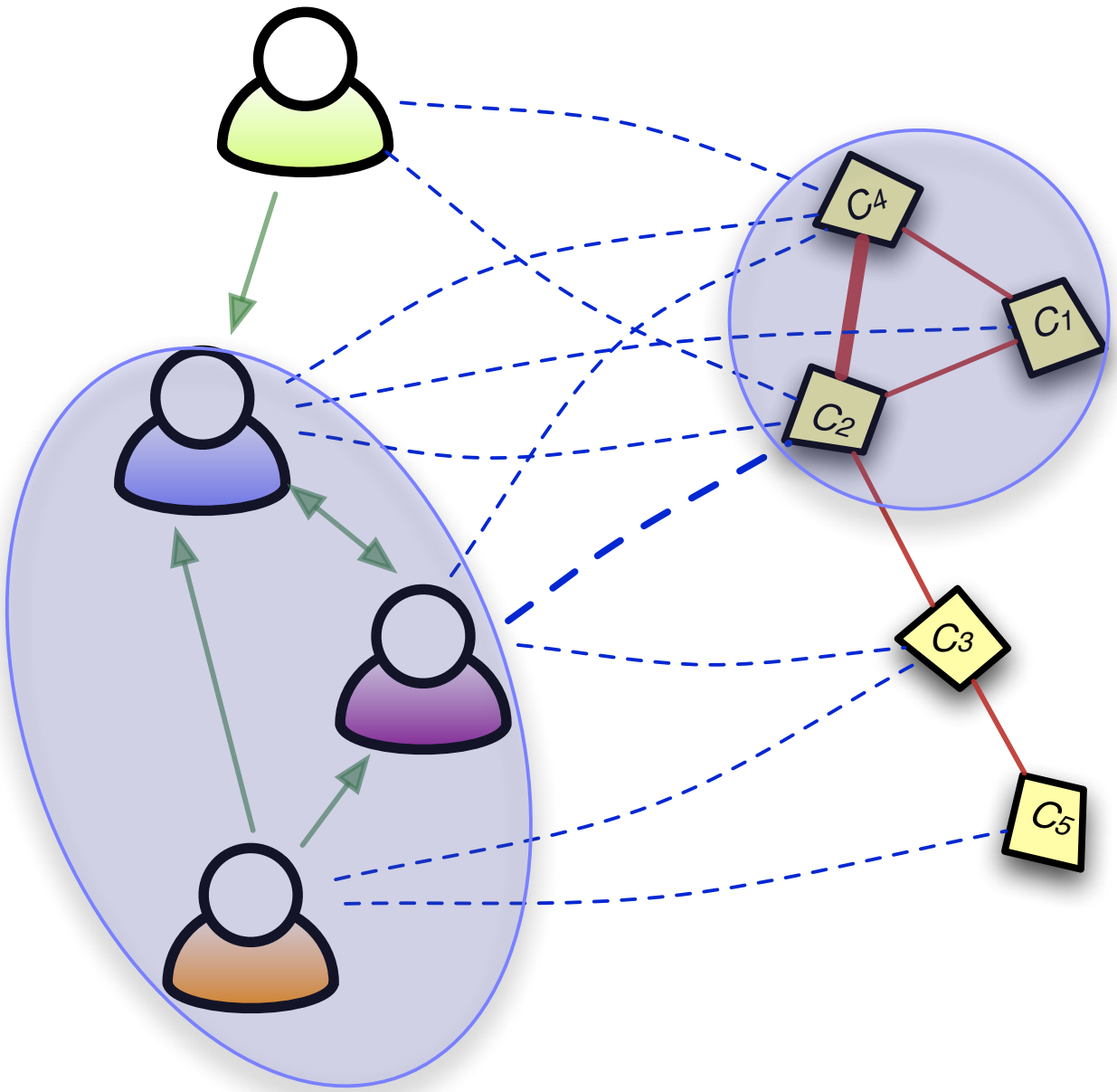
From traces to landscapes



Streams reconstruct socio-semantics dynamics

A(ttra)ctors dynamics

Heterogeneous cooccurrences



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

Understanding proximity metrics

It's all about context

" *Seule la proposition a un sens ; ce n'est que lié dans une proposition que le nom a une signification* "

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

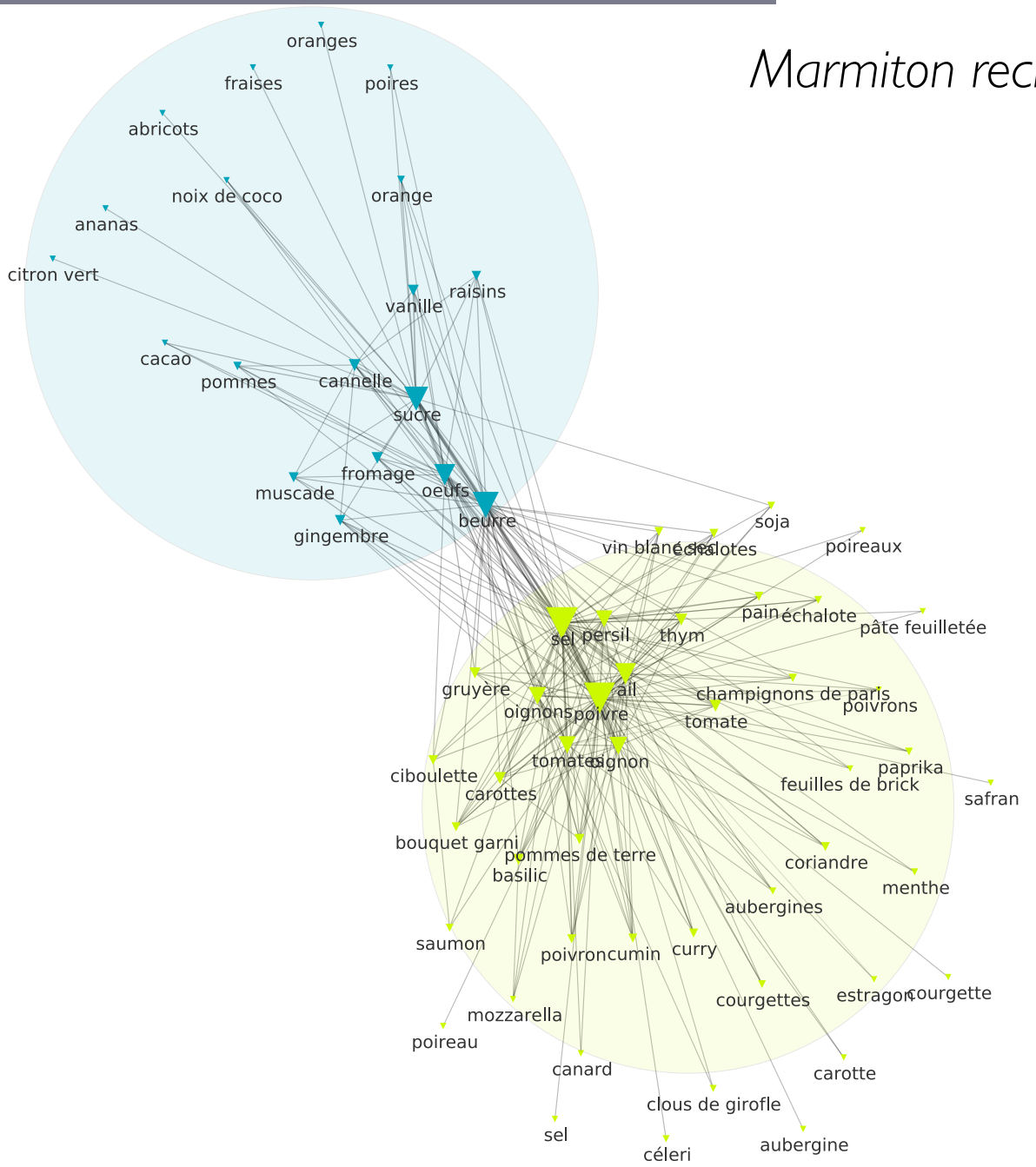
" *A word is characterized by the company it keeps* "

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

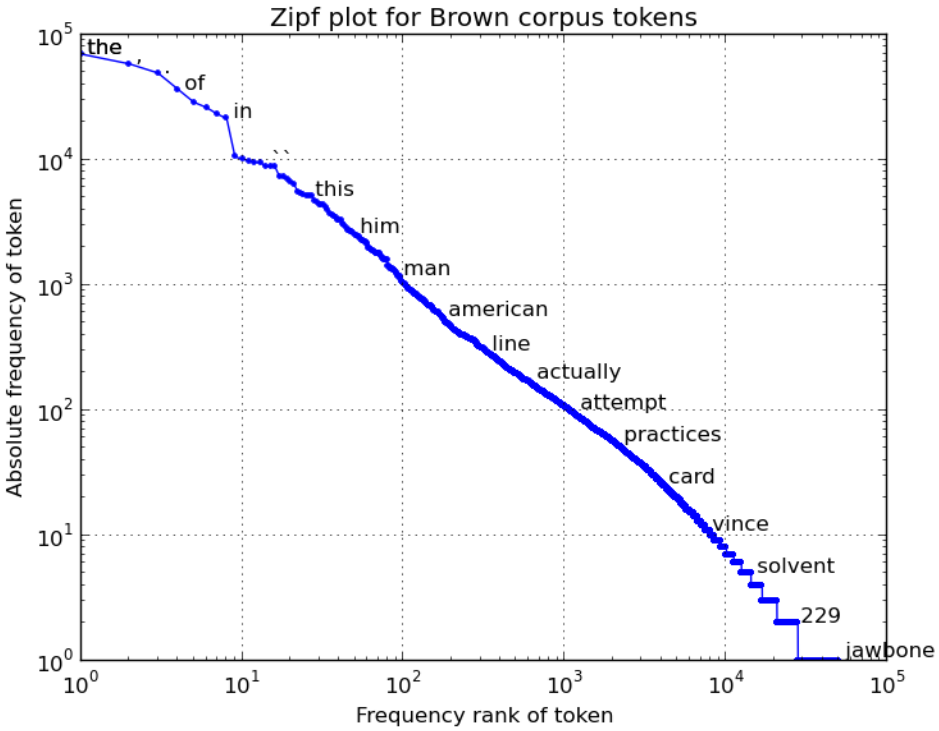
Metrics

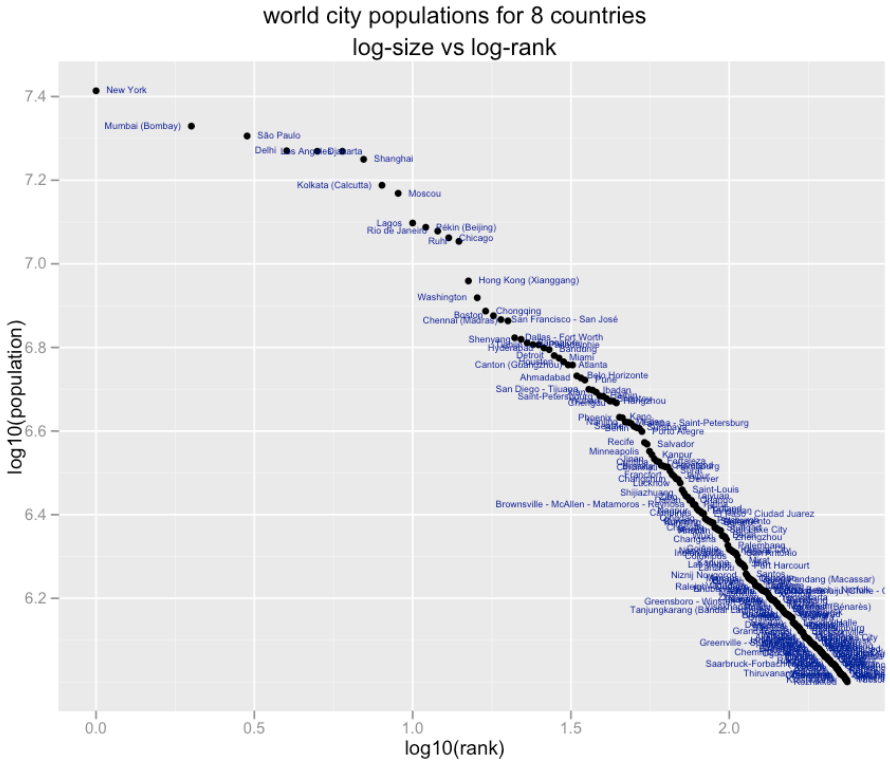
Marmiton recipes

- Raw cooccurrence measure



Heterogeneity curse

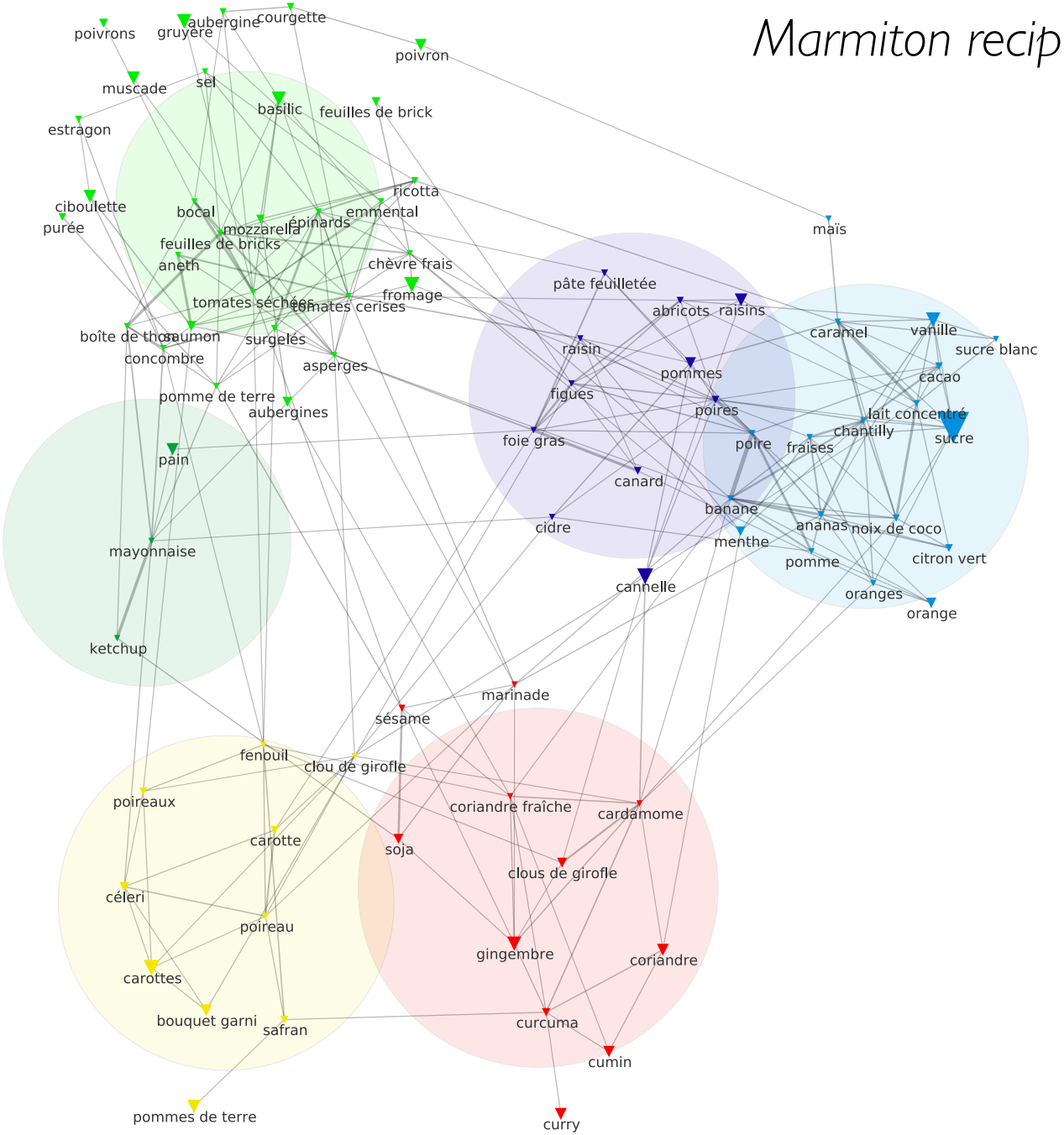




Metrics

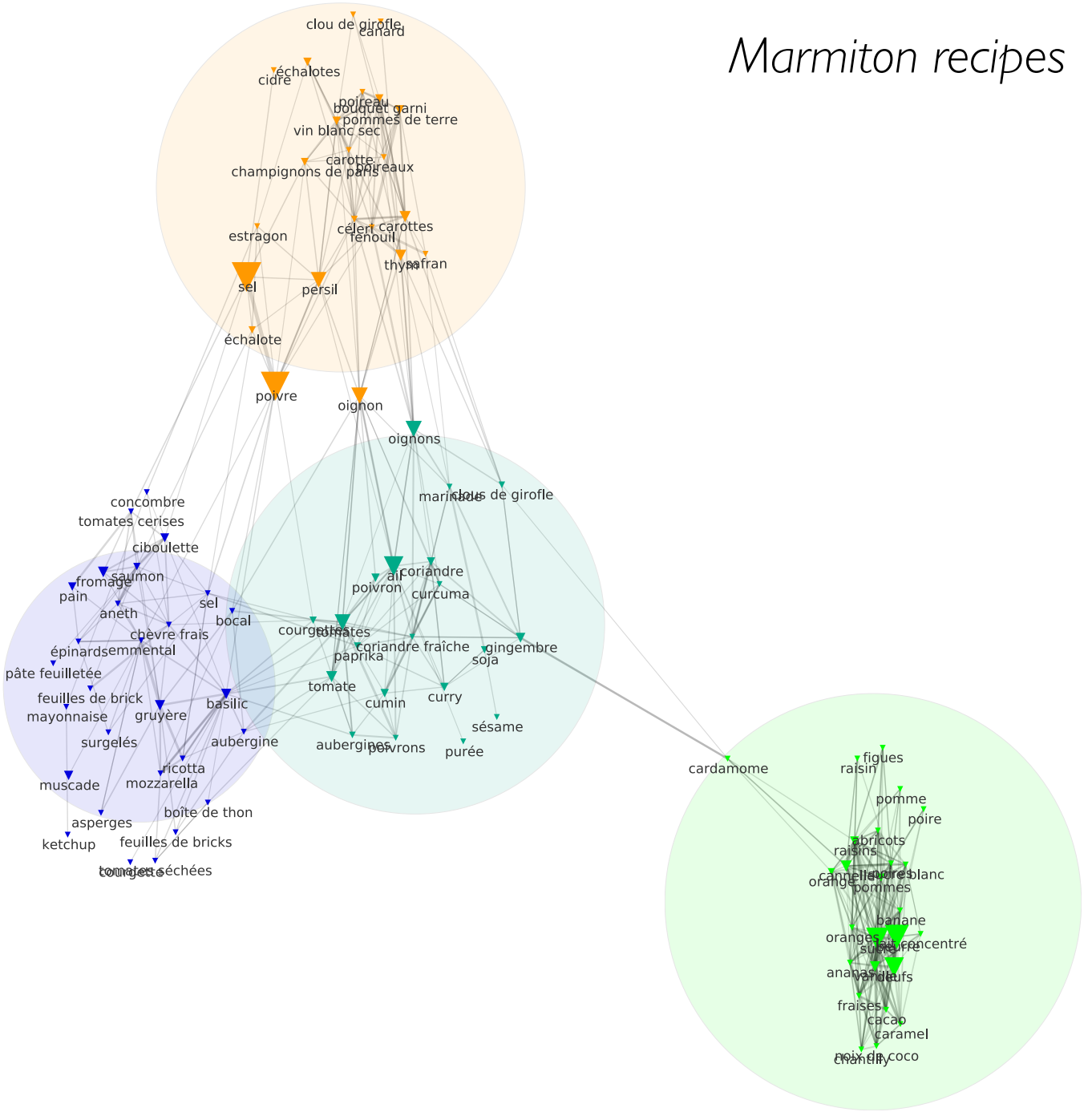
- Cramer

Marmiton recipes



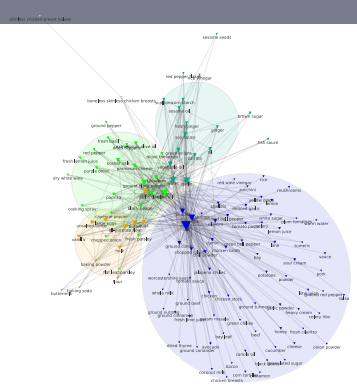
Metrics

- Distributional Measure

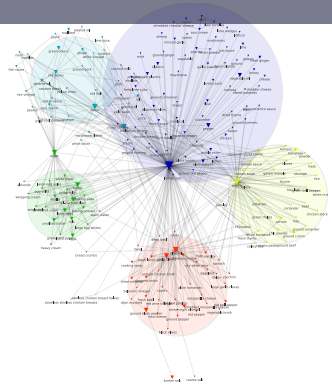


Marmiton recipes

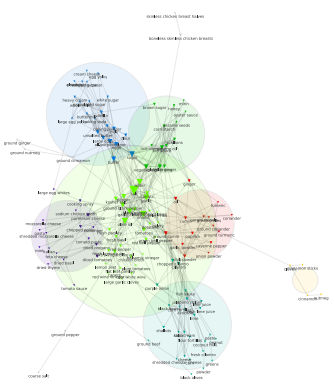
Metrics



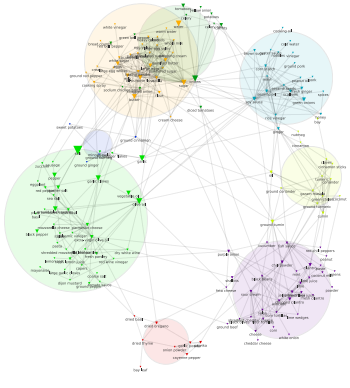
coocurrences brutes ($Q = 0.2$)



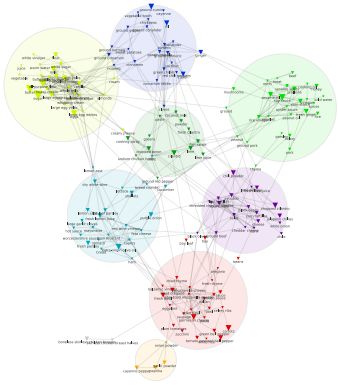
indice d'inclusion ($Q = 0.38$)



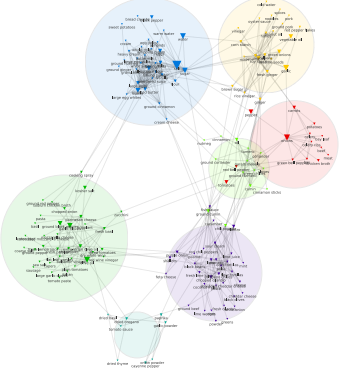
coefficient de Jaccard ($Q = 0.56$)



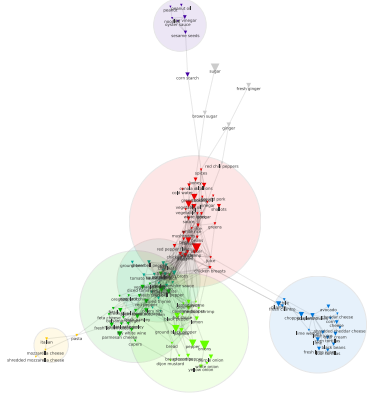
χ^2 ($Q = 0.70$)



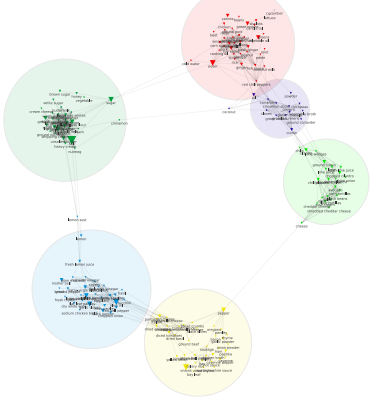
IM ($Q = 0.69$)



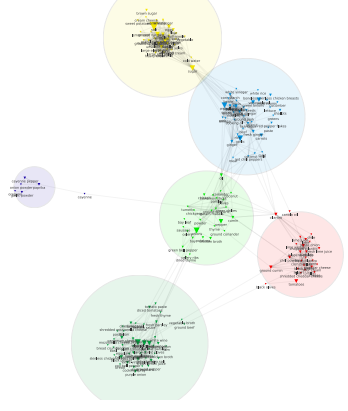
(LLR) ($Q = 0.67$)



cosine ($Q = 0.52$)



s_dIM ($Q = 0.75$)



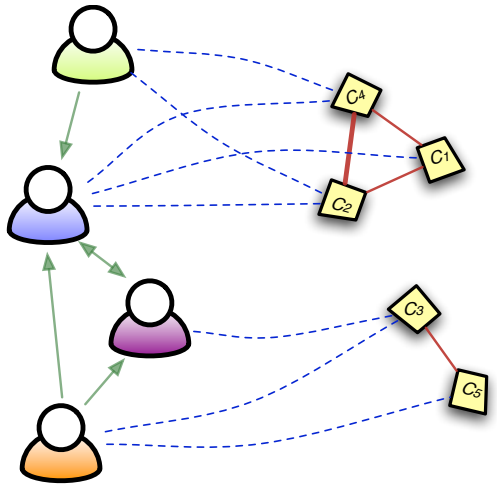
s_dLLR ($Q = 0.73$)

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 appearences of two items in the same document. $C = O^t O$

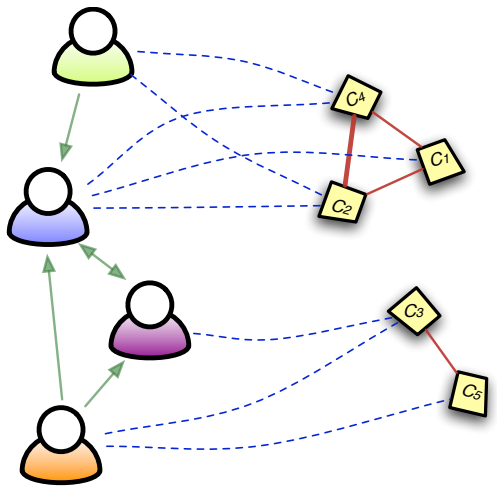


documents / items	A	B	C	D
C1		1		
C2	1	1		
C3			1	1
C4	1	1		
C5				1

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 appearences 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 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 coccurrences : $e_{ij} = \frac{s_i s_j}{N}$

	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 coccurrences : $e_{ij} = \frac{s_i s_j}{N}$

	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 coccurrences : $e_{ij} = \frac{s_i s_j}{N}$

	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 coccurrences : $e_{ij} = \frac{s_i s_j}{N}$

	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 coccurrences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- *Raw cooccurrences :* $S_R(i,j) = c_{ij}$
- *Association strength :* $S_A(i,j) = \frac{c_{ij}}{s_i s_j}$
- *Mutual Information :* $S_{MI}(i,j) = \log(\frac{c_{ij}}{e_{ij}})$
- *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 coccurrences : $e_{ij} = \frac{s_i s_j}{N}$

Indirect Measures of Similarity :

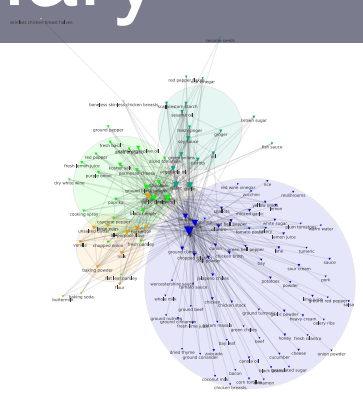
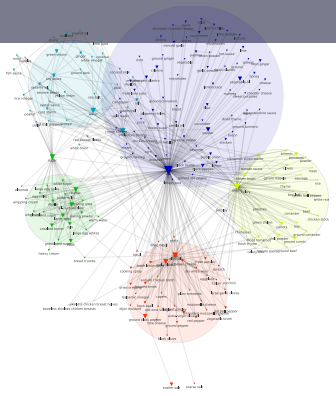
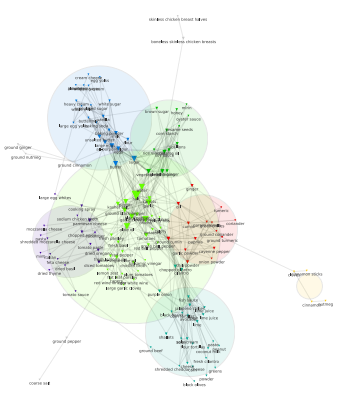
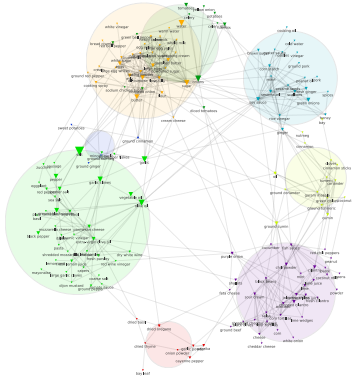
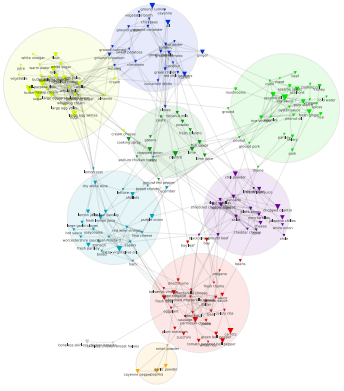
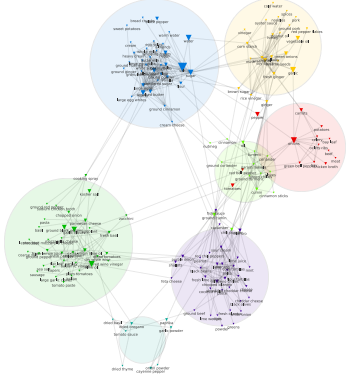
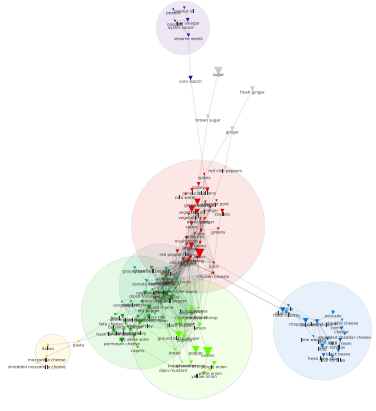
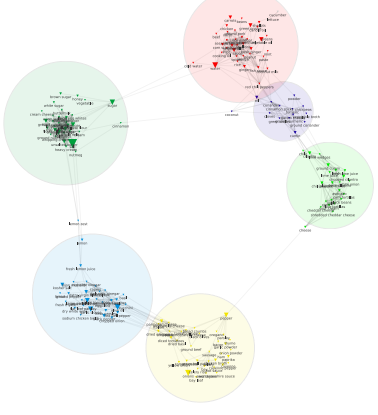
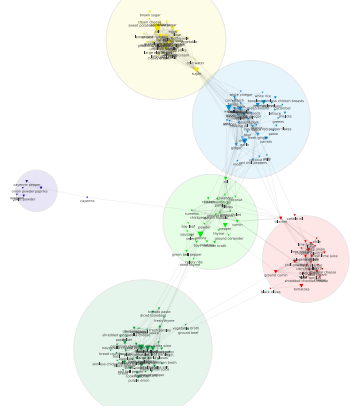
- Cosine (distributional) :

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

- 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

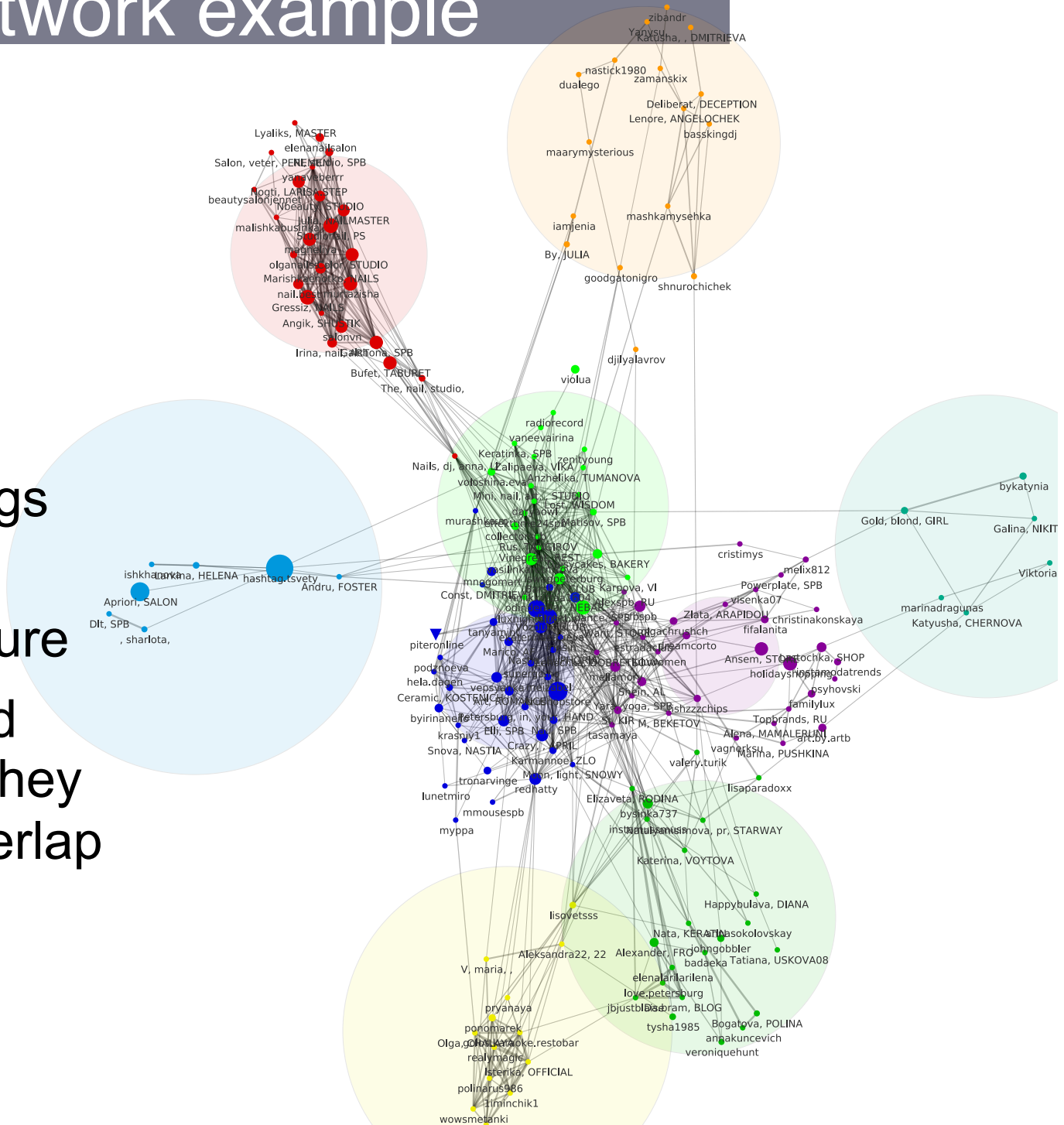
Co-word	 coocurrences brutes ($Q = 0.2$)	 indice d'inclusion ($Q = 0.38$)	 coefficient de Jaccard ($Q = 0.56$)
Direct (syntagmatic)	 χ^2 ($Q = 0.70$)	 IM ($Q = 0.69$)	 (LLR) ($Q = 0.67$)
Indirect (paradigmatic)	 cosine ($Q = 0.52$)	 s_dIM ($Q = 0.75$)	 s_dLLR ($Q = 0.73$)

Metrics Summary

proximity measures	type of network	normalisation	special properties
raw	interaction network (*e.g.* 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

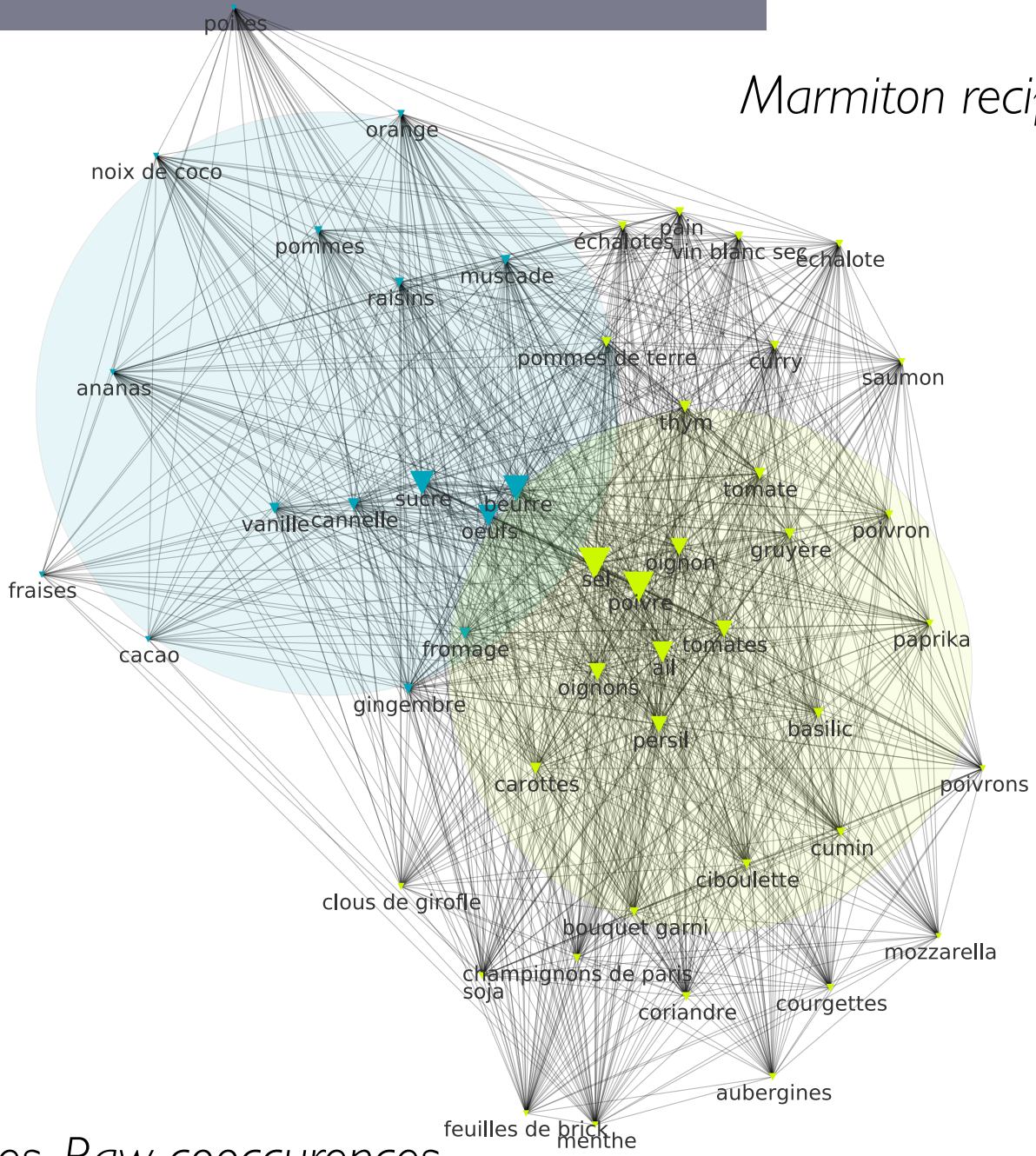
- parameters:
  - –Field1: Users
  - –Field 2: Hashtags
  - – cosine_het proximity measure
- Two Users are linked when the hashtags they use have a large overlap



Filtering

- No filtering

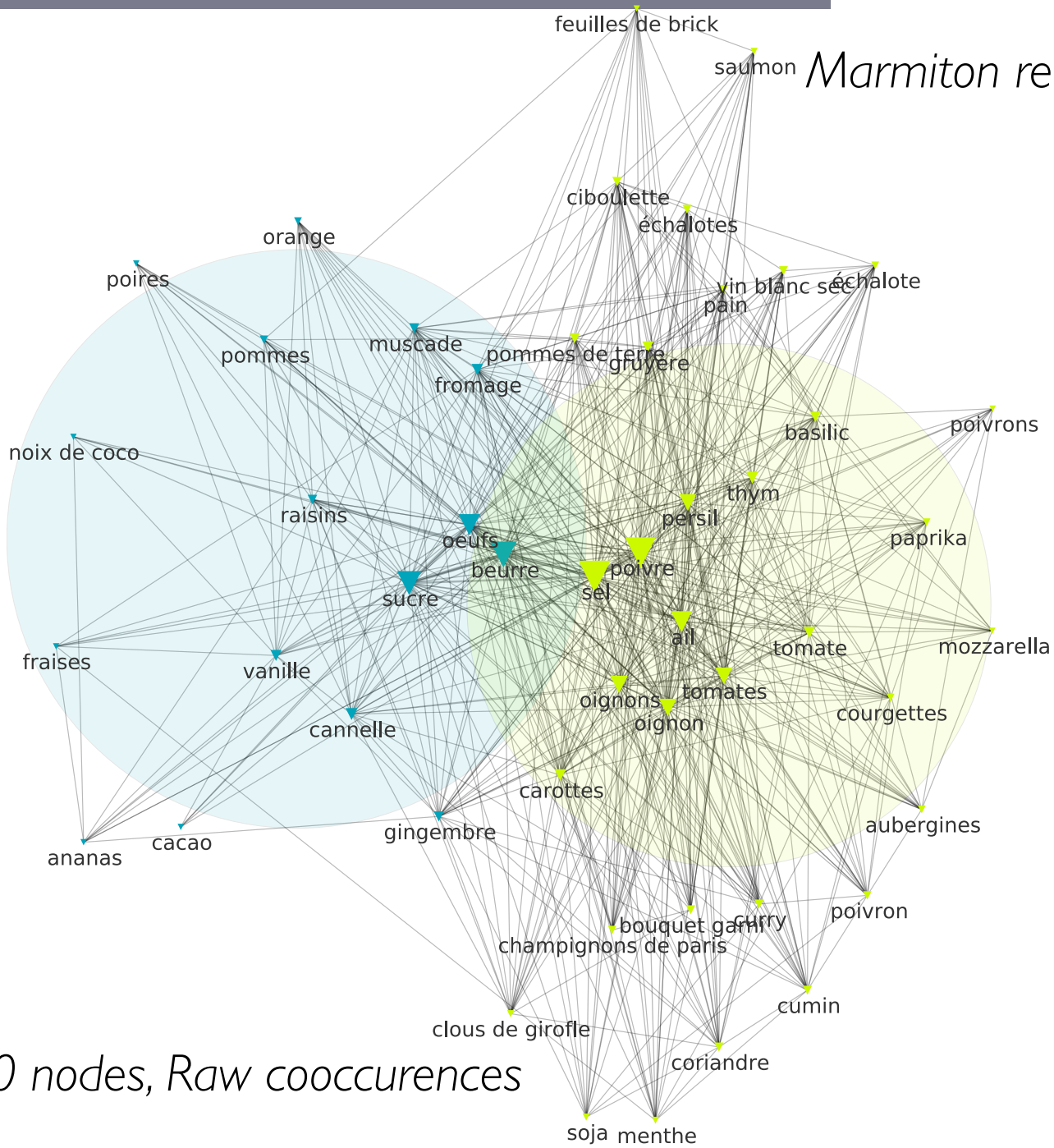
Marmiton recipes



50 nodes, Raw cooccurences

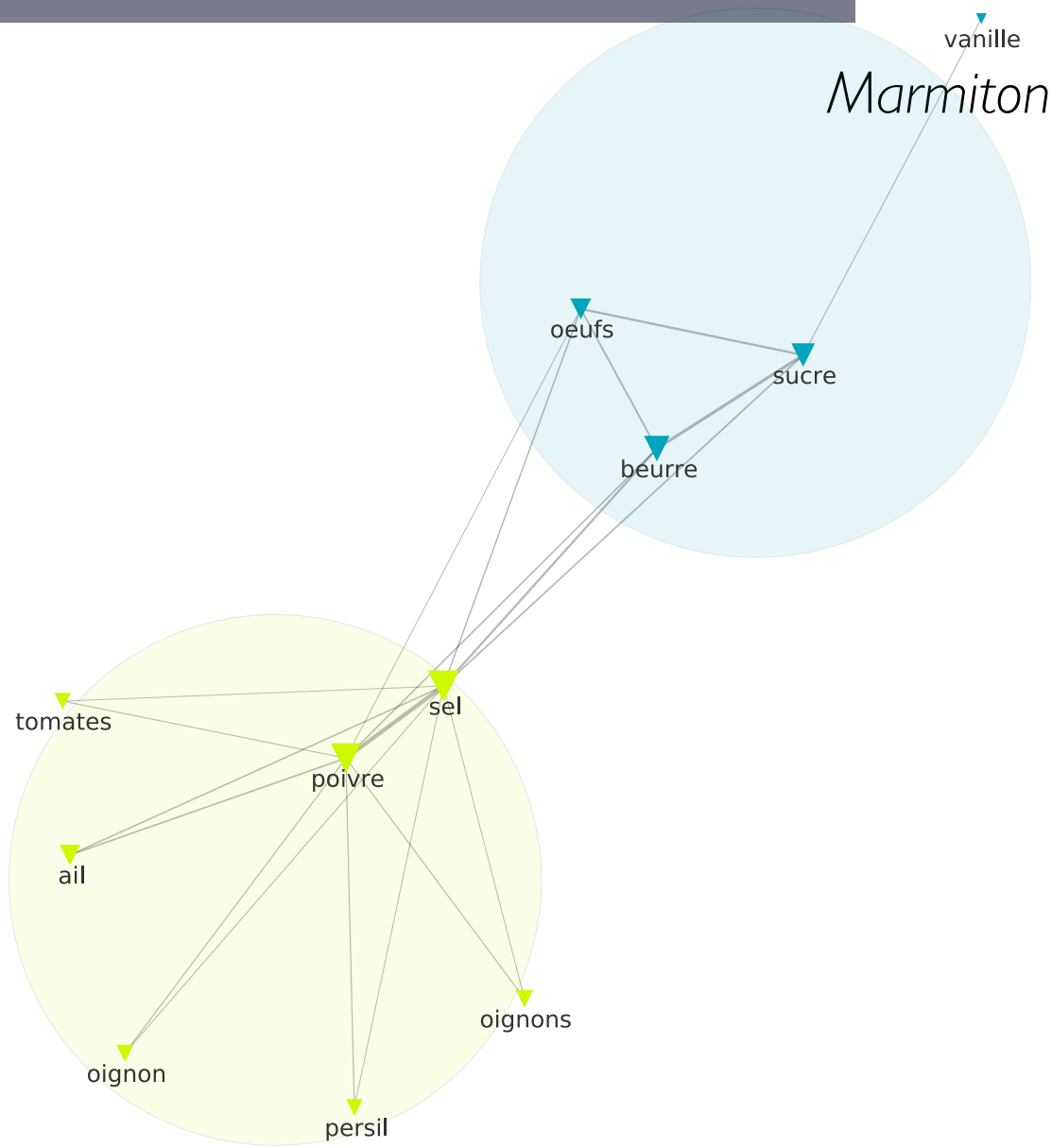
Filtering

- filtering on edge strength values: 10 cooc at least



Filtering

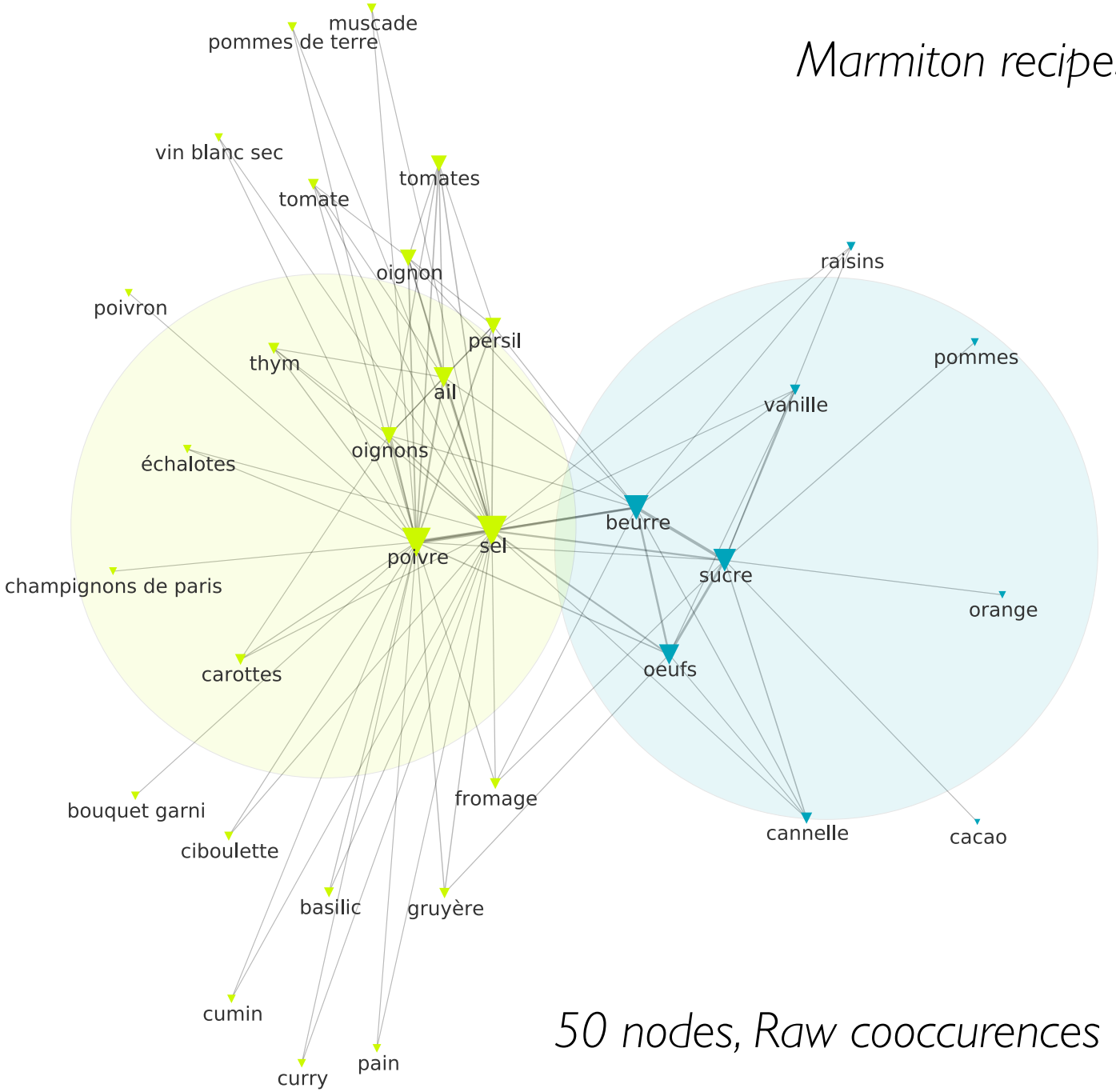
- filtering on the number of edges: 20 Top edges



50 nodes, Raw cooccurences

Filtering

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

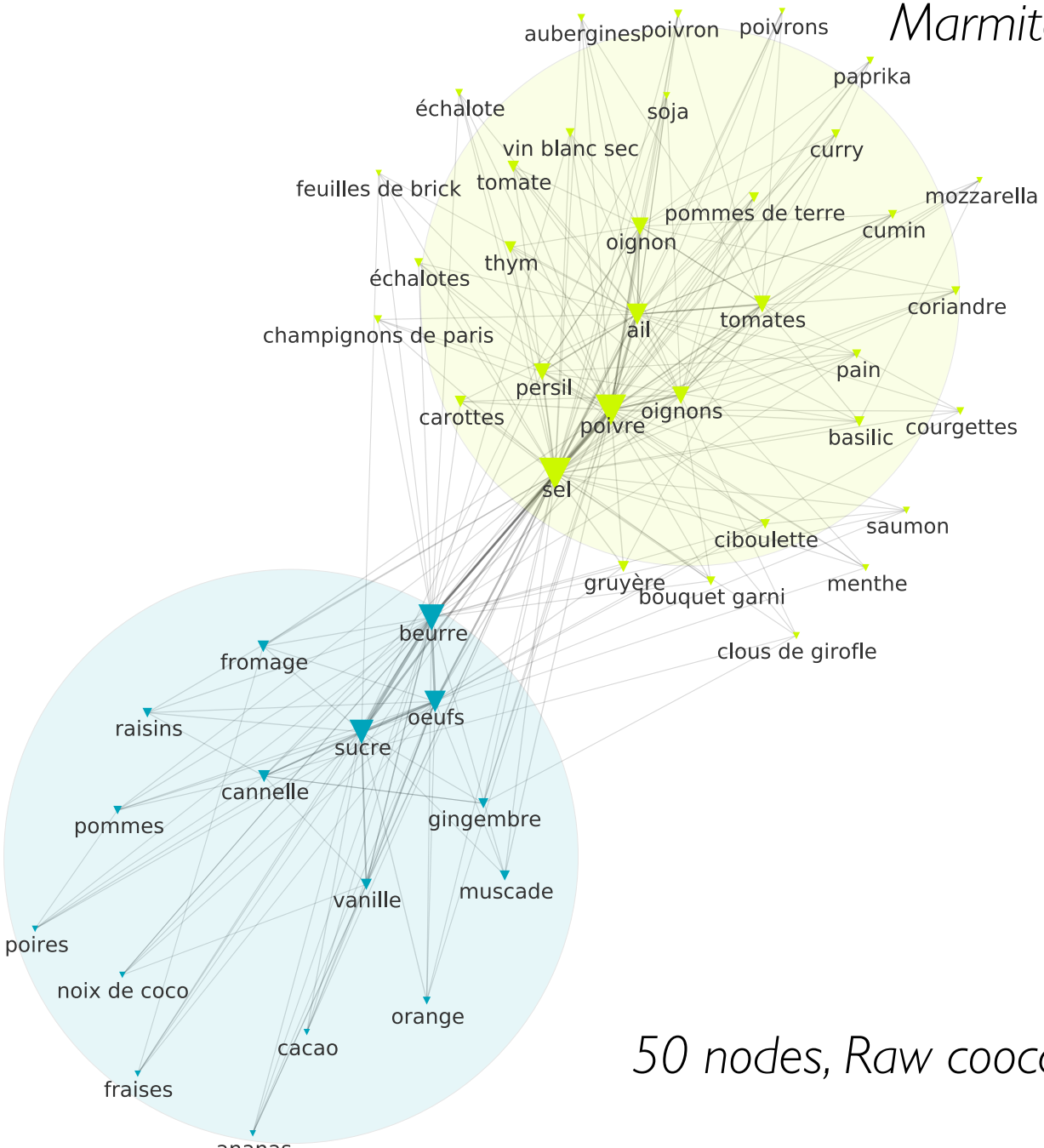


Marmiton recipes

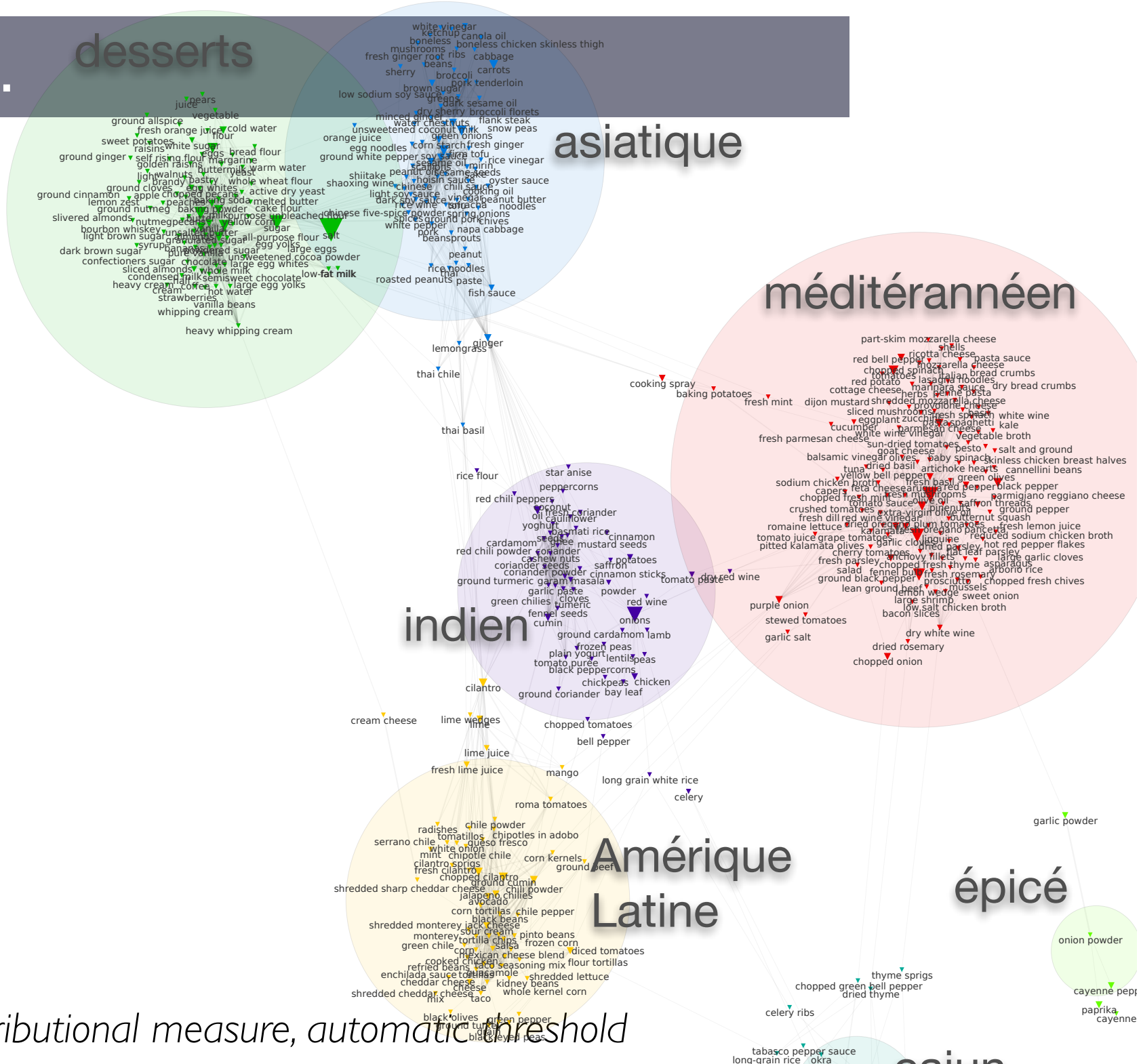
50 nodes, Raw cooccurences

Filtering

- Local Filtering: 5 top neighbours



Recipes…

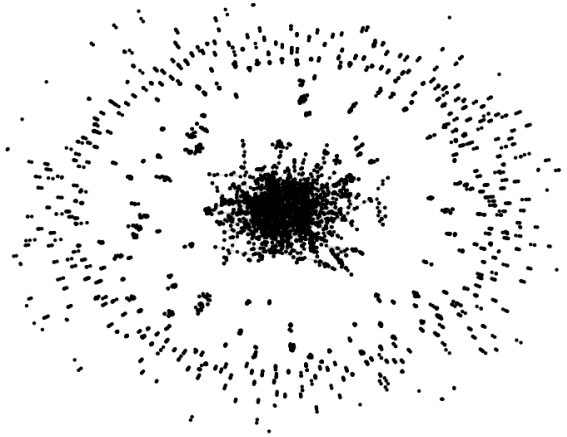


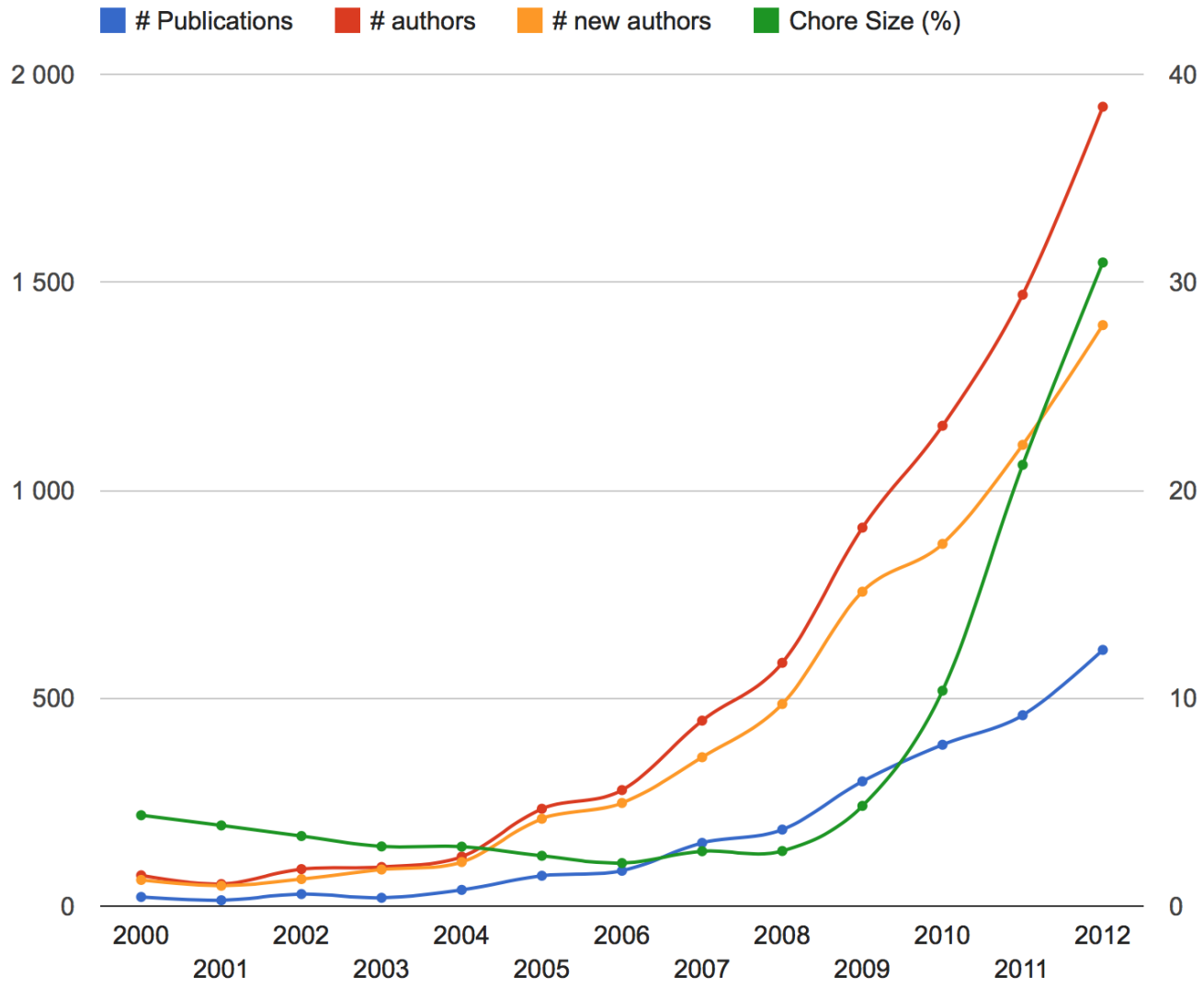
500 nodes, Distributional measure, automatic threshold

Synthetic Biology Emergence

Collaboration Network Dynamics

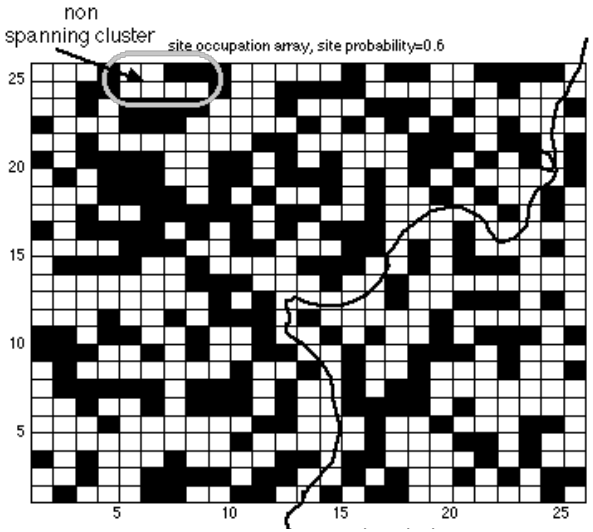
- Unstabilized field at the crossroads between different disciplinary origins
- Exponential growth of publications with a very high rate of newcomers
- Although a central community is emerging

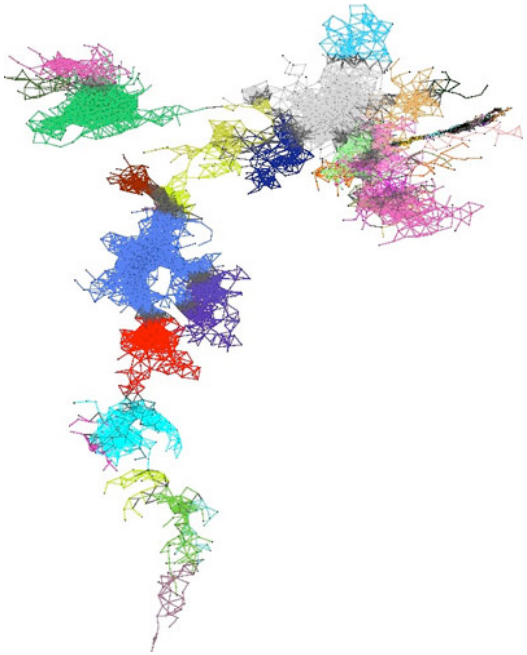


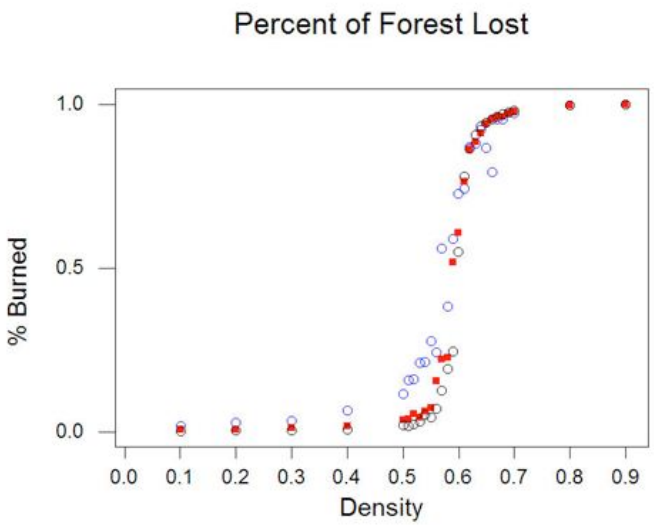


Collaboration Network Dynamics

- 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



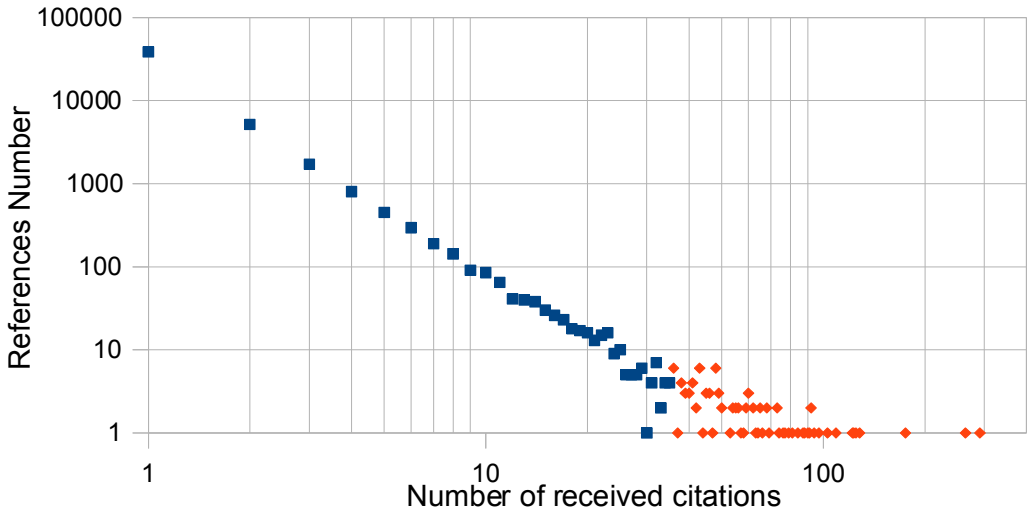




«percolation threshold» above which the giant connected component collapse in disconnected subgraphs

Scientific references co-citation map

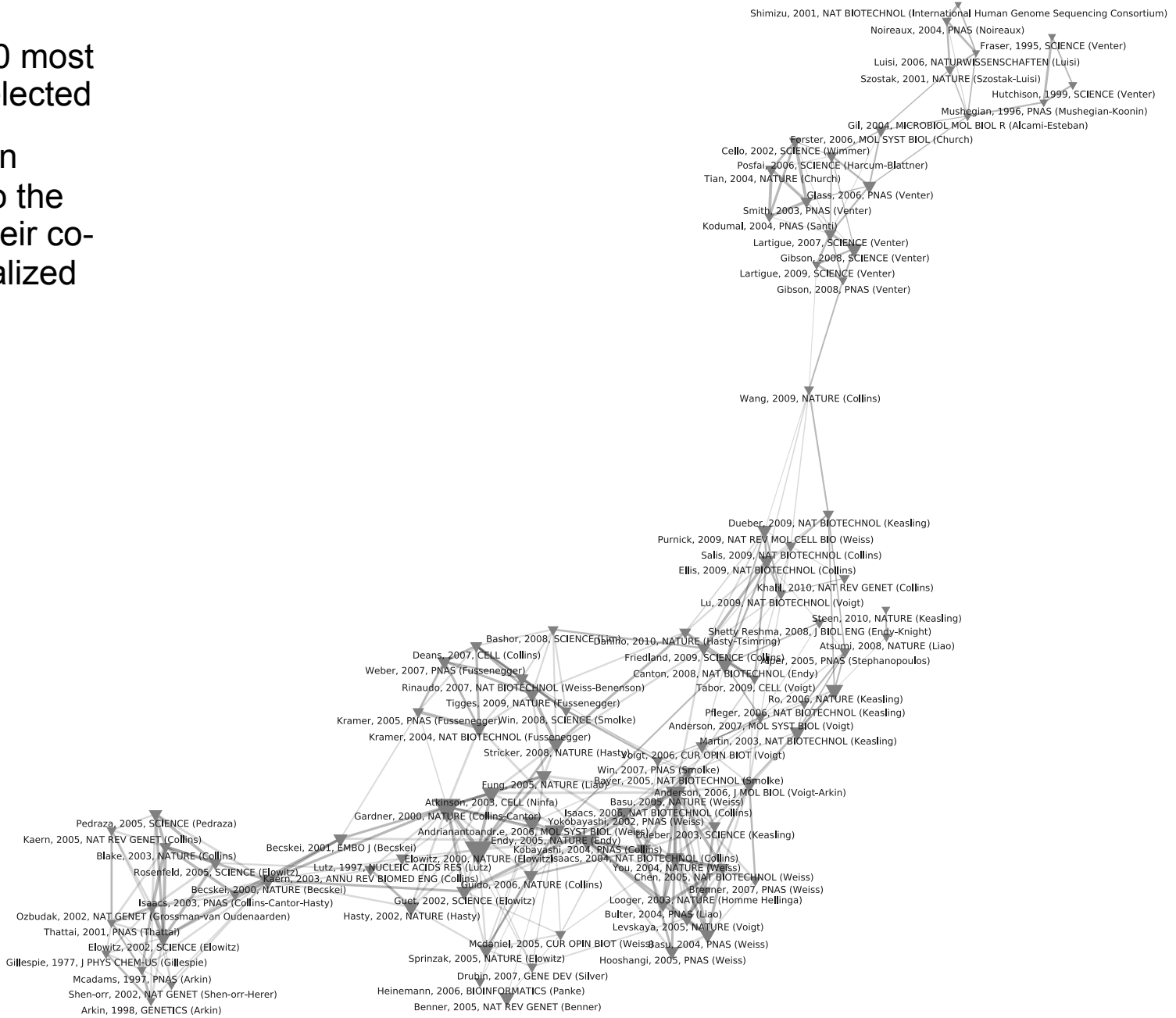
- **Selection**: only the 100 most cited references are selected



Received citations distributions

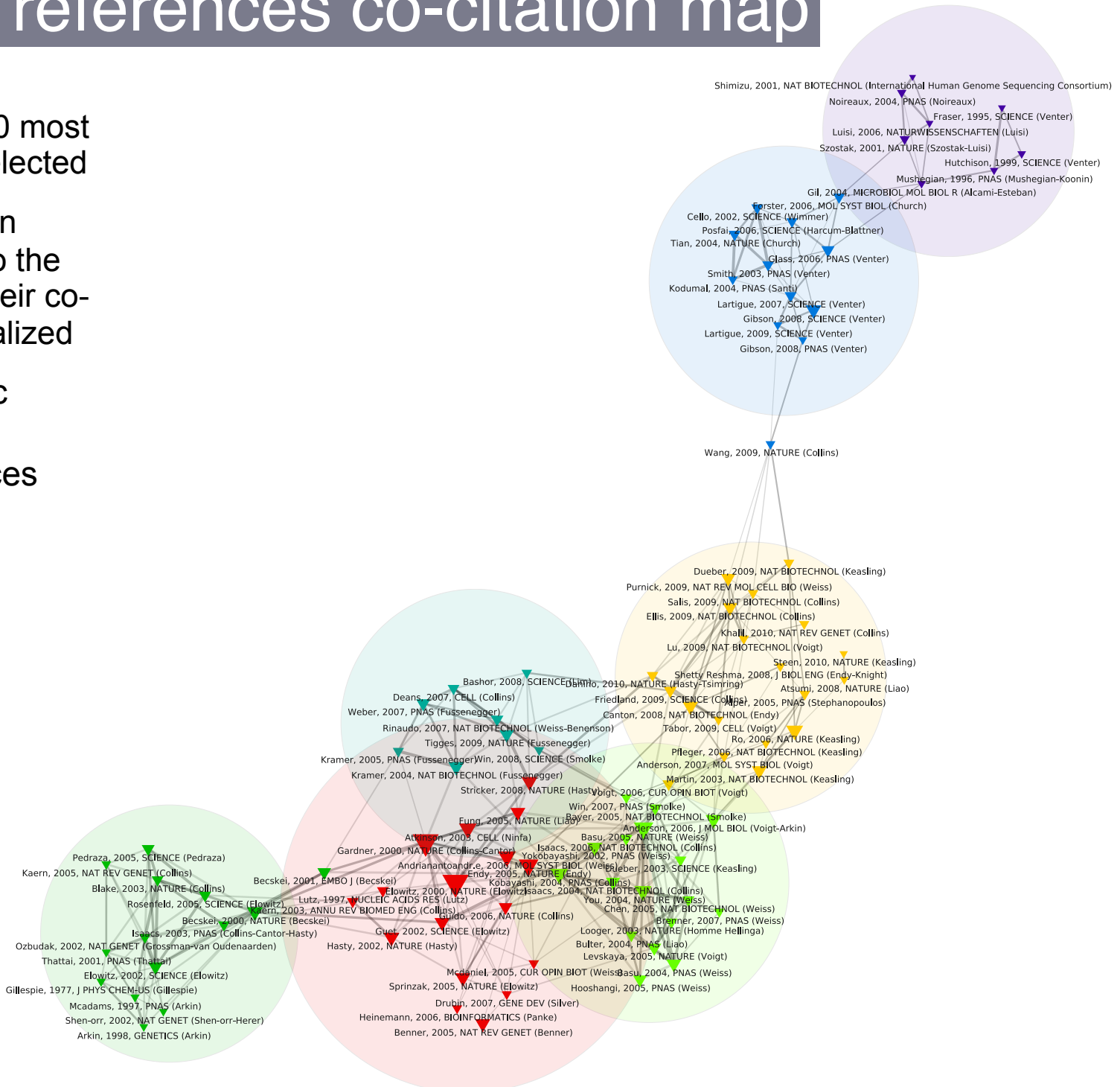
Scientific references co-citation map

- **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



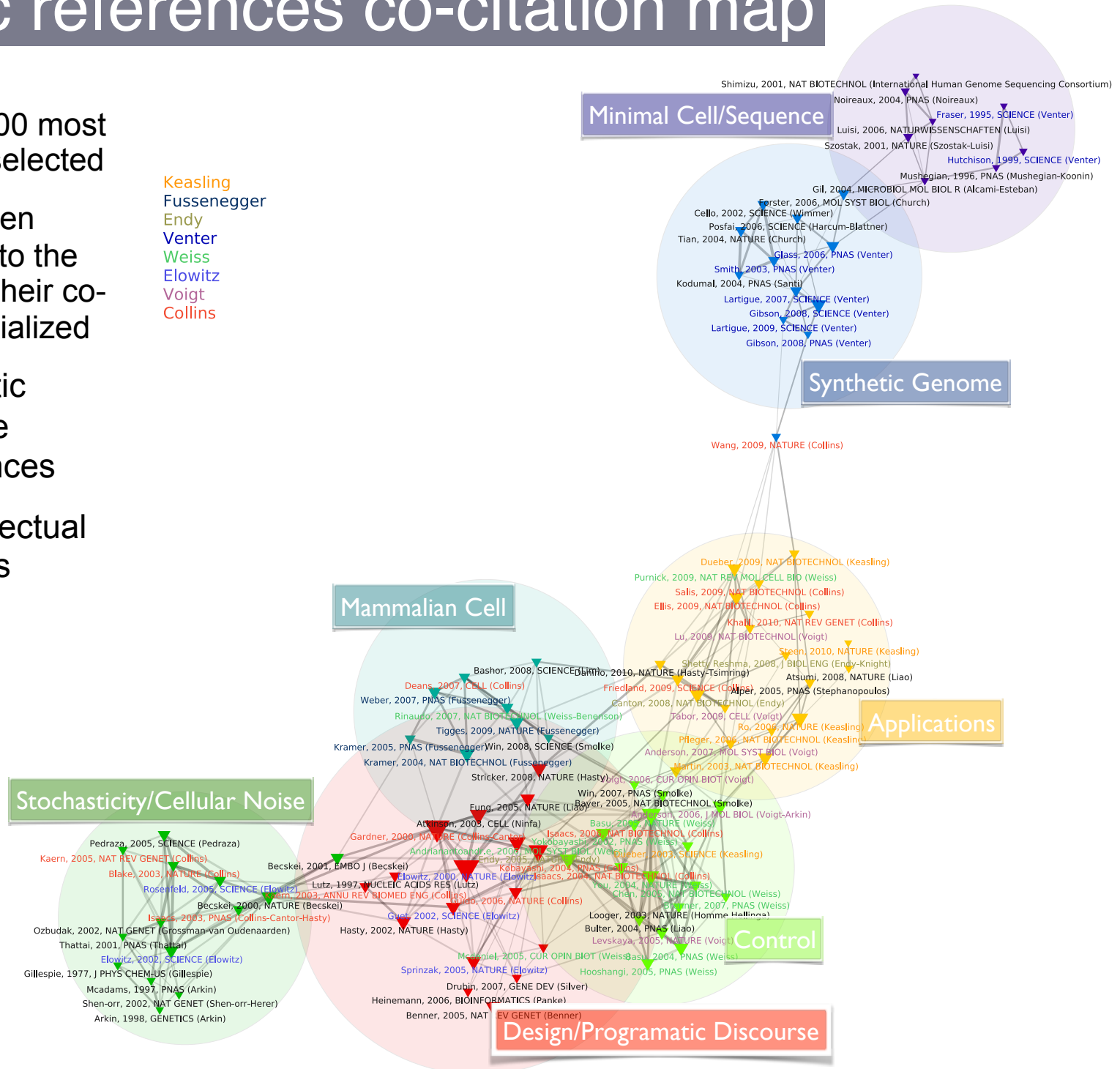
Scientific references co-citation map

- **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



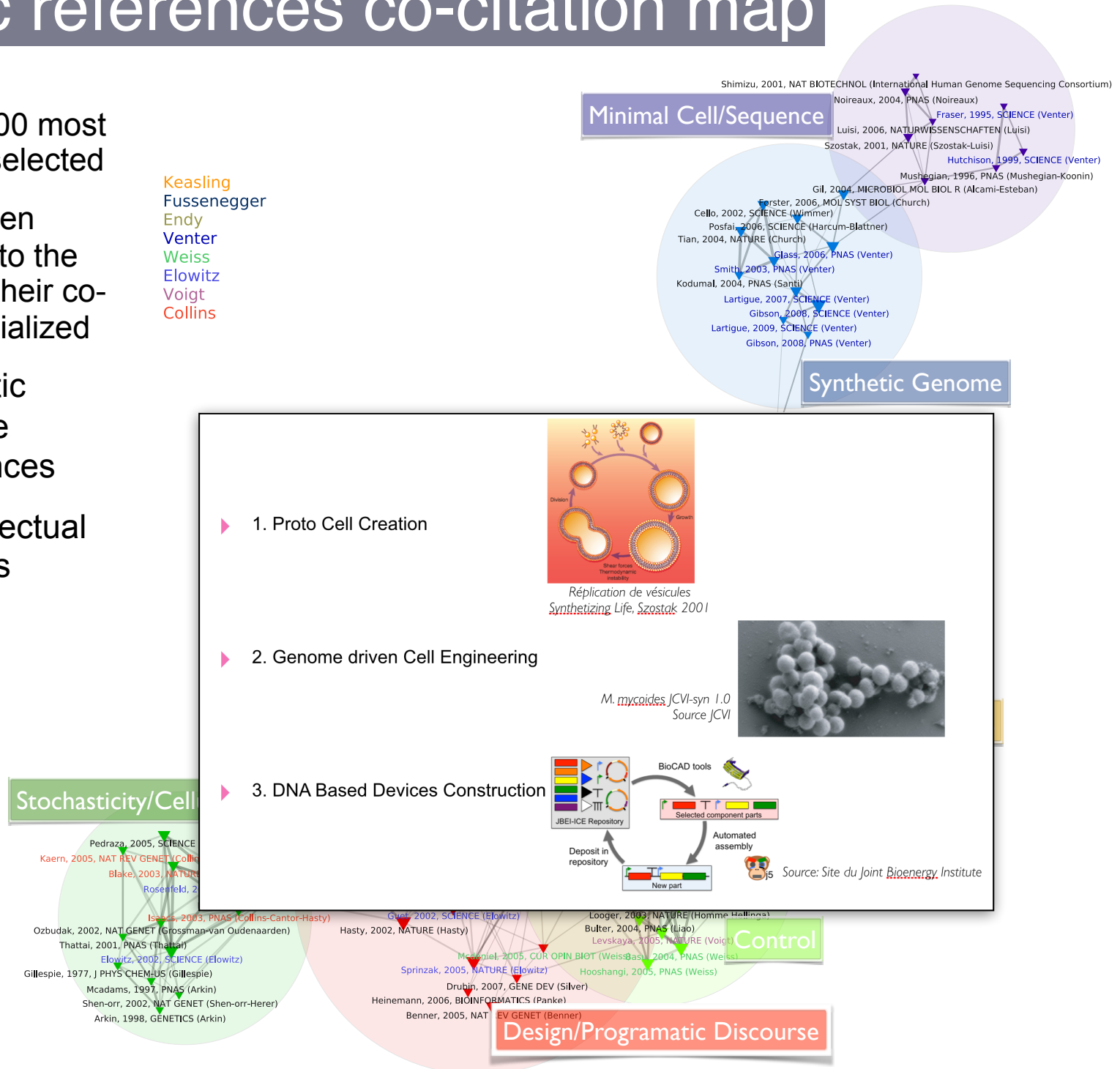
Scientific references co-citation map

- **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

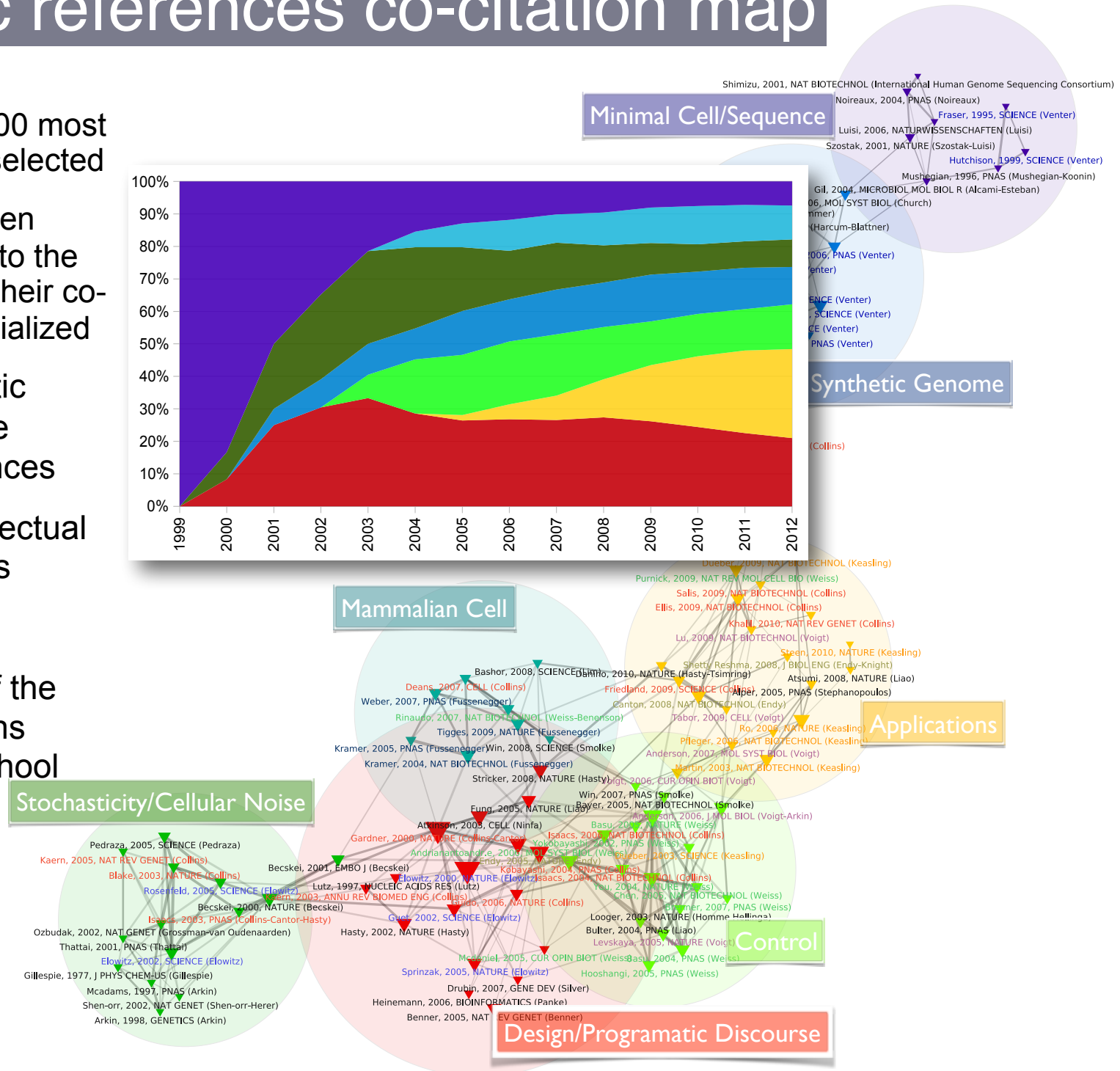
- **Selection**: only the 100 most cited references are selected
- **Network**: They are then connected according to the relative frequency of their co-occurrences and spatialized
- **Clustering**: Automatic detection of cohesive subgroups of references
- **Interpretation**: intellectual schools, major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



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

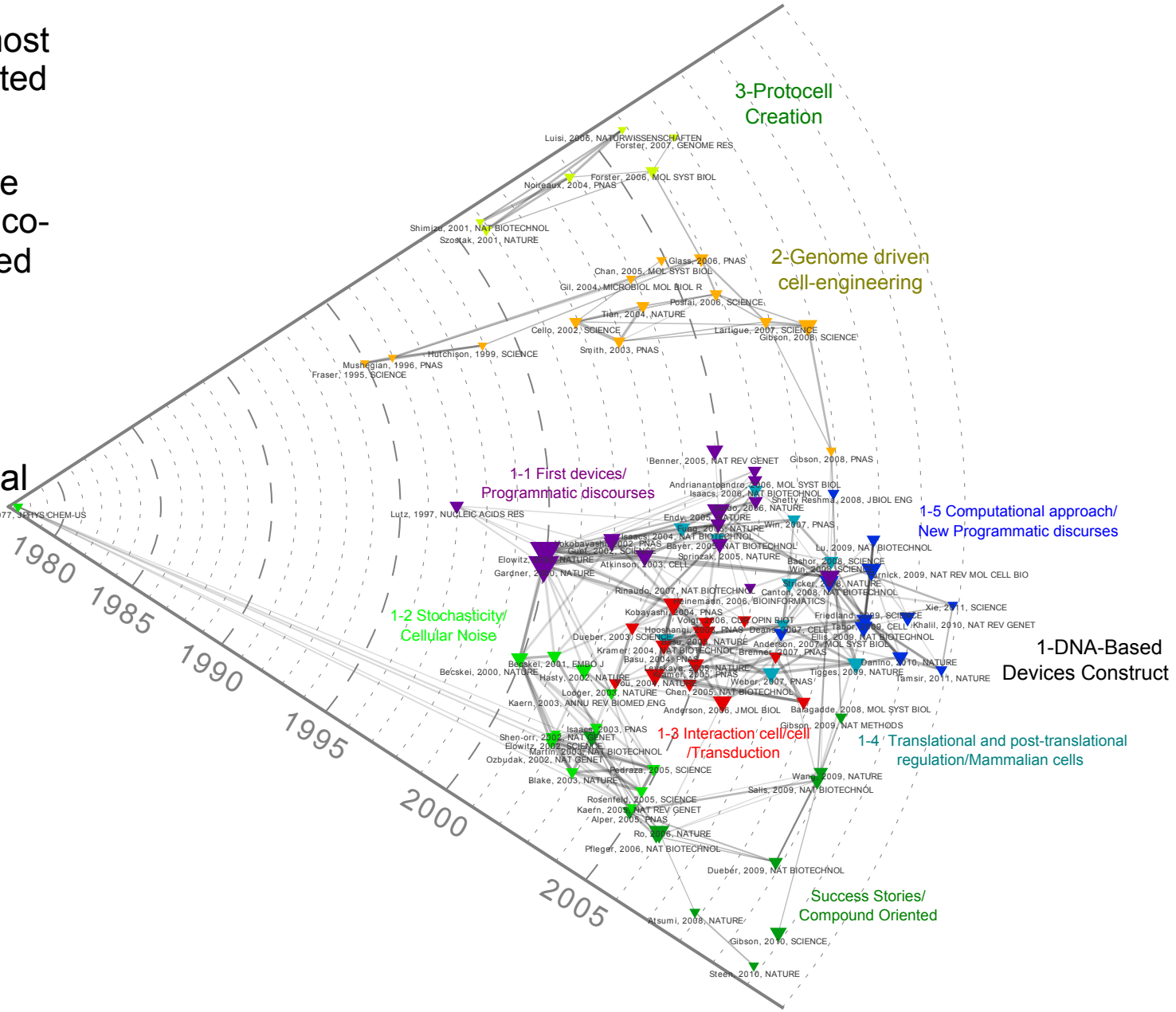


Scientific references co-citation map

- **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

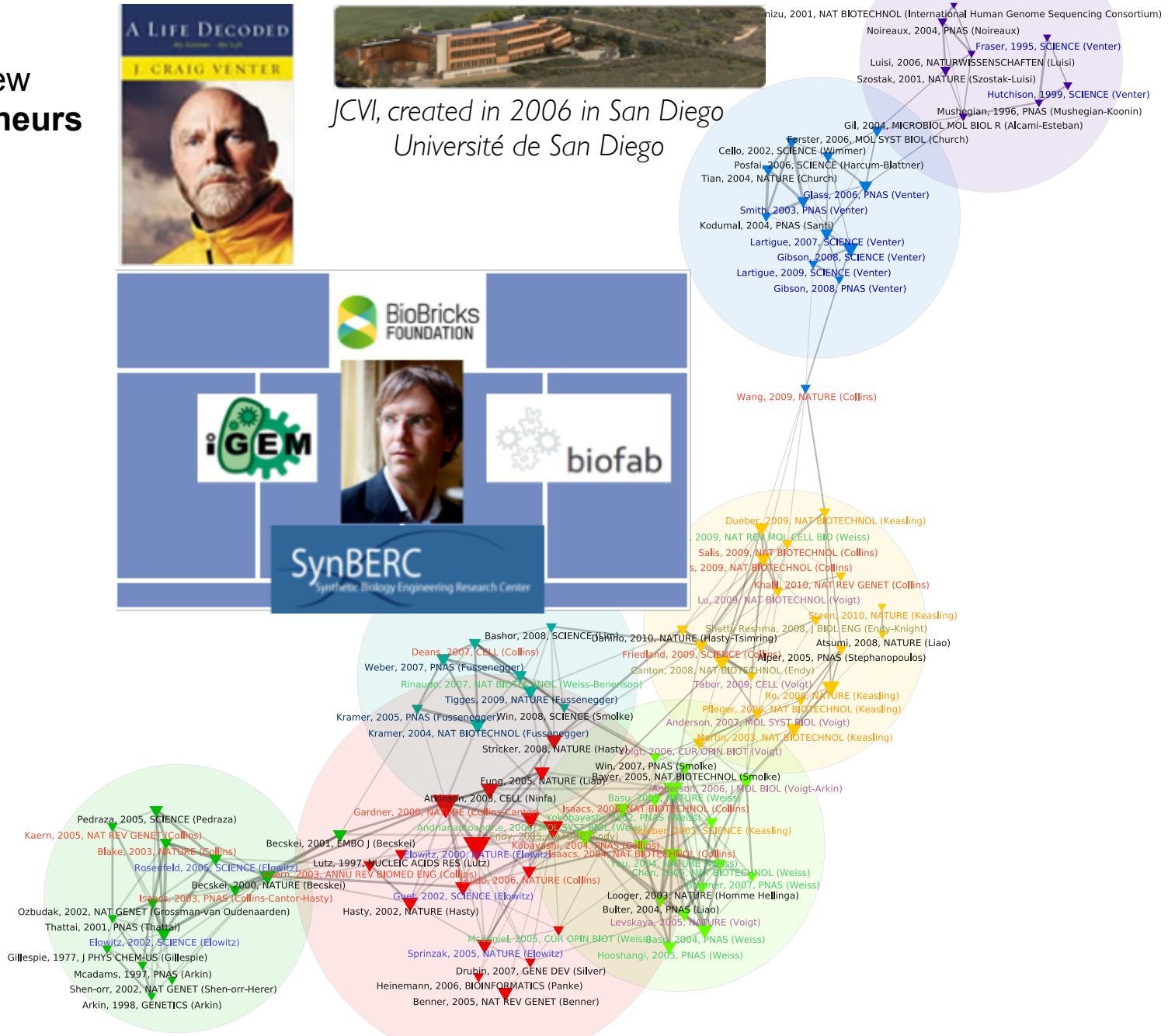
  -> radar representation



Actors and discourses

- Determining role of a few **institutional entrepreneurs**

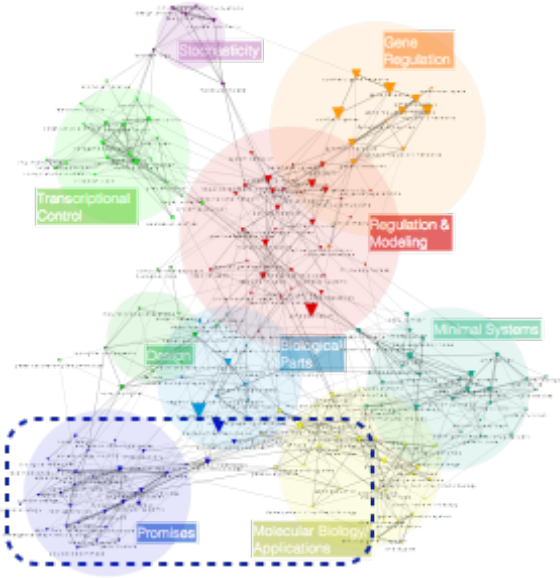
JCVI, created in 2006 in San Diego
Université de San Diego

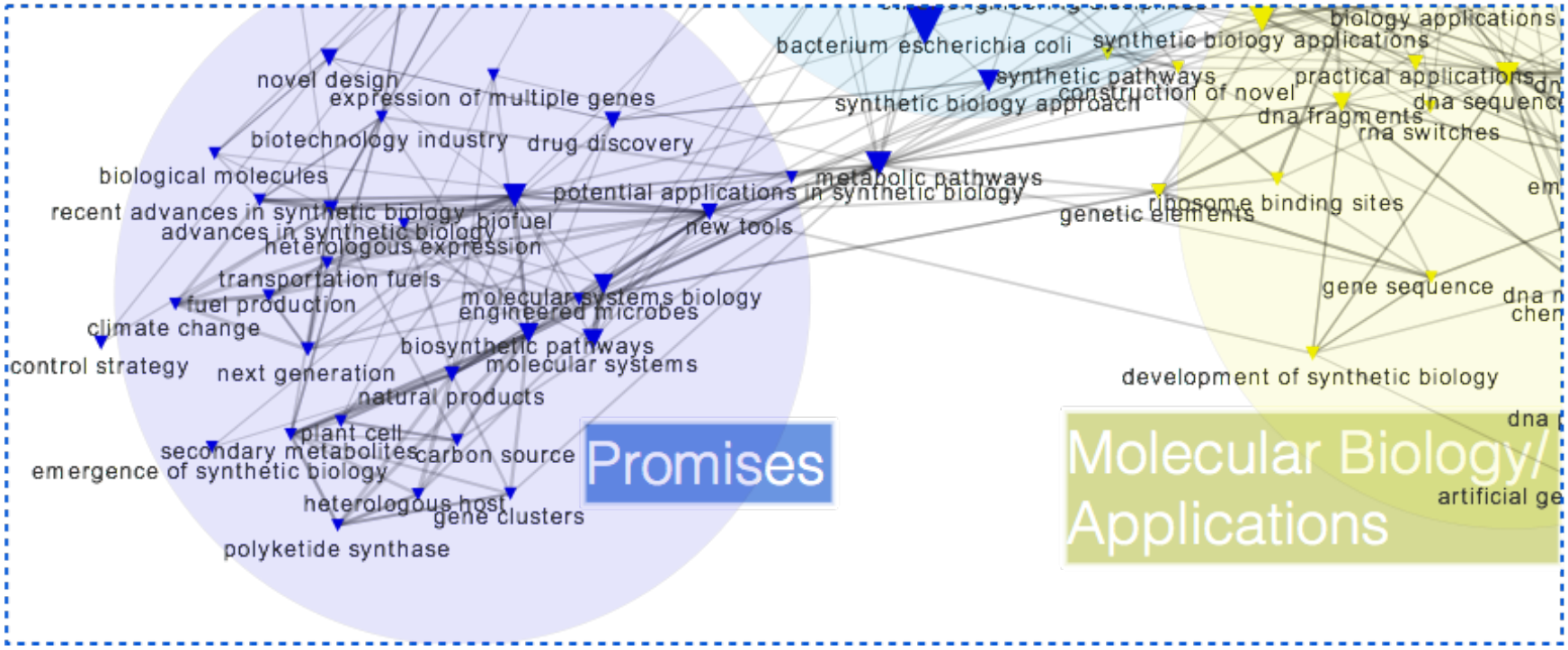


Actors and discourses

- 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



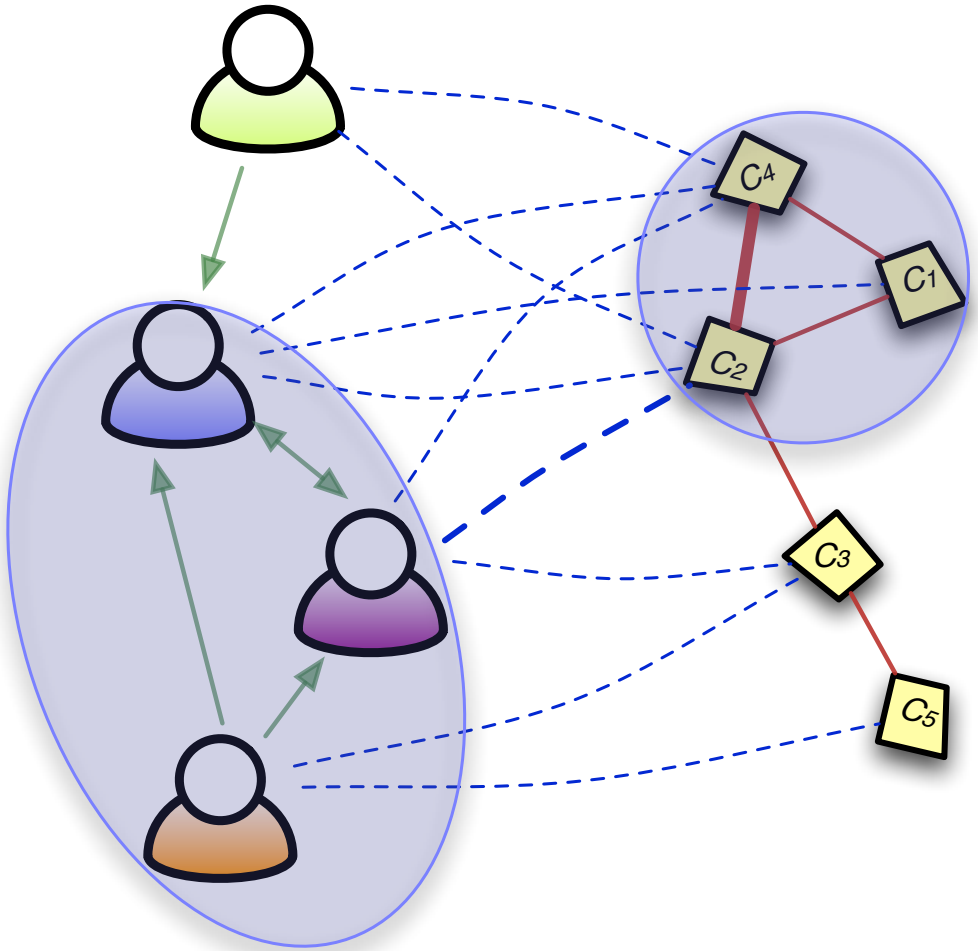


Unveiling Networks Community Structure

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

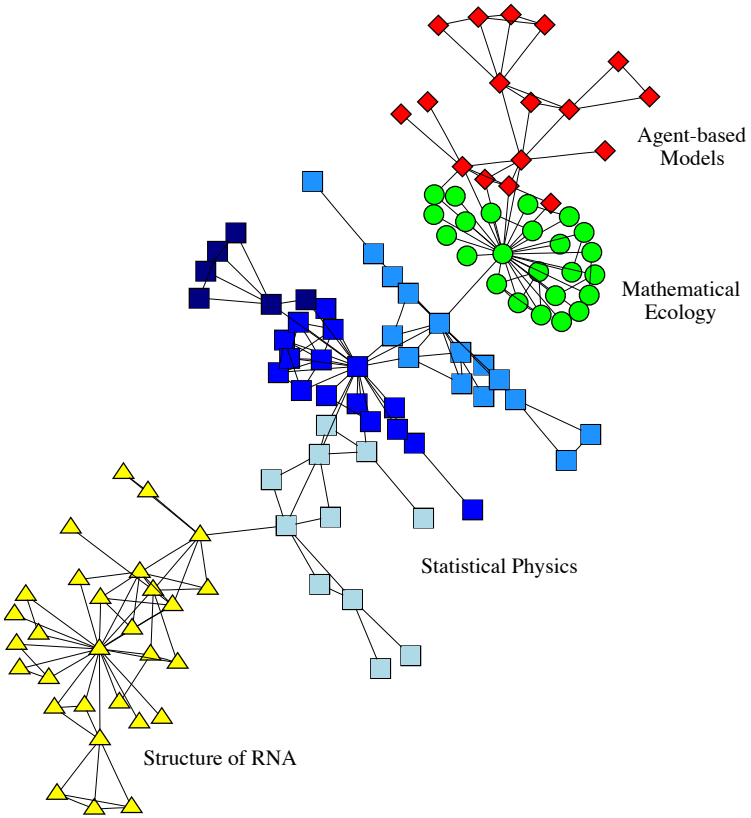
Community Detection Algorithms

- Modularity optimization (algorithmic definition)

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

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

Network Modularity Measure

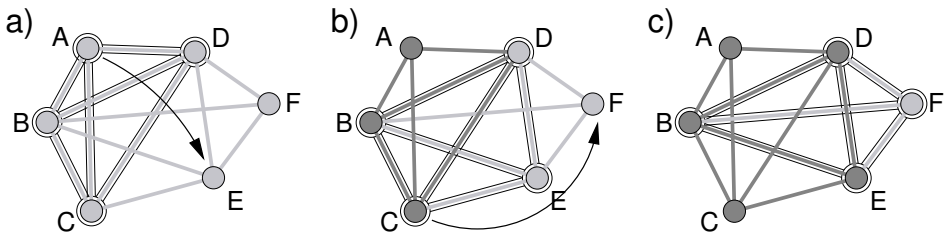


Santa Fe Collaboration Network

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.

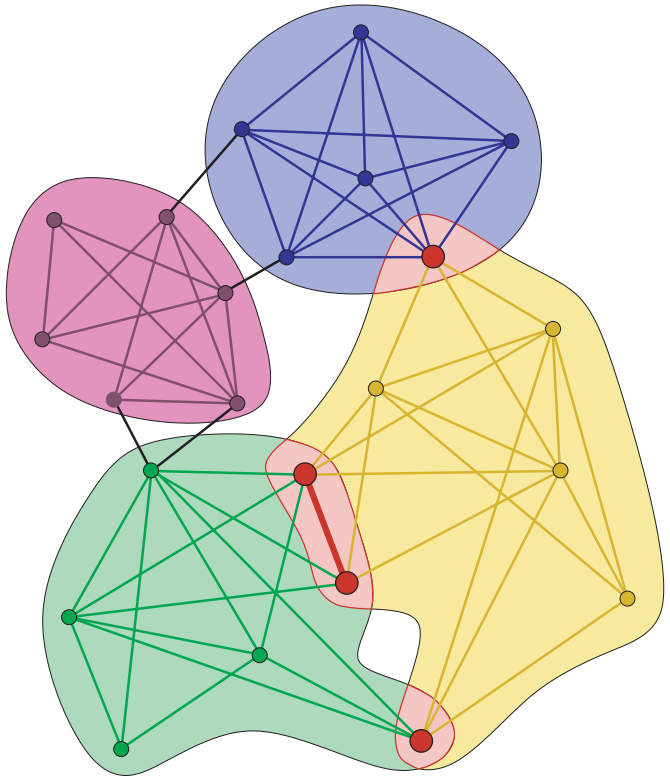
Community Detection Algorithms

- Modularity optimization (algorithmic definition)



Rolling 4-clique

- Clique Percolation (algebraic definition)

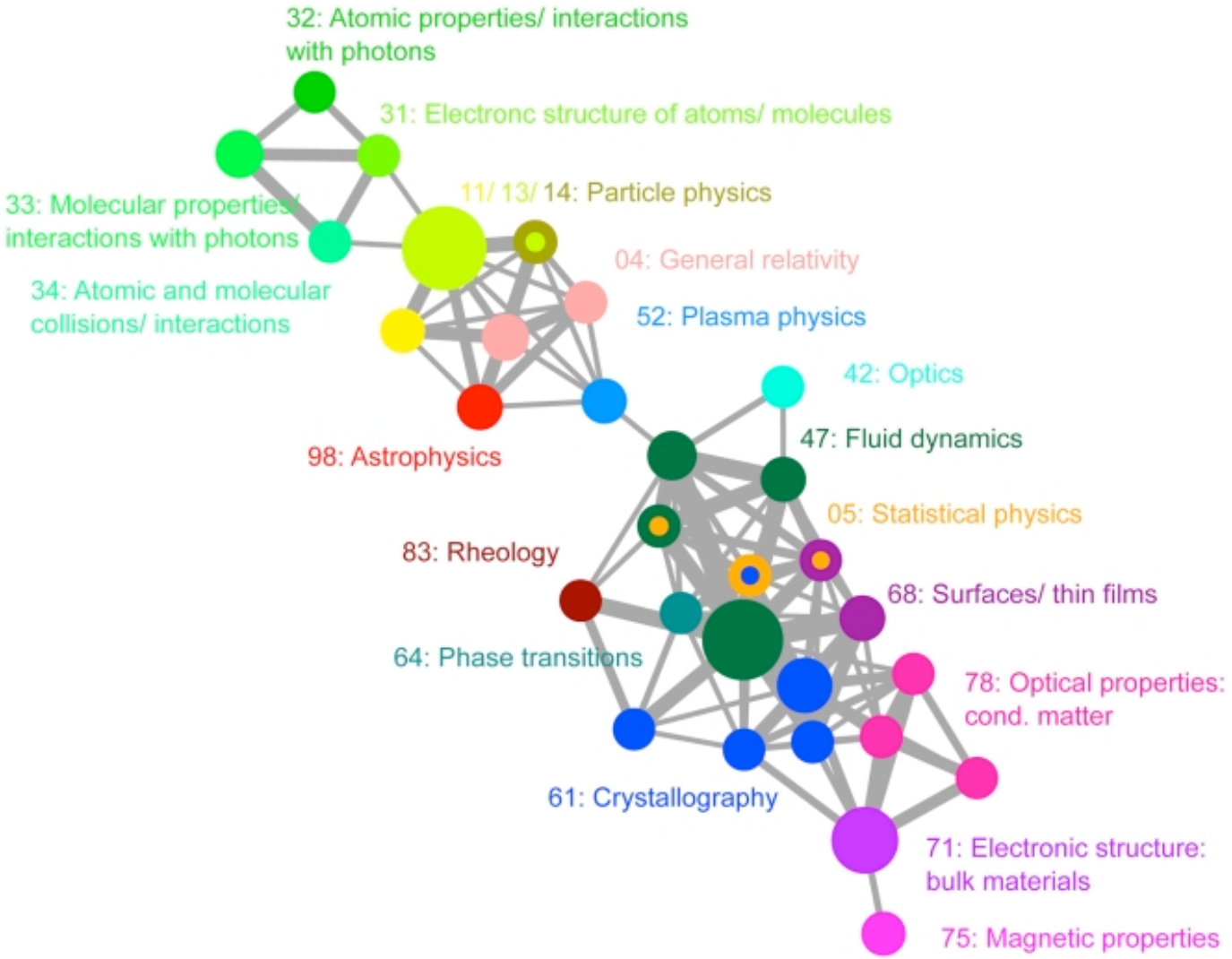


Clique percolation communities

Palla, G., Derenyi, I., Farkas, I. J., & Vicsek, T. A. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435

Community Detection Algorithms

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

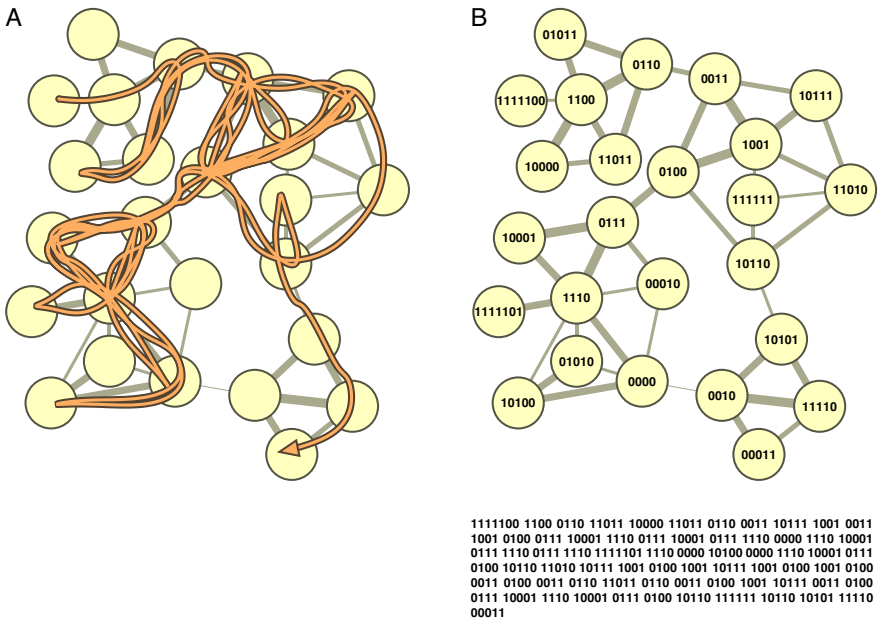


Map based on PACS code of APS publications

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

Community Detection Algorithms

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



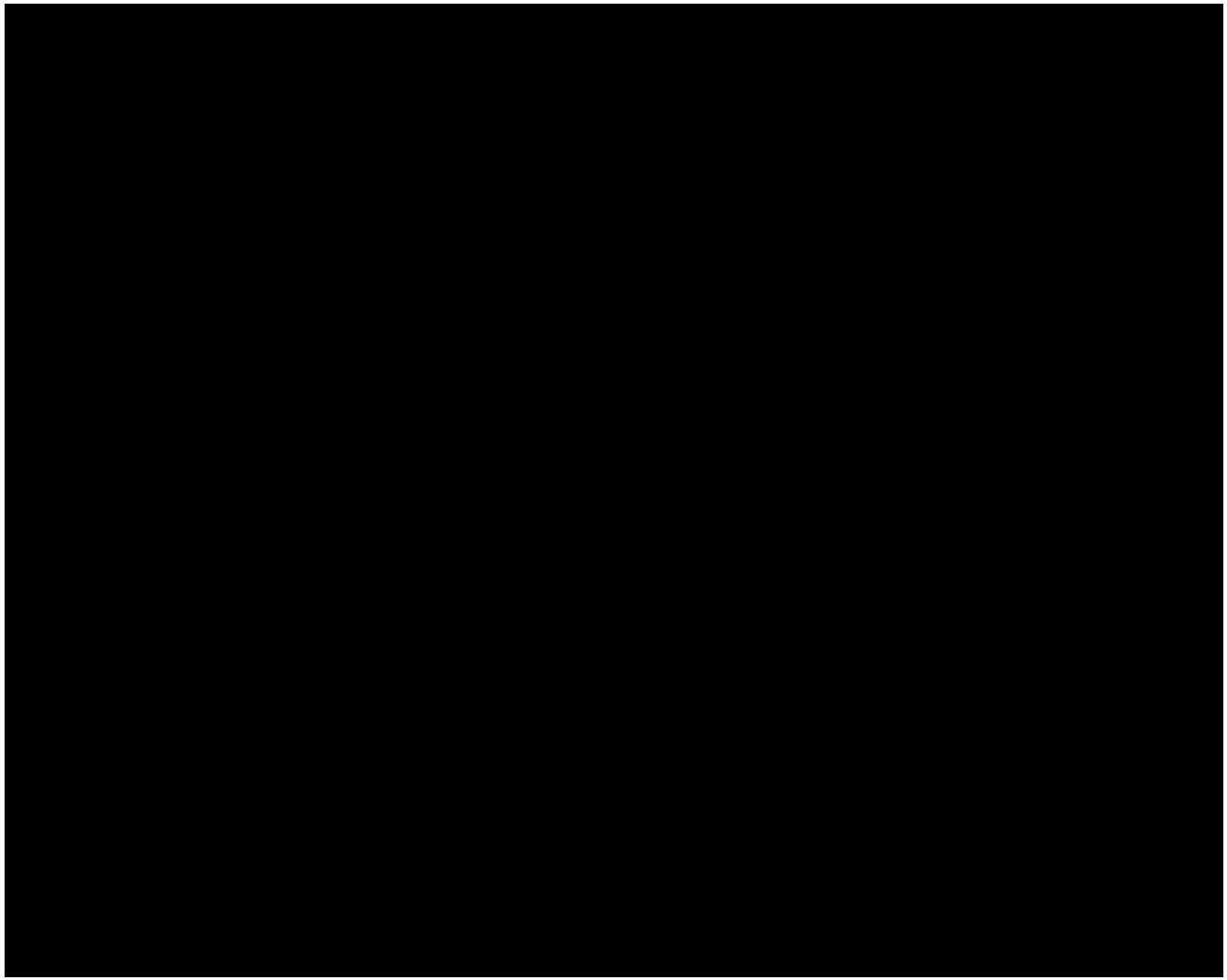
Rosvall, M., & Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences of the United States of America, 105(4), 1118.

compressing probability flow description

Spatializing Networks

Mapping Networks

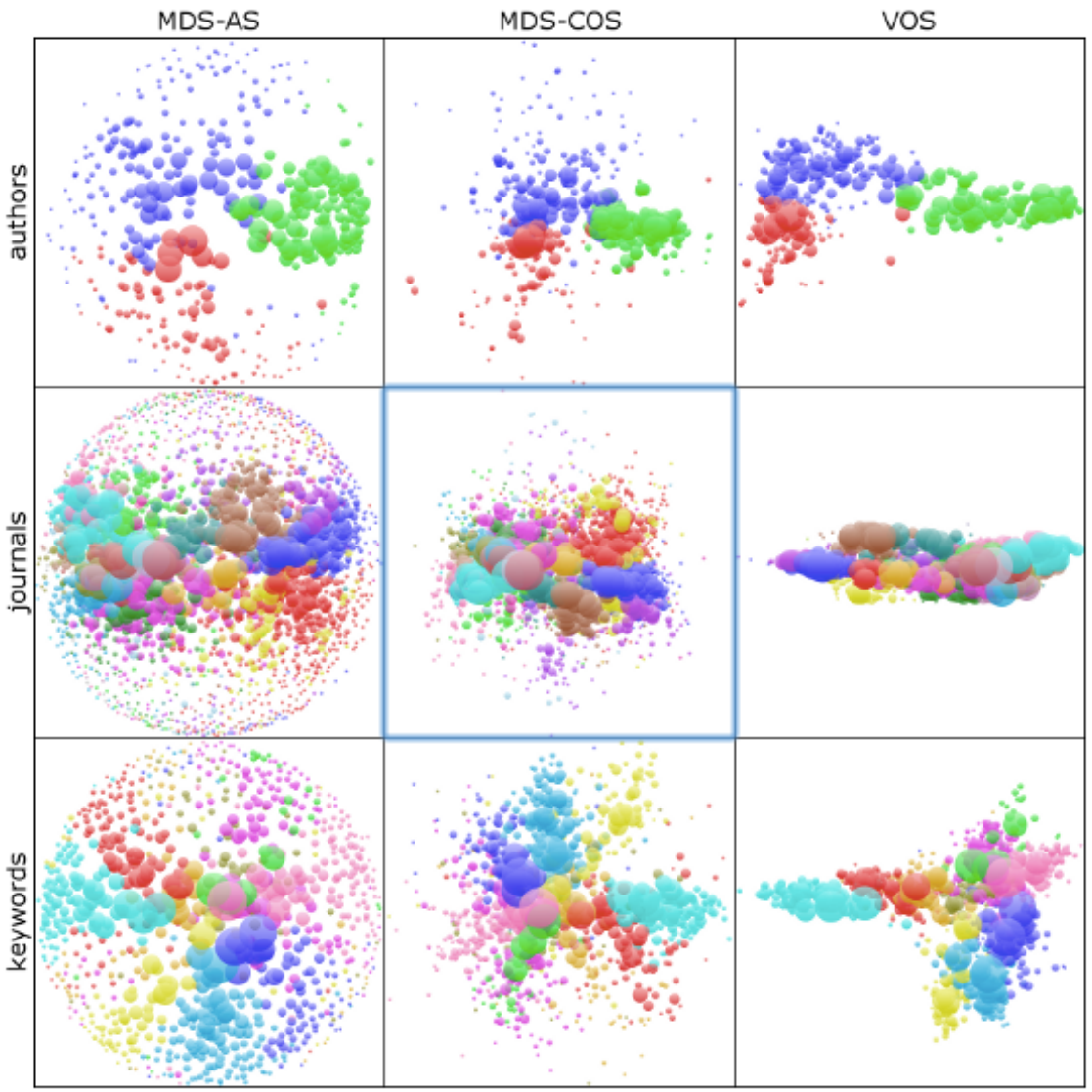
- Force-directed graph drawing



Gephi demo

Mapping Networks

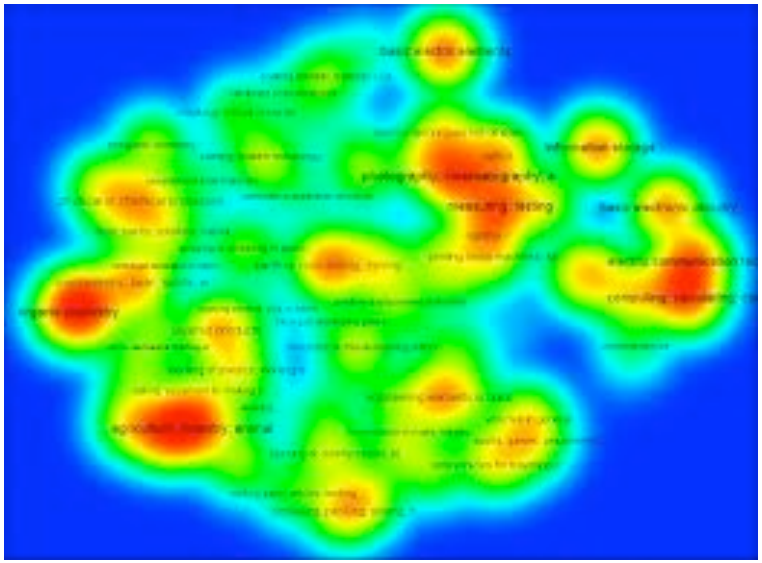
- Force-directed graph drawing
- Multidimensional Scaling and related techniques



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

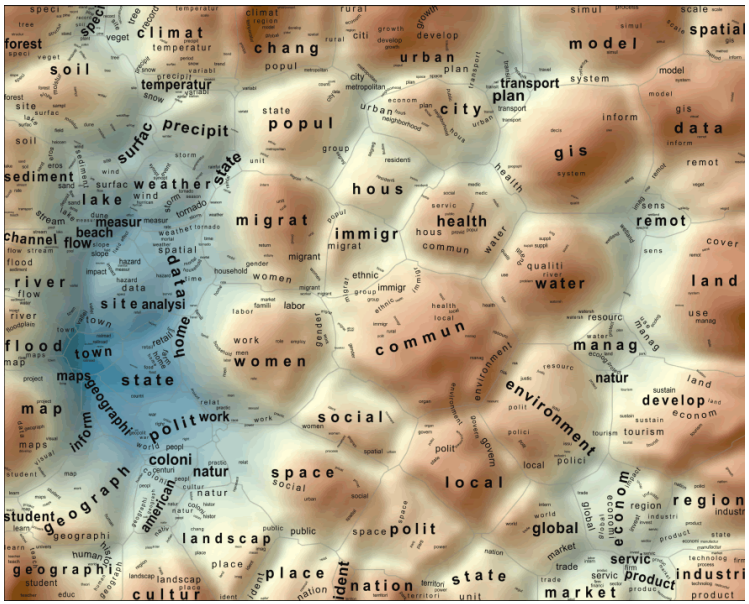
Mapping Networks

- Force-directed graph drawing



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

- Multidimensional Scaling and related techniques



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

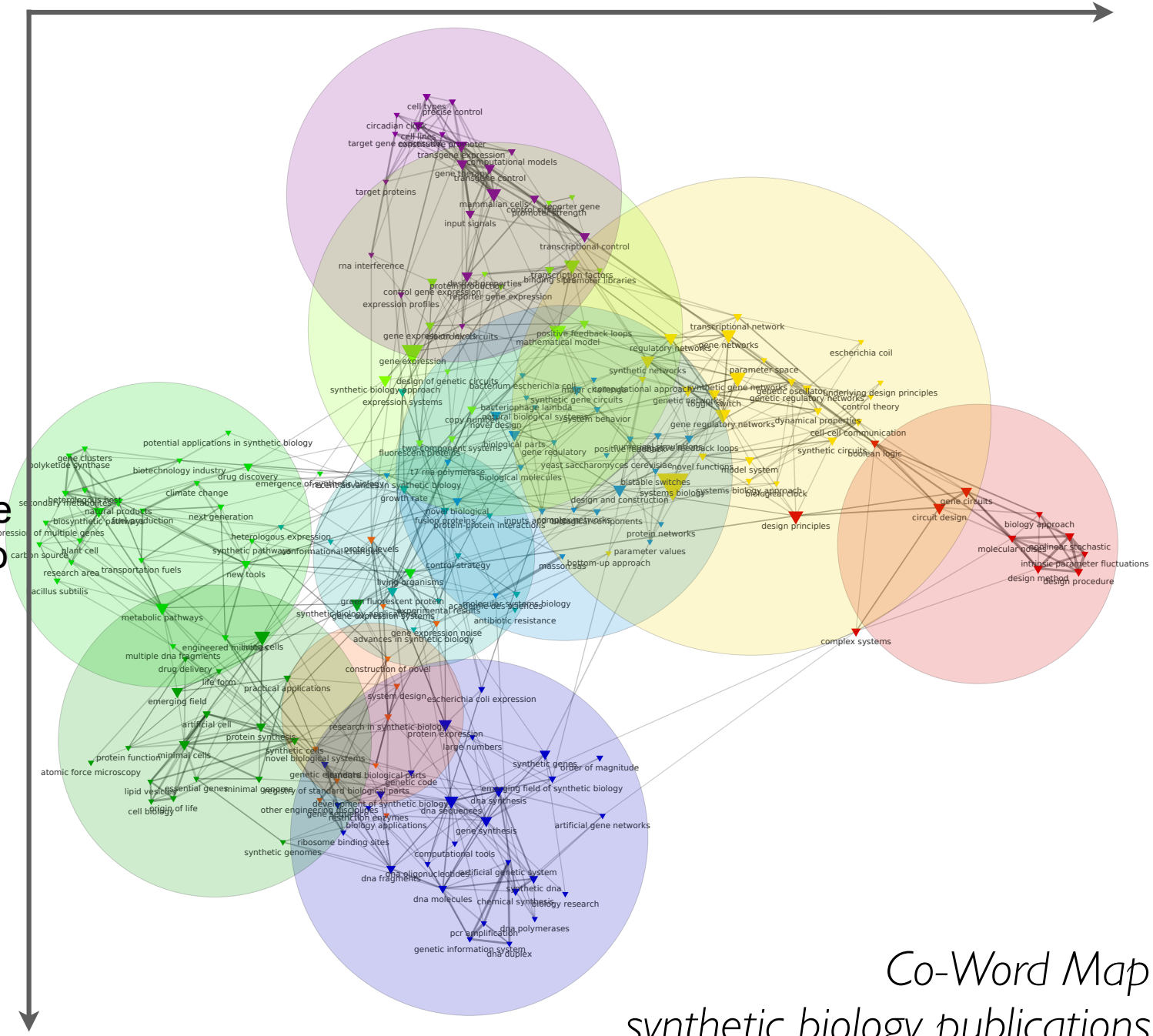
From Communities to Network Mapping

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

Cognitive landscape - Synthetic Biology Publications

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

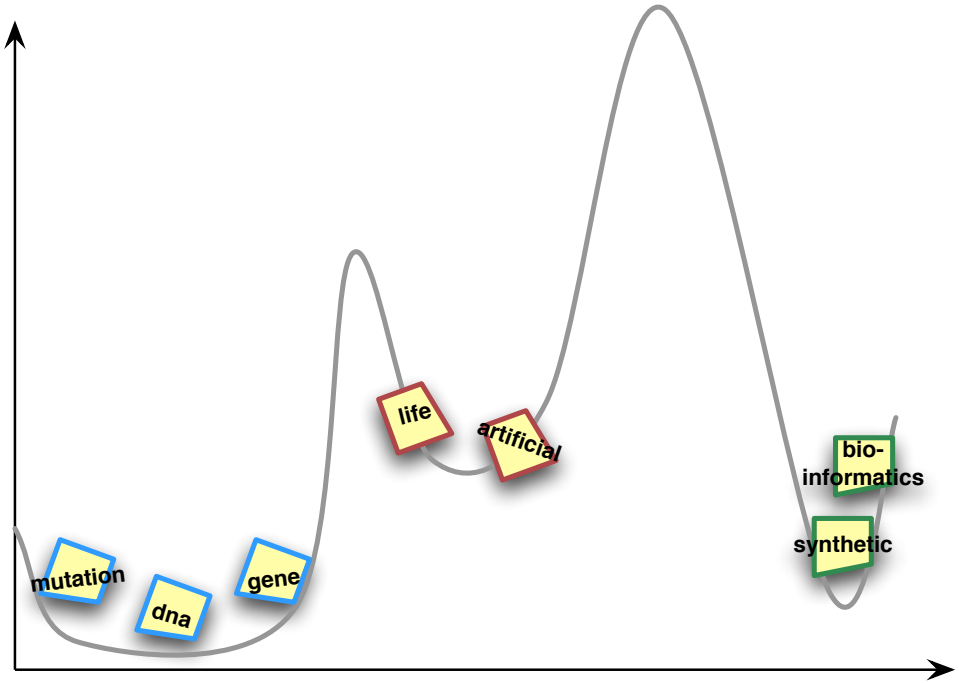
Textual Analysis

«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 ofscience and technology: Sociology ok Science in the real world.* London: The Macmillan Press 1,td.

what it is about a text that is interesting?

- **grammatical** criterion, candidate terms are usually limited *noun phrases*,
- **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, Filetering 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

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

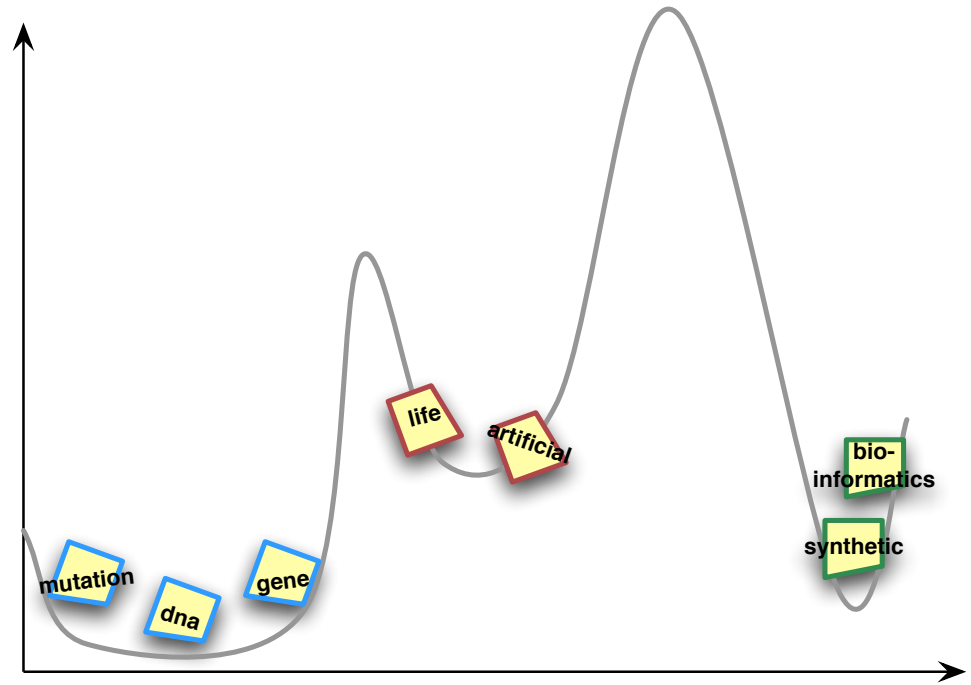
- Two possible strategies to assess how pertinent a term may be:

  - A term which local frequency in documents it occurs in is high is likely to be informative: *gf.idf*, pigeonhole score

  - A term which distribution is slanted toward certain type of documents or sub-vocabularies is likely to be pertinent: LLR, χ^2

Textual Analysis

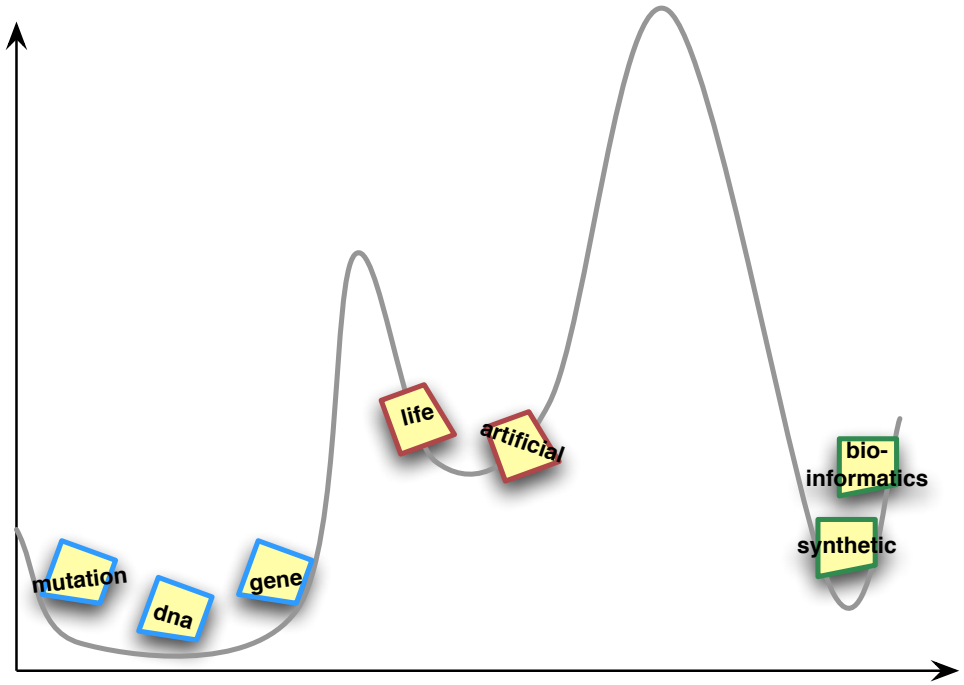
NLP analysis outcome



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

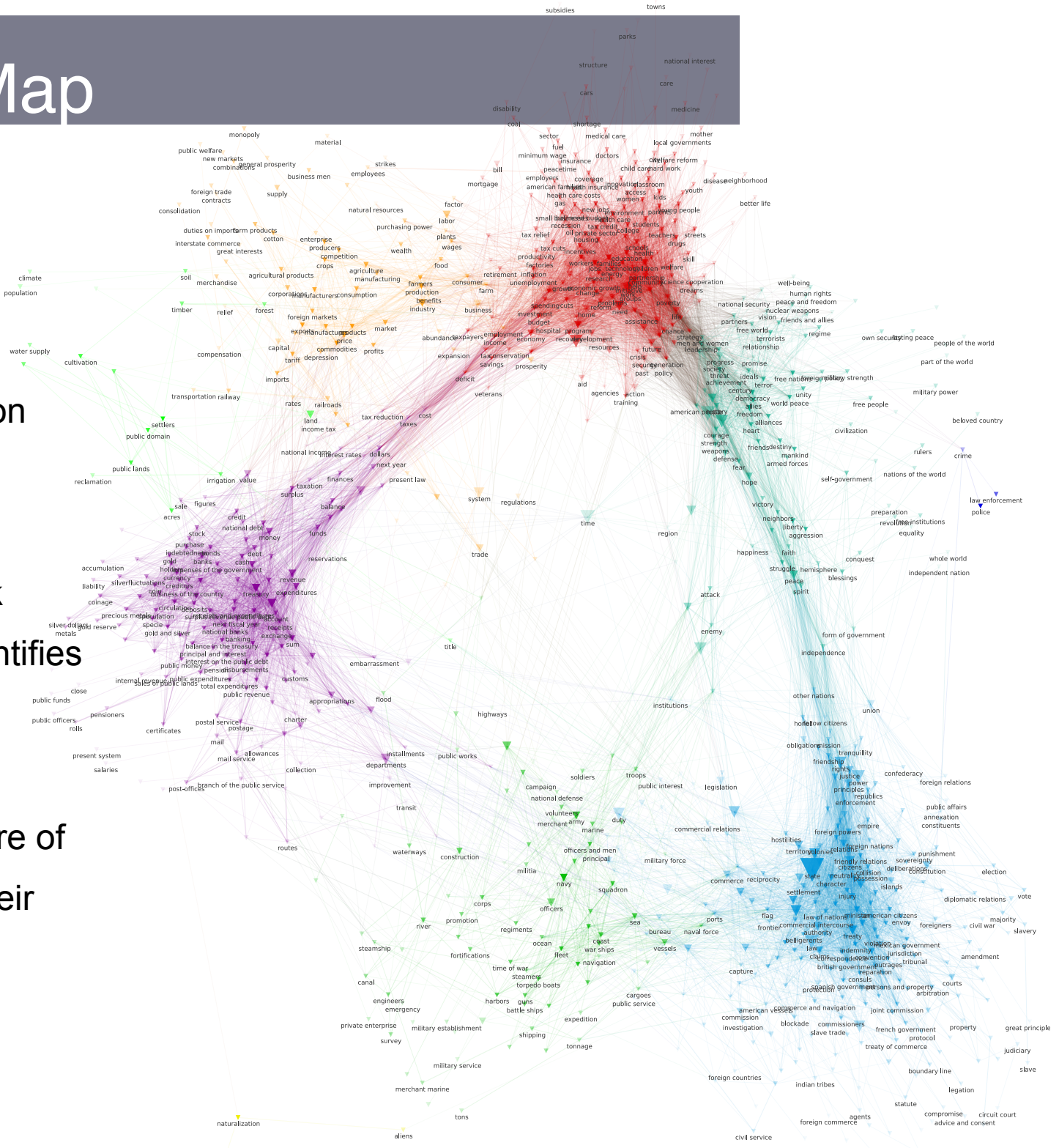


label	forms
abundance	abundance
acres	acres|&|acre
aliens	aliens|&|alien
alliances	alliances|&|alliance
allies	allies|&|ally
American life	American life|&|American lives
armed forces	armed forces|&|armed force|&|force of arms|&|arms and force
boundary line	boundary line|&|line of boundary
business men	business men|&|business man|&|men of business|&|man of business
children	children|&|child
commerce and navigation	commerce and navigation|&|navigation and commerce|&|navigation or commerce
construction work	construction work|&|construction of works|&|work of construction|&|works construction|&|construction of these works|&|work in construction
crews	crews|&|crew
crime	crime|&|crimes
crisis	crisis|&|crises
crops	crops|&|crop
cruisers	cruisers|&|cruiser
democracy	democracy|&|democracies
diplomatic relations	diplomatic relations
expenses of the Government	expenses of the Government|&|Government expenses|&|Government expense|&|expenses of Government|&|expense of the Government
farm products	farm products|&|products of the farm|&|products of farm
great importance	great importance|&|greater importance|&|greatest importance
health care costs	health care costs|&|cost of health care|&|health care cost
income tax	income tax|&|tax the income|&|tax on the income
peace and freedom	peace and freedom|&|freedom and peace|&|peace with freedom
property rights	property rights|&|property right|&|rights of property|&|right of property|&|rights and property

Semantic Map

Four steps process:

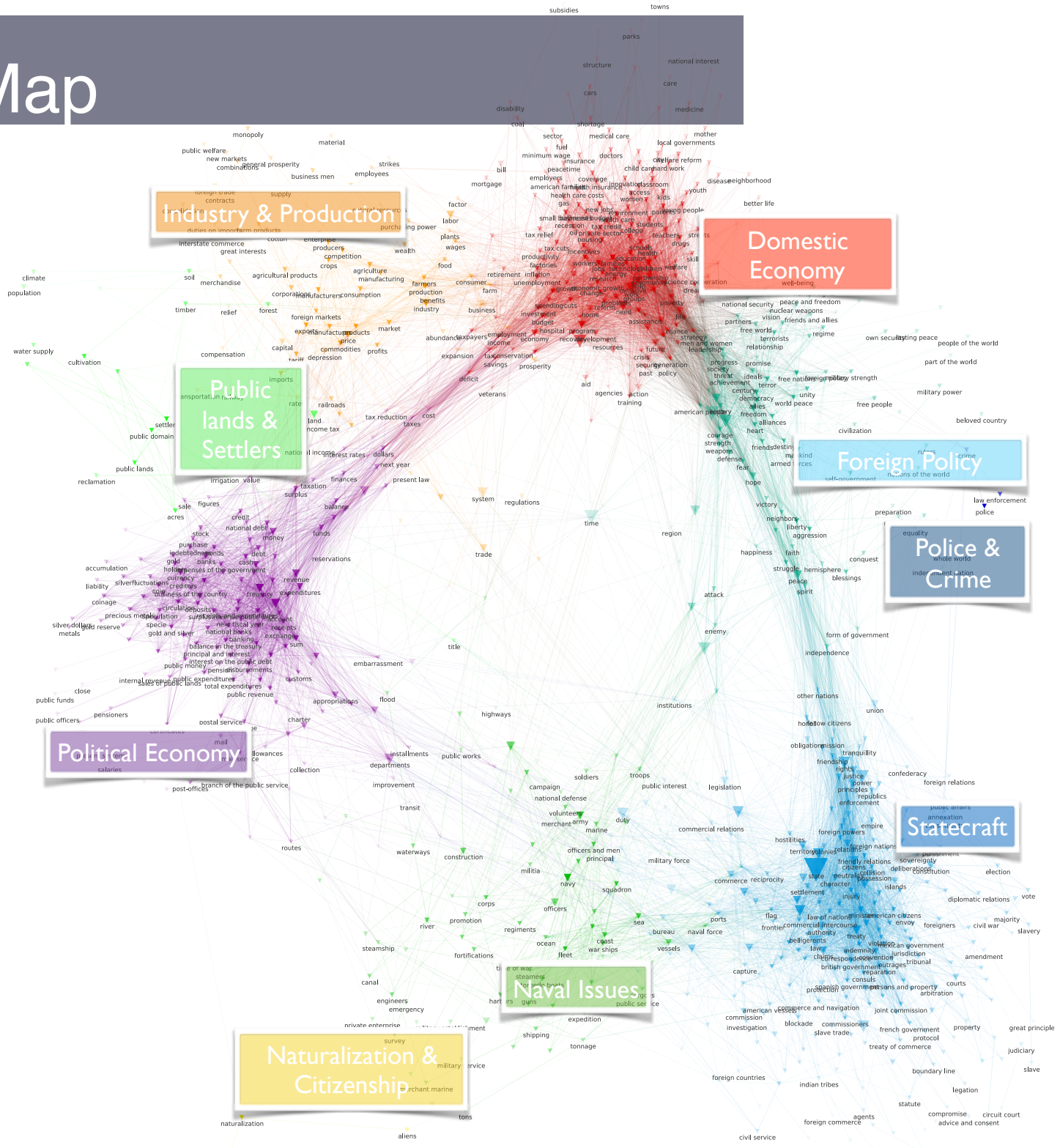
1. **Similarity Network** based on terms coocurrences
2. **Filter Edges** to obtain the sparsest connected network
3. **Community detection** identifies cohesive thematic areas
4. **Mapping** allows to visually investigate the inner structure of those clusters, as well as their relationships



Semantic Map

9 Topics:

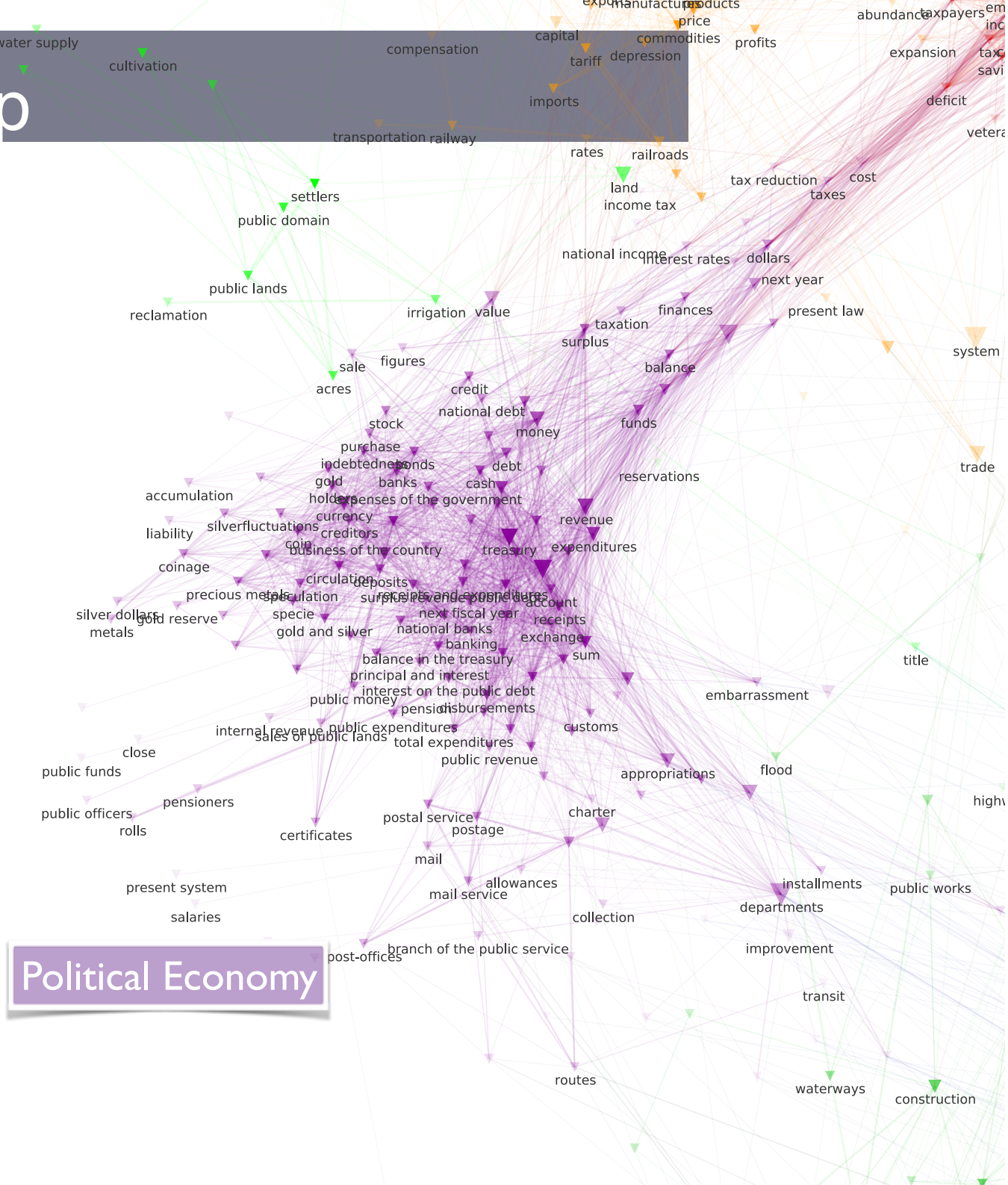
- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy**— “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”,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



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 inforcement”

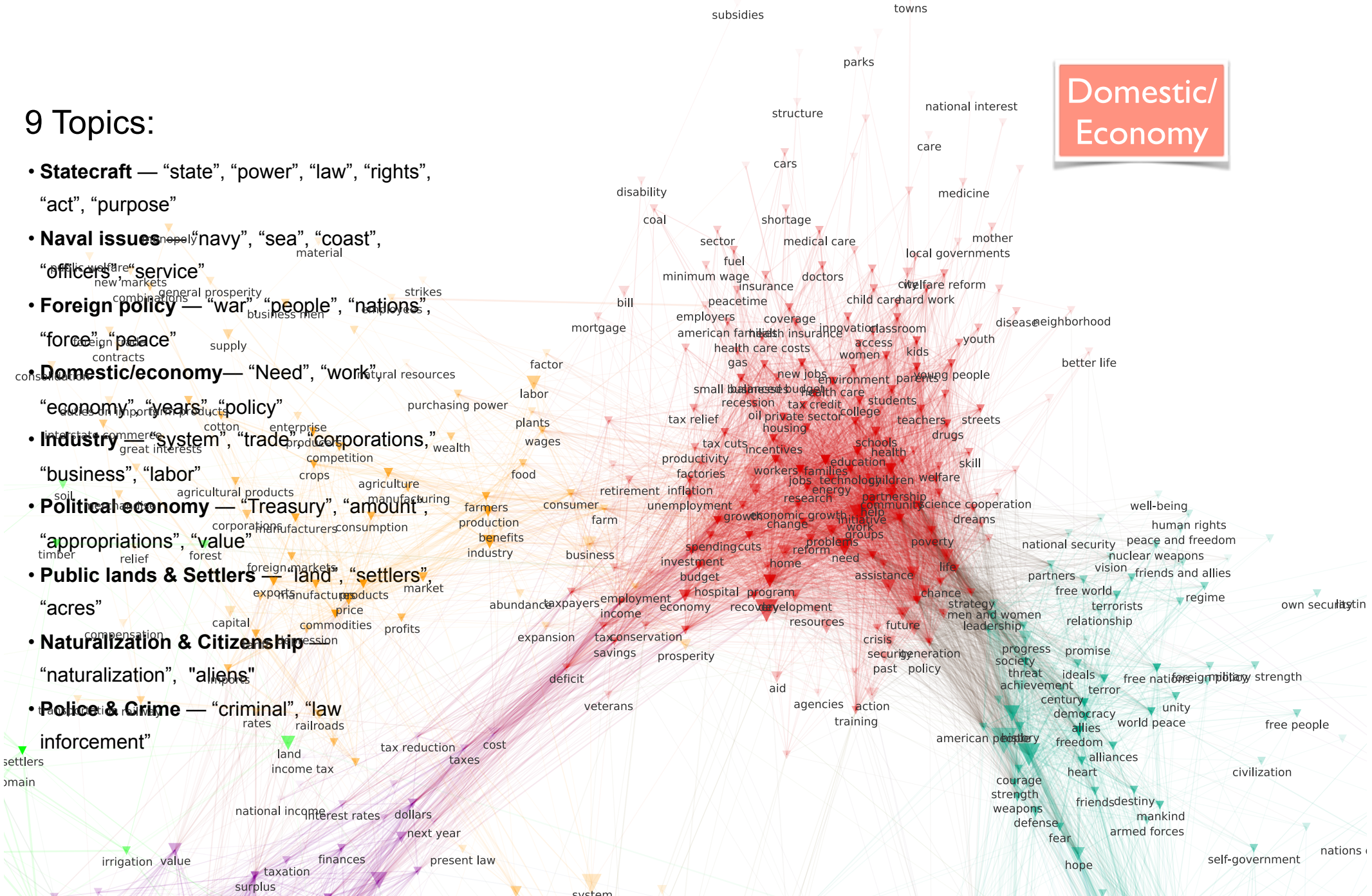


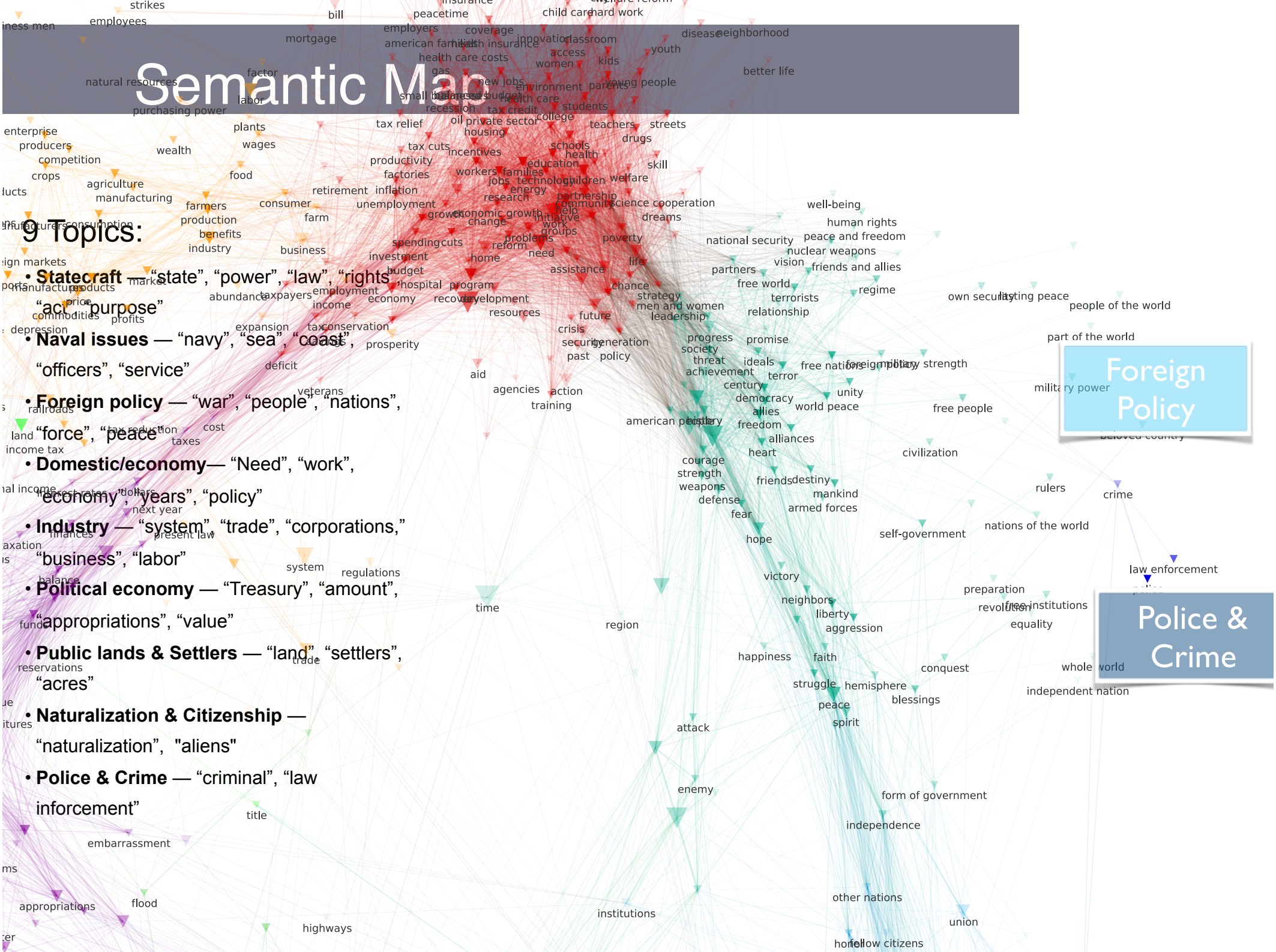
Political Economy

Semantic Map

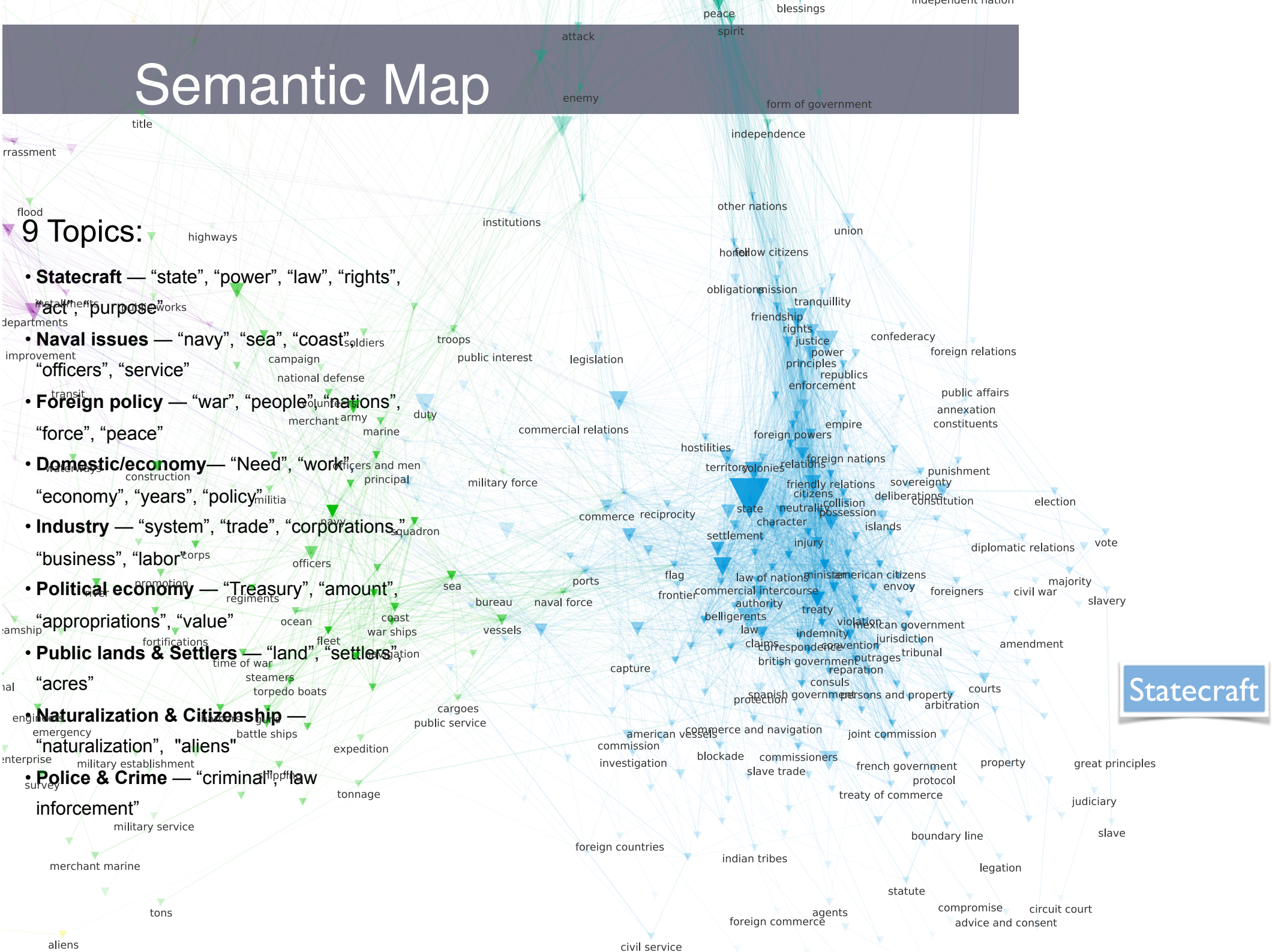
9 Topics:

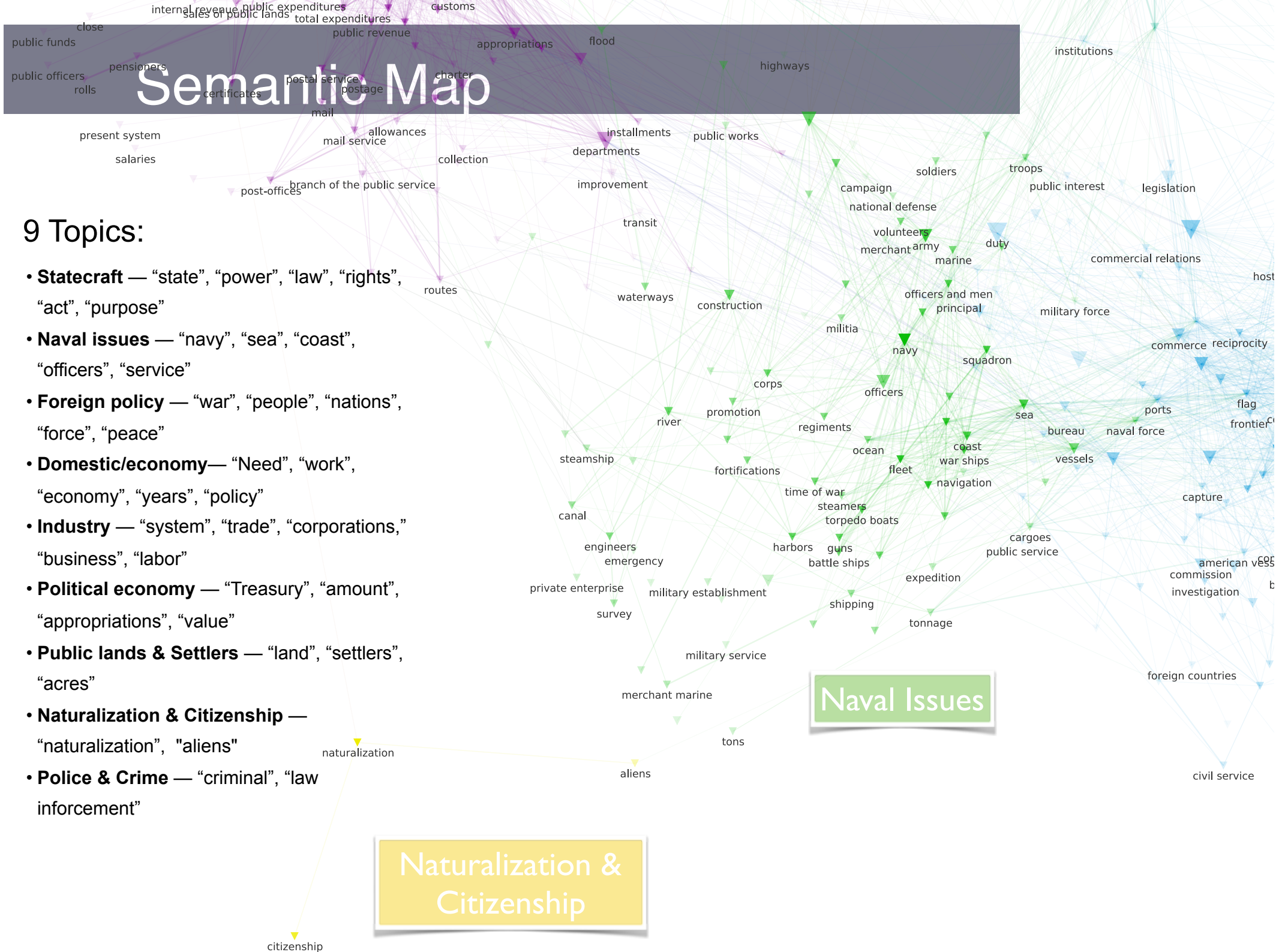
- **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
- **Naval issues** “navy”, “sea”, “coast”, “officers”, “service”
- **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
- **Domestic/economy**— “Need”, “work”, “economy”, “years”, “policy”
- **Industry** — “system”, “trade”,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




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 inforcement”
Foreign Policy
Police & Crime

Semantic Map

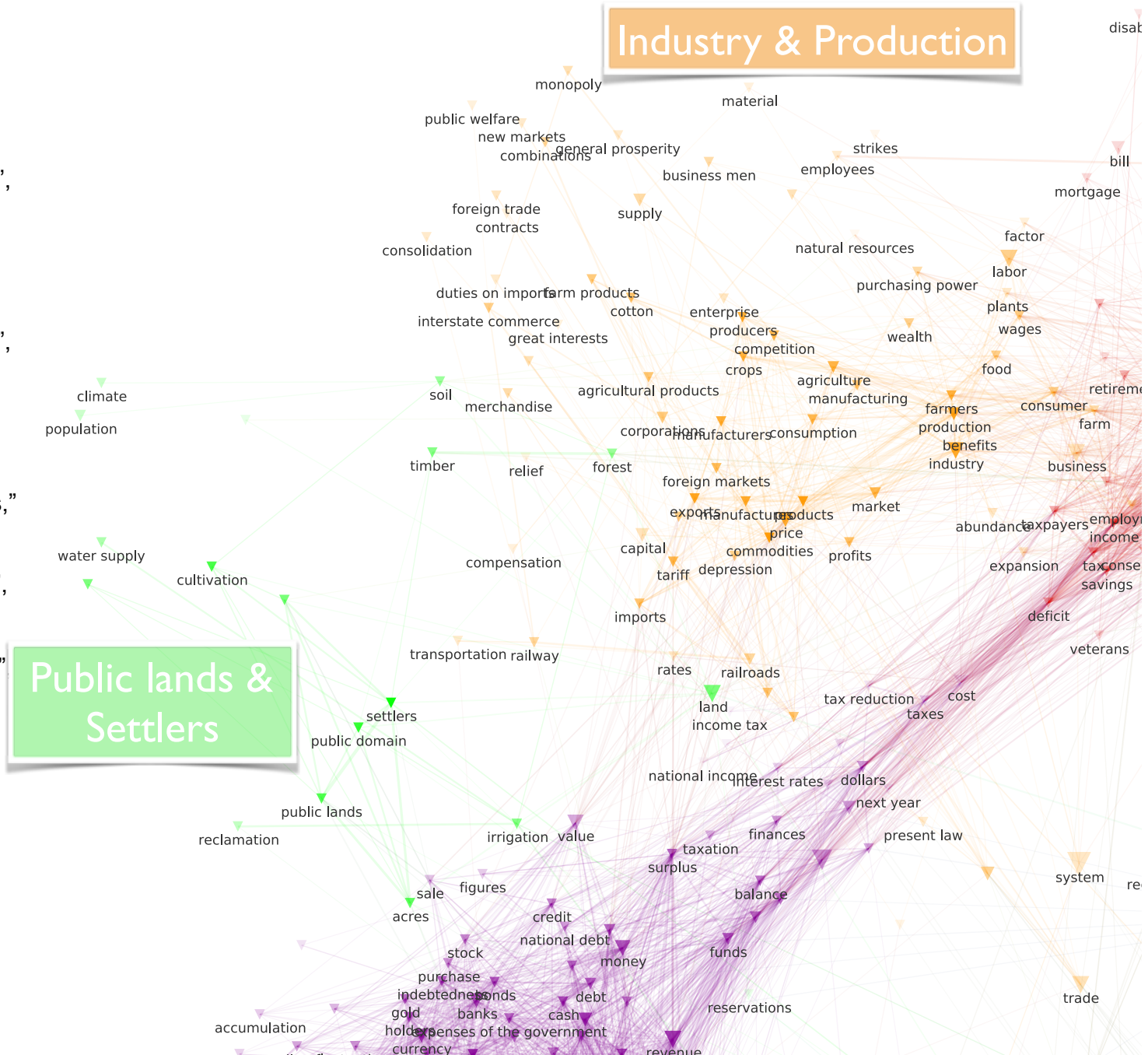



Semantic Map
9 Topics:
• **Statecraft** — “state”, “power”, “law”, “rights”, “act”, “purpose”
• **Naval issues** — “navy”, “sea”, “coast”, “officers”, “service”
• **Foreign policy** — “war”, “people”, “nations”, “force”, “peace”
• **Domestic/economy**— “Need”, “work”, “economy”, “years”, “policy”
• **Industry** — “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 inforcement”
Naval Issues
Naturalization & Citizenship
internal revenue
public expenditures
sales of public lands
total expenditures
customs
close
public funds
public revenue
appropriations
flood
institutions
pensioners
public officers
rolls
postal service
charter
certificates
postage
highways
mail
present system
allowances
mail service
installments
public works
salaries
departments
collection
soldiers
troops
branch of the public service
post-offices
improvement
campaign
public interest
legislation
national defense
transit
volunteers
army
merchant
duty
marine
commercial relations
routes
waterways
construction
officers and men
principal
military force
militia
navy
commerce
reciprocity
squadron
corps
officers
river
promotion
sea
ports
flag
regiments
bureau
naval force
ocean
coast
war ships
steamship
vessels
fortifications
fleet
navigation
time of war
capture
steamers
canal
torpedo boats
cargoes
engineers
harbors
guns
public service
emergency
battle ships
expedition
private enterprise
military establishment
commission
investigation
shipping
survey
tonnage
military service
foreign countries
merchant marine
tons
naturalization
aliens
civil service
citizenship

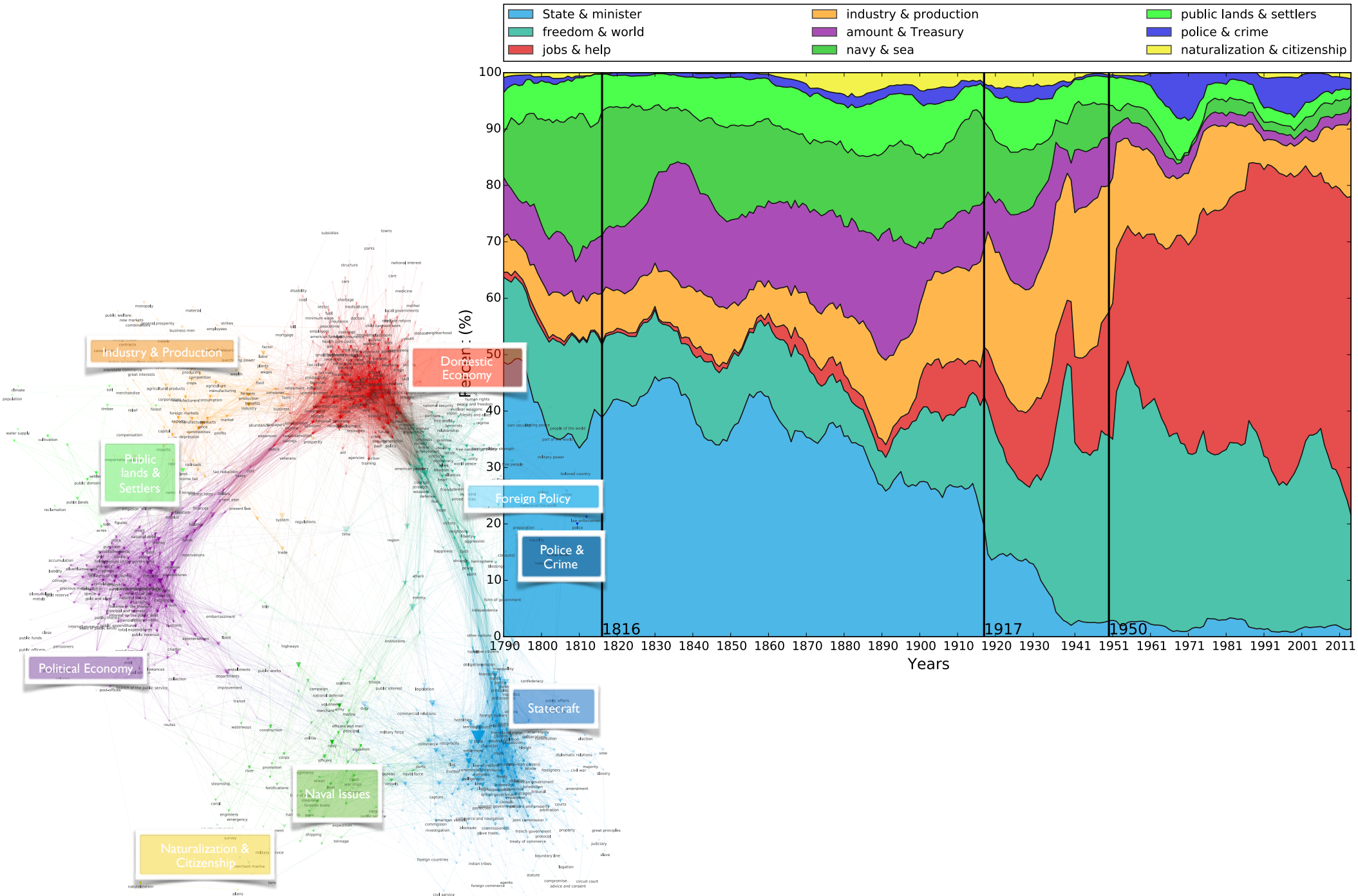
Semantic Map

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 inforcement”

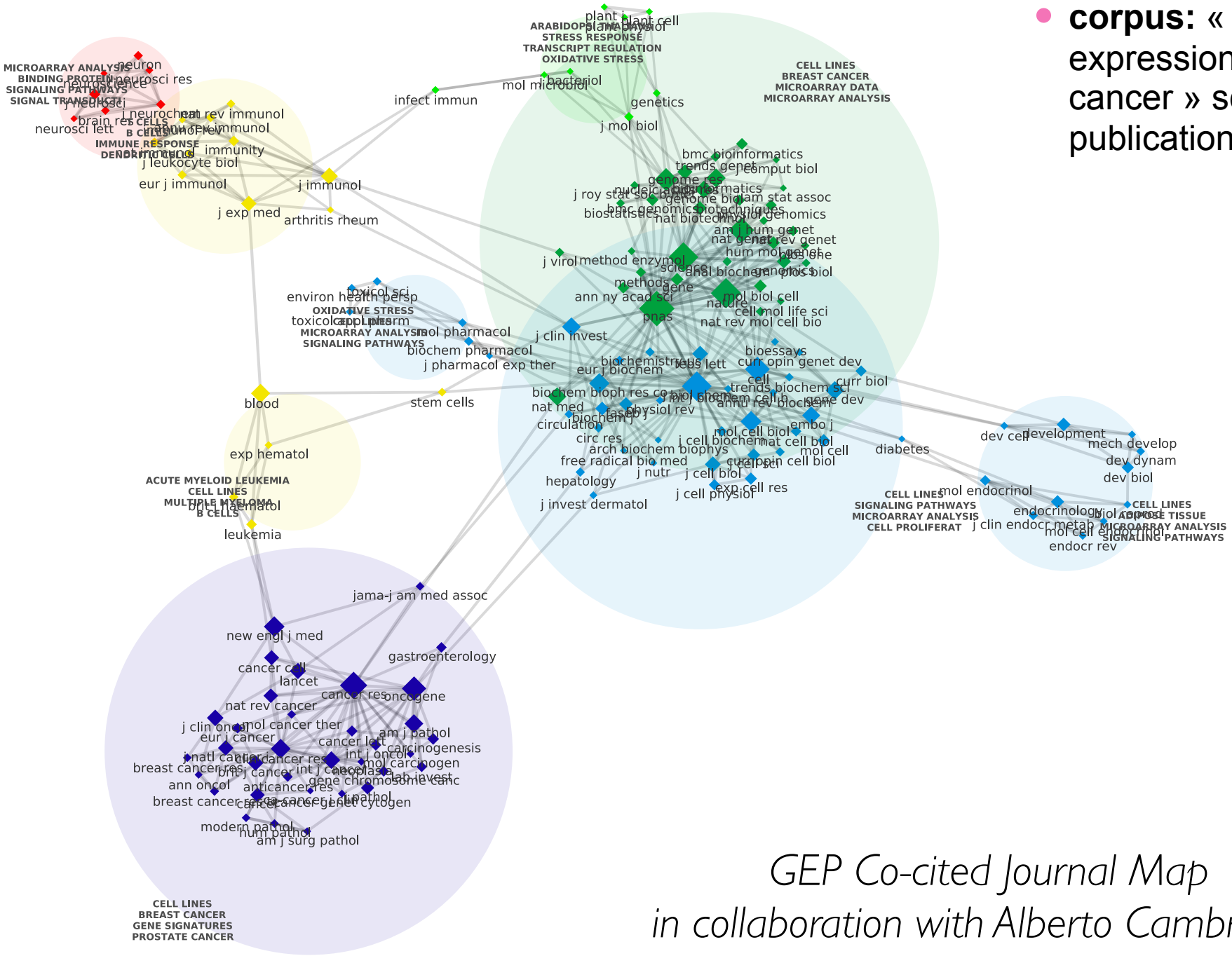


Categories in Time



From heterogeneous structure to dynamics

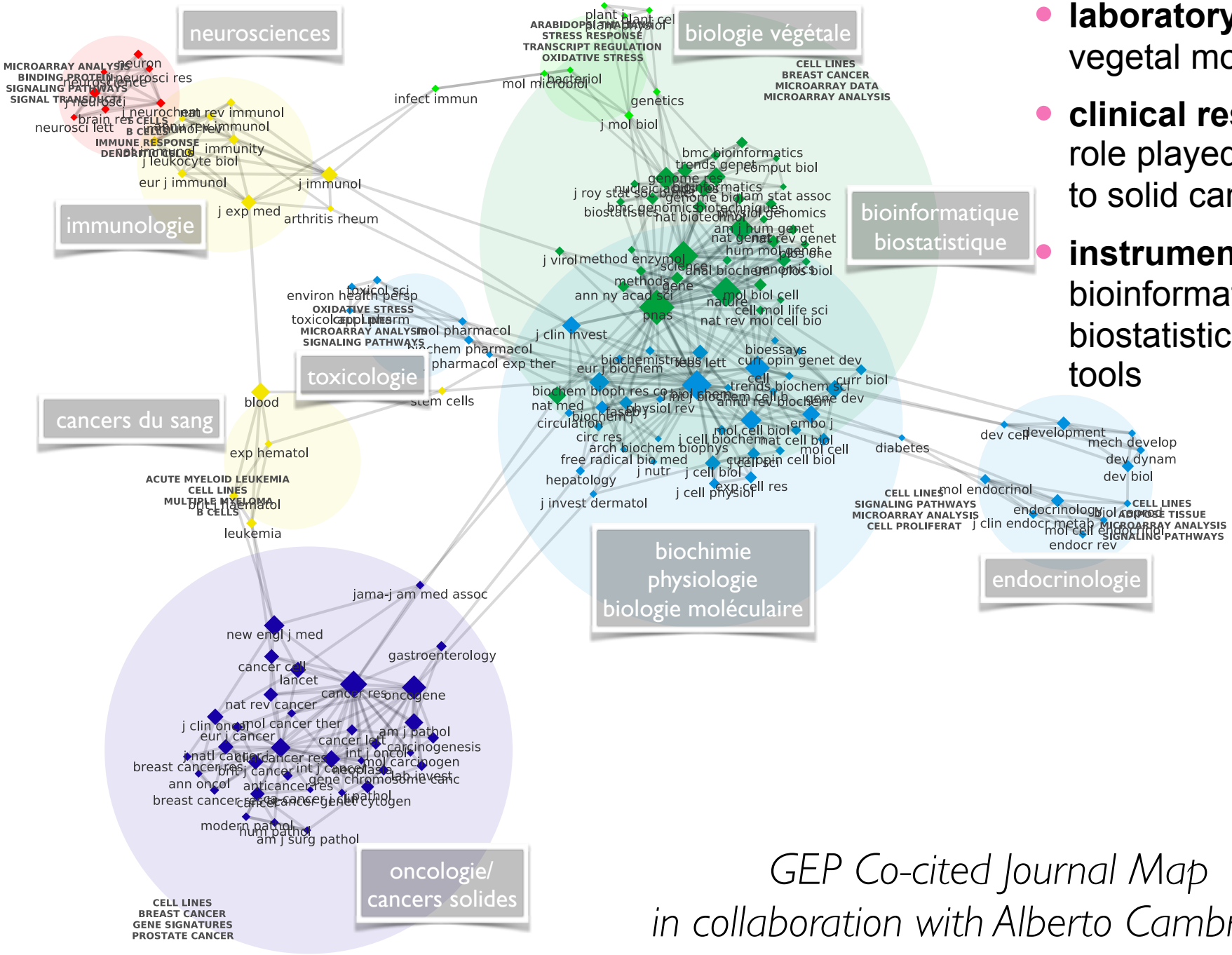
Translational Research Triangle



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

GEP Co-cited Journal Map
in collaboration with Alberto Cambrosio

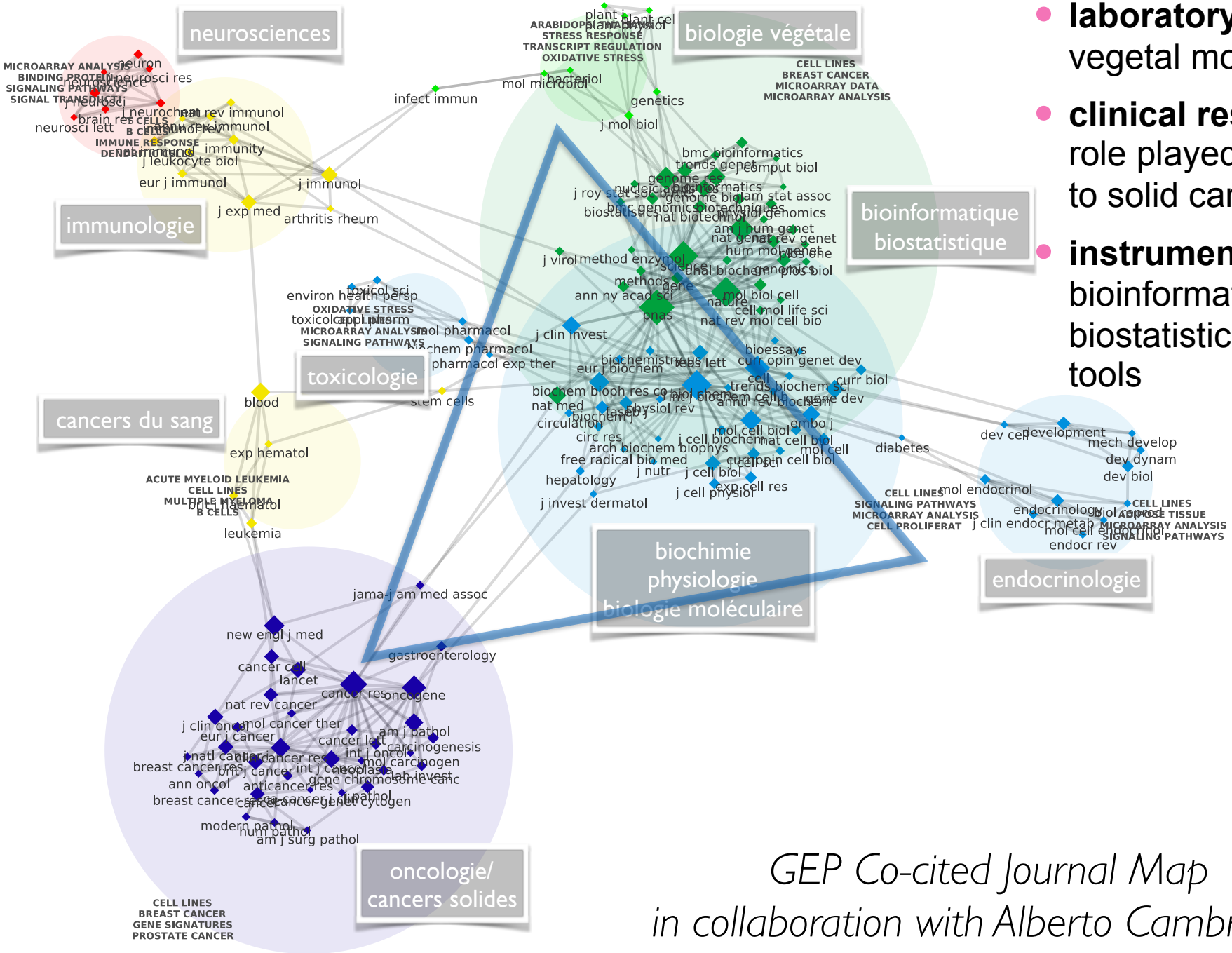
Translational Research Triangle



- **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

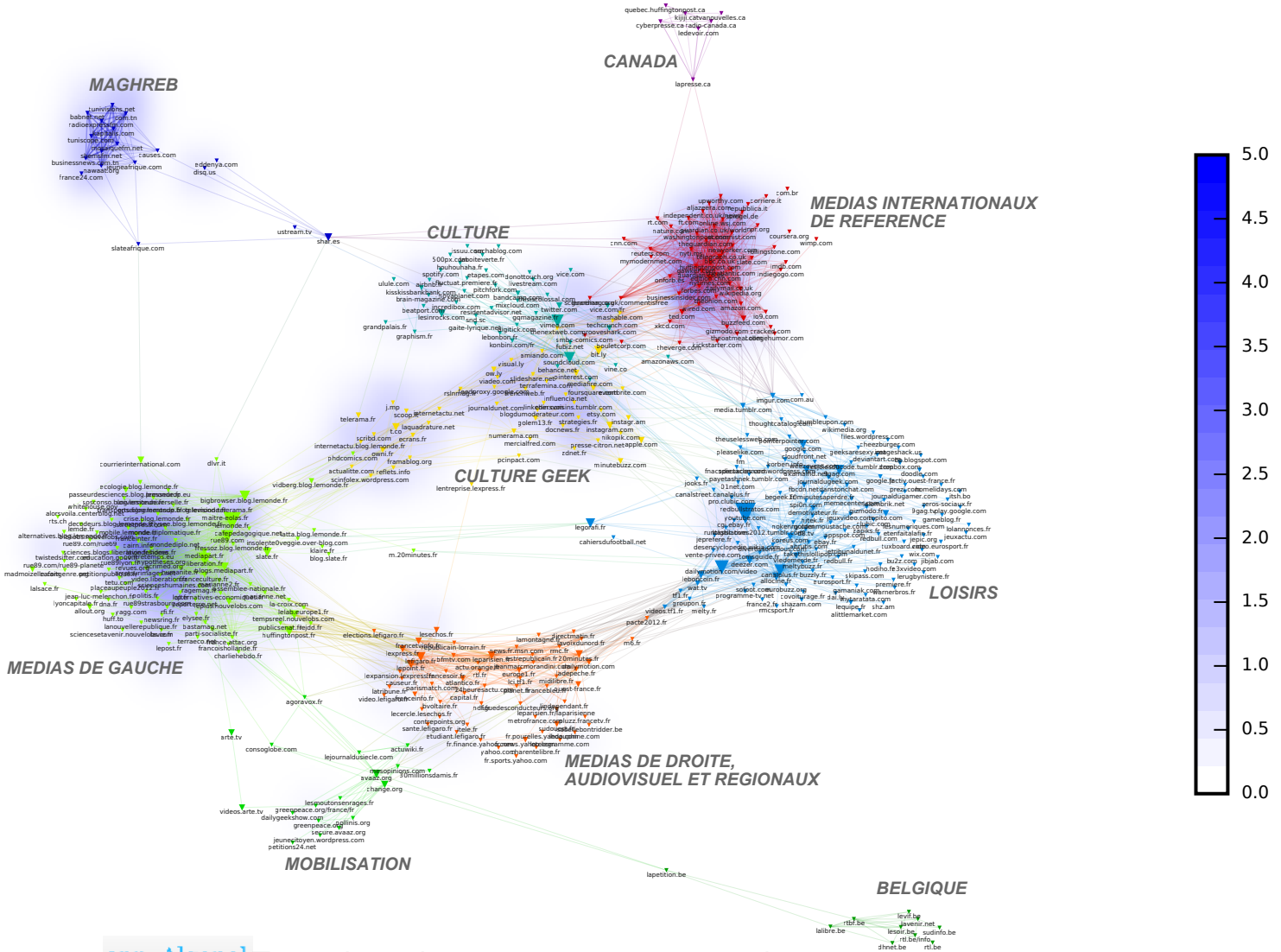
Translational Research Triangle



- **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

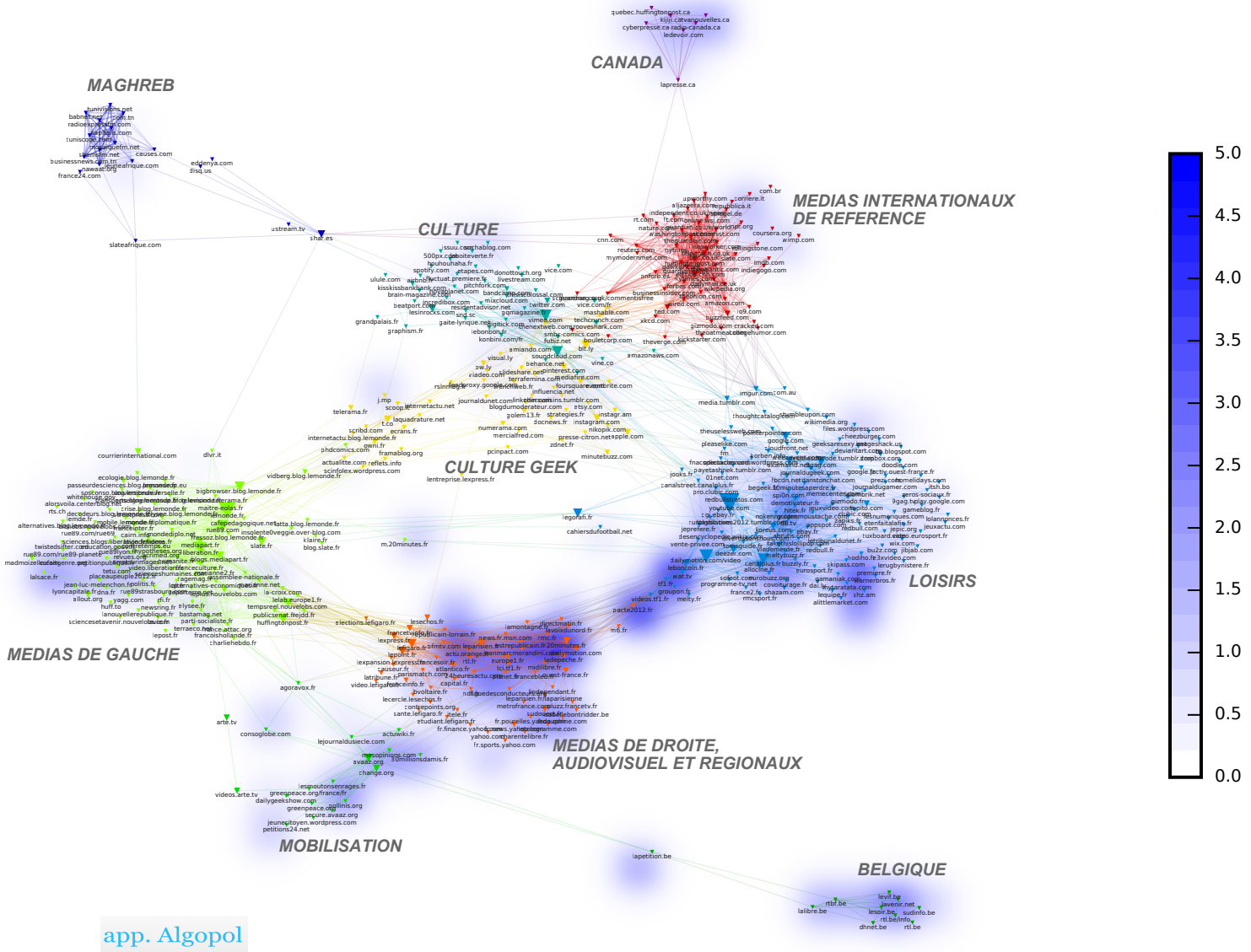


app. Algopol

Facebook citation network

CORTEXT

Mixing dimensions of analysis

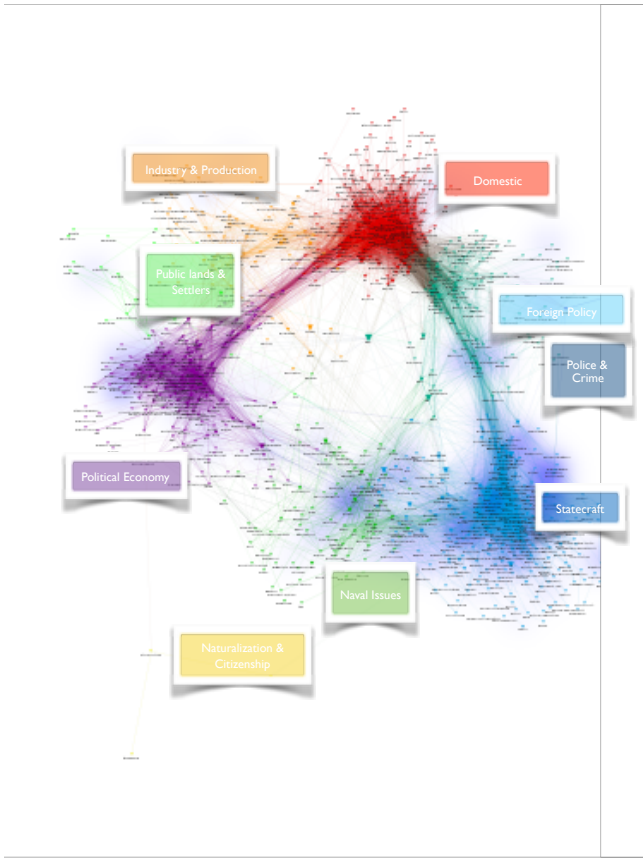


app. Algopol

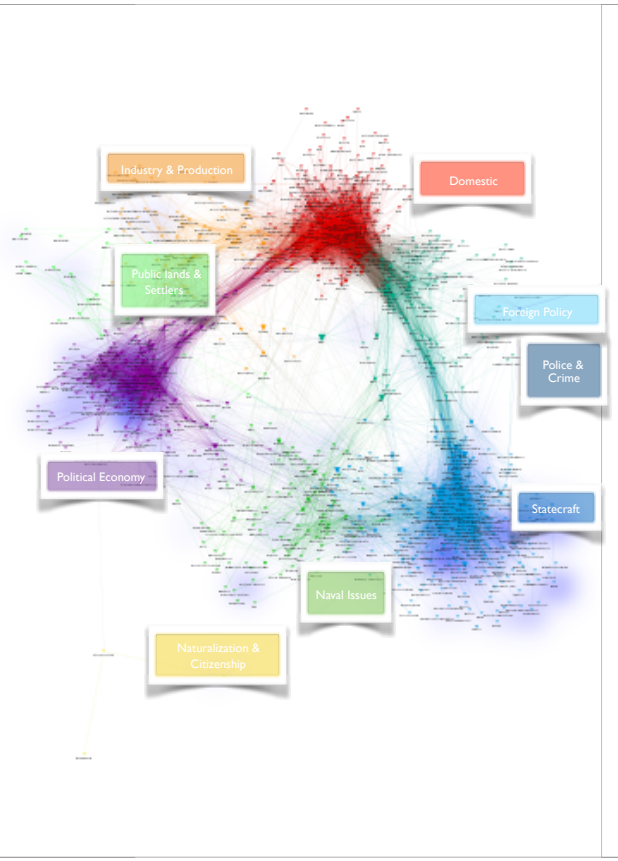
CORTEXT

Mapping individual presidents

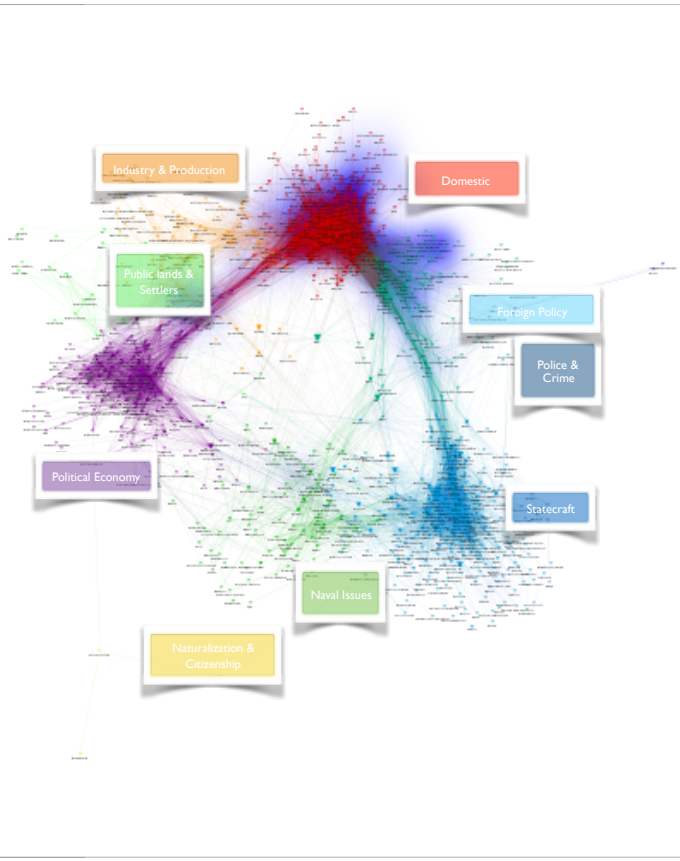
George Washington, 0-3



Abraham Lincoln, 0-3



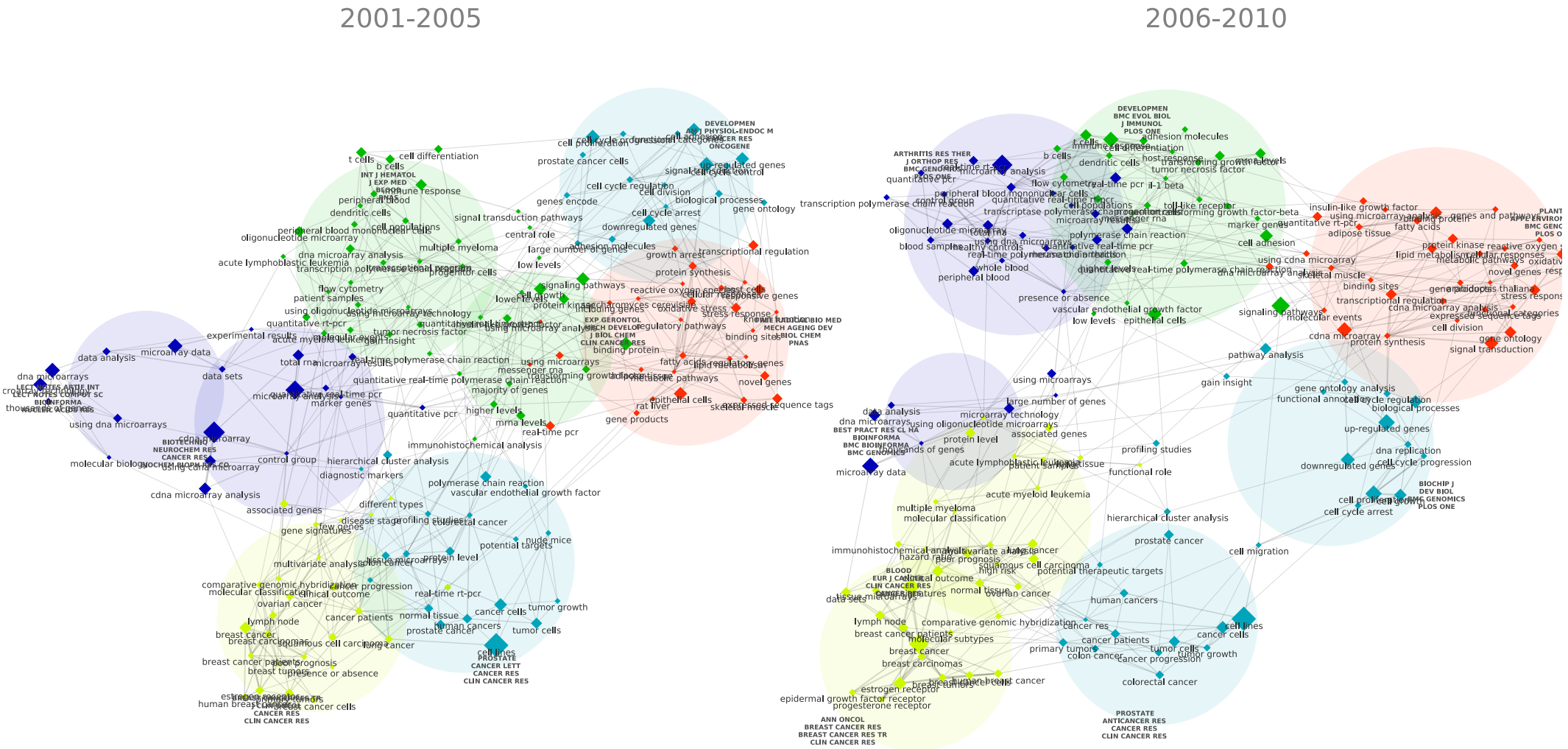
Ronald Reagan, 0-3



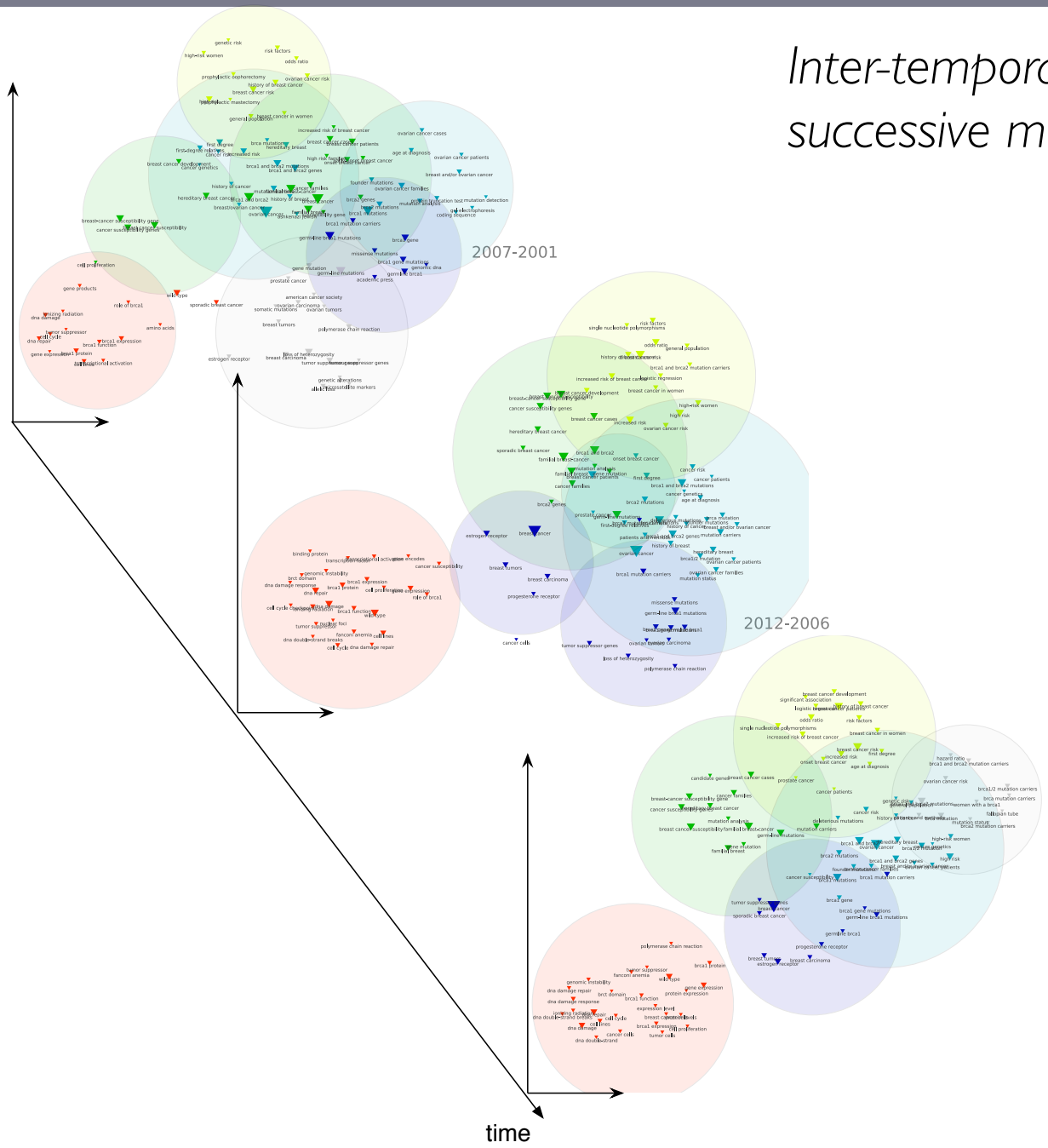
Comparing political discourses across time

Dynamical mapping

Inter-temporal matching between two maps



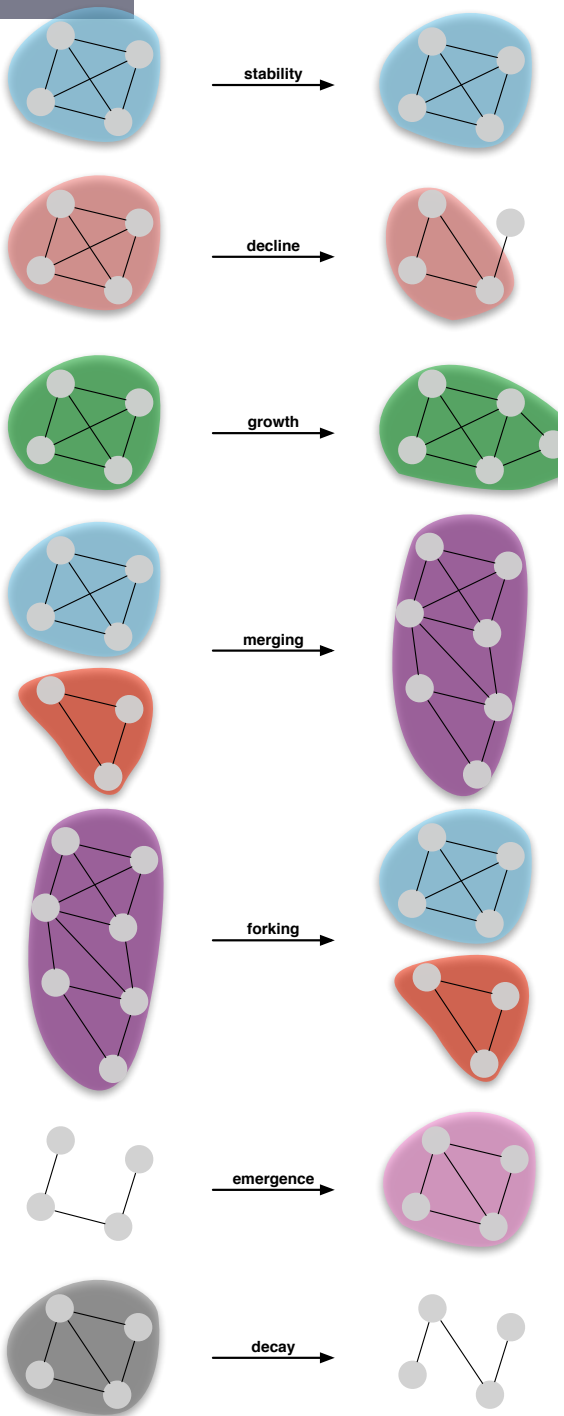
Dynamical mapping



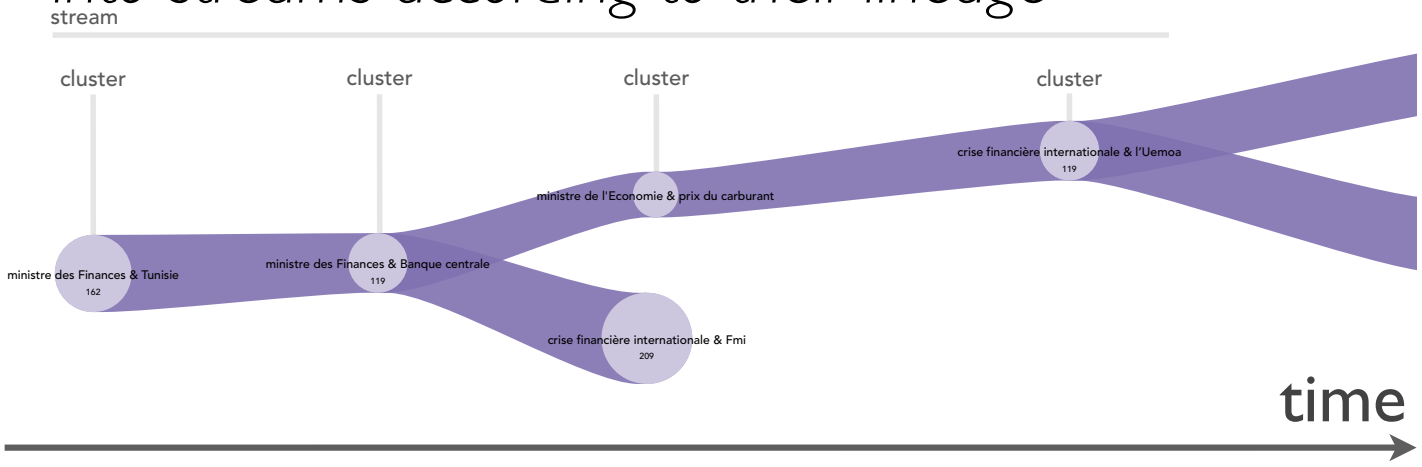
Inter-temporal matching between successive maps

Stream Reconstruction

Possible events at mesoscopic level

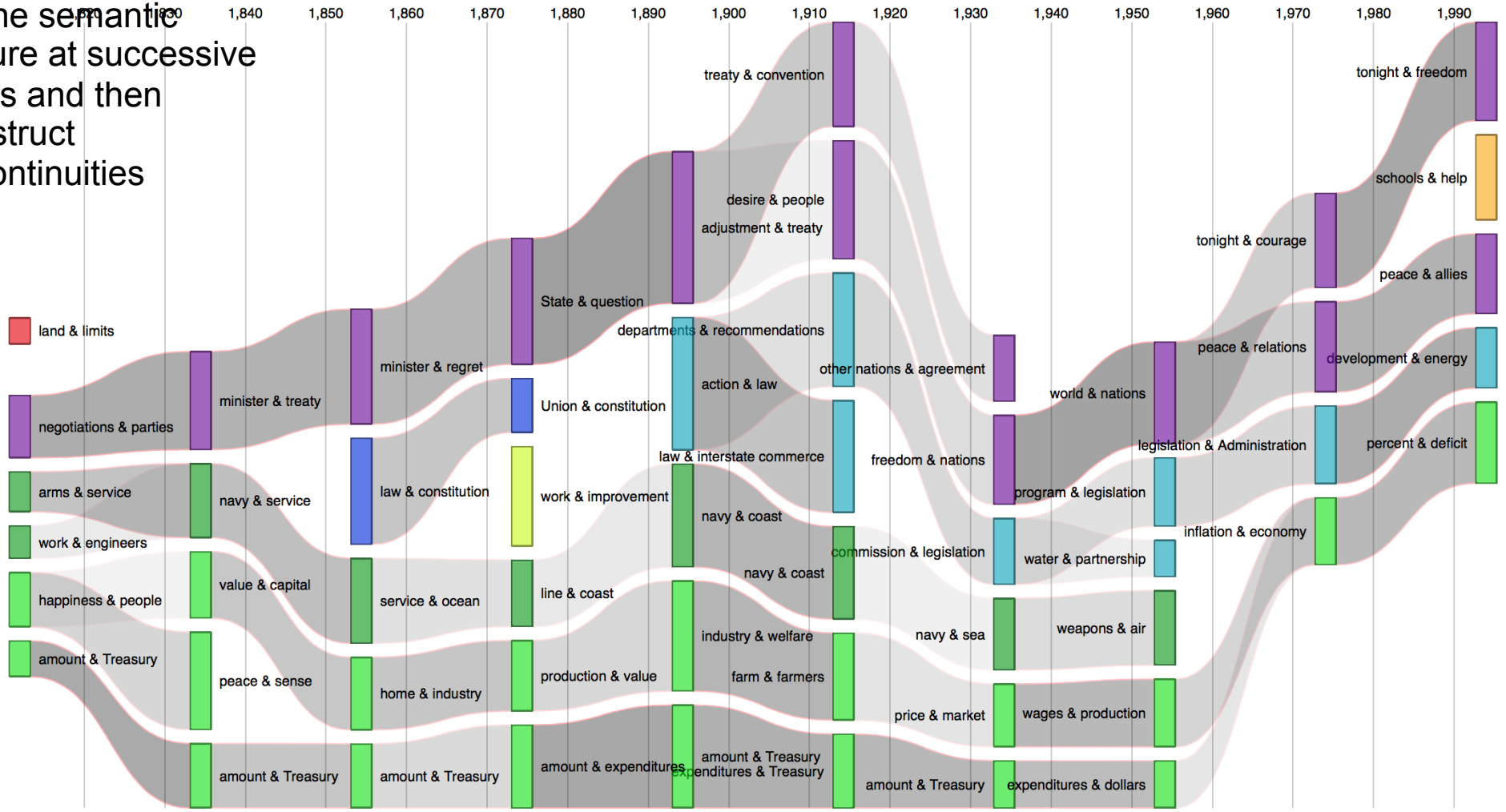


Clusters are intertemporally grouped into streams according to their lineage



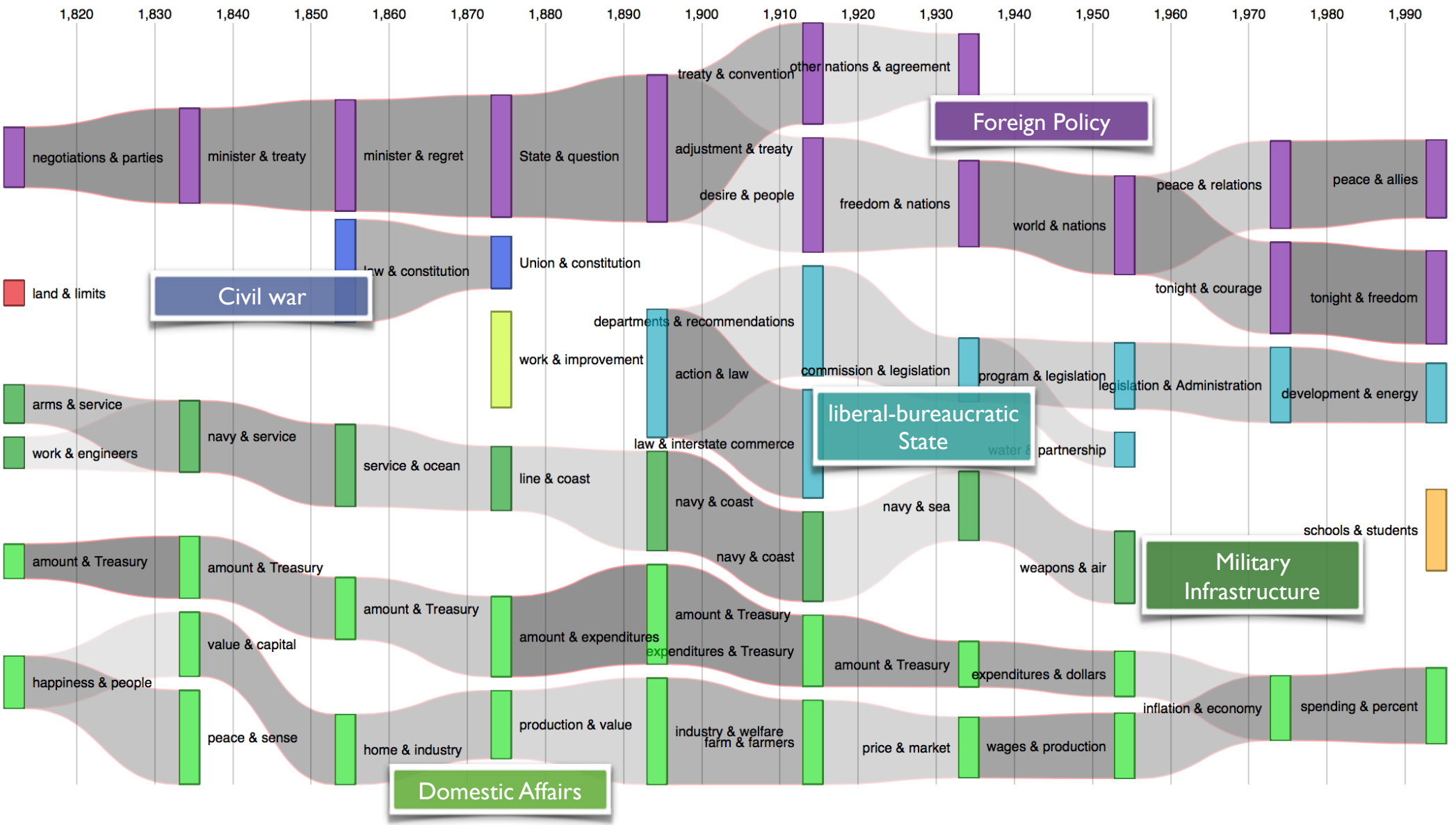
River Networks

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

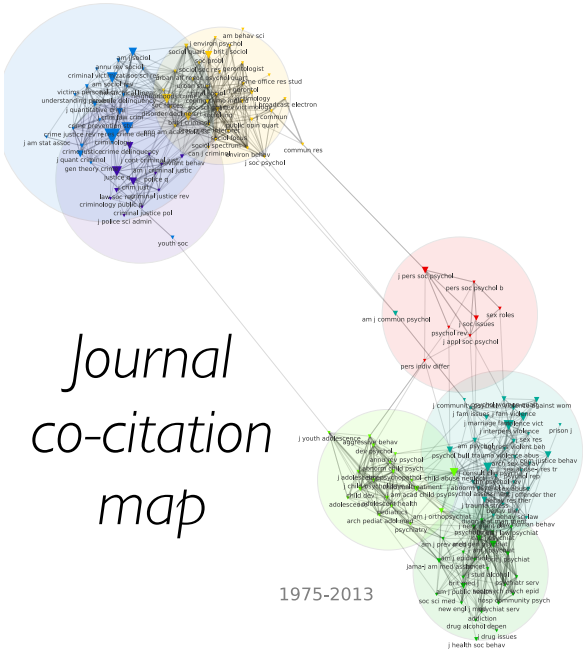


River Network

River Network

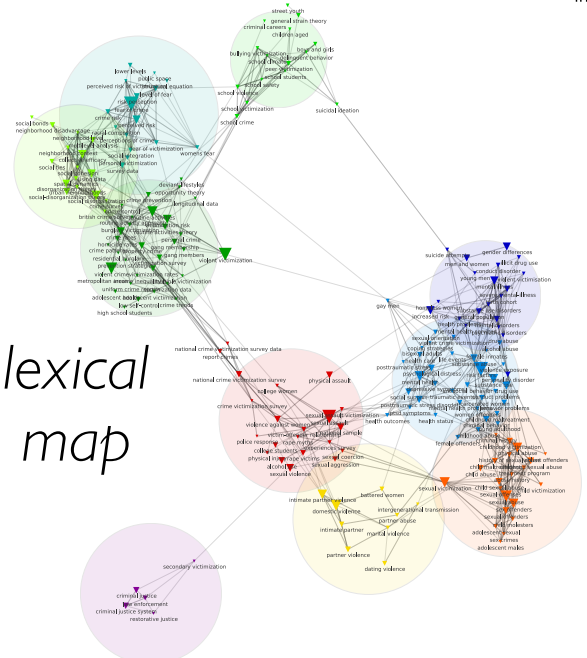


Mixing dimensions of analysis

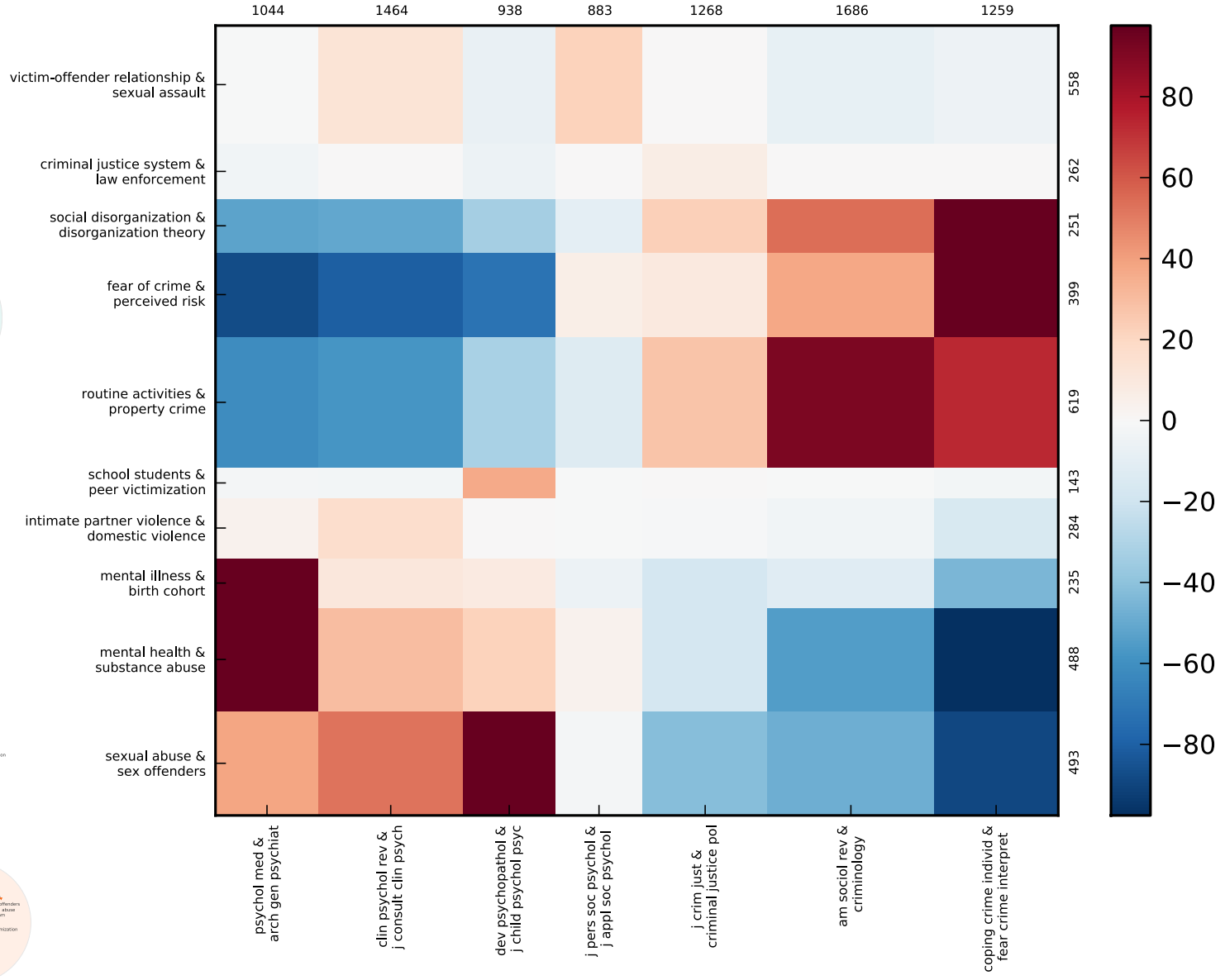


Journal co-citation map

1975-2013



lexical map



crime science corpus