

RECONSTRUCTION AND MONITORING OF THE SCIENTIFIC DEBATES FROM DIGITAL TRACES THE CASES OF SCIENCE AND POLITICS

David Chavalarias

Directeur de Recherche CNRS

EHESS/CAMS & Institut des Systèmes Complexes de Paris IdF

<http://chavalarias.com>

SEMANTIC ANALYSIS FOR INNOVATION POLICY

OECD - March 12 2018



DIGITAL SPACES & AND SOCIAL BIG DATA

Velocity

TIME SCALES

Seconde

- > 6k pageview/s
- > 40k/s requêtes



Minutes

- > 347k/s WhatsApp messages
- > 280k tweets/min
- > 375k SMS/min



Hours

- > 4k/h blog post
- > 6,25k edits/h on Wikipedia



Days

- > 3,2k AFP news/d



Years

- 79k/y french papers
- (> 2M peer reviewed papers/an)



Audience

DIGITAL SPACES & AND SOCIAL BIG DATA

Telephony data
mobility,
consumption,
medical data,
navigation data,
employment data
engagement data (petitions),
bank data,
sexual preferences (meeting sites),
social relationships (address books,
social networks, e-mails, etc.),
professional relationships, etc...

Velocity

TIME SCALES

Seconde

- > 6k pageview/s
- > 40k/s requêtes



Minutes

- > 347k/s WhatsApp messages
- > 280k tweets/min
- > 375k SMS/min



Hours

- > 4k/h blog post
- > 6,25k edits/h on Wikipedia



Days

- > 3,2k AFP news/d



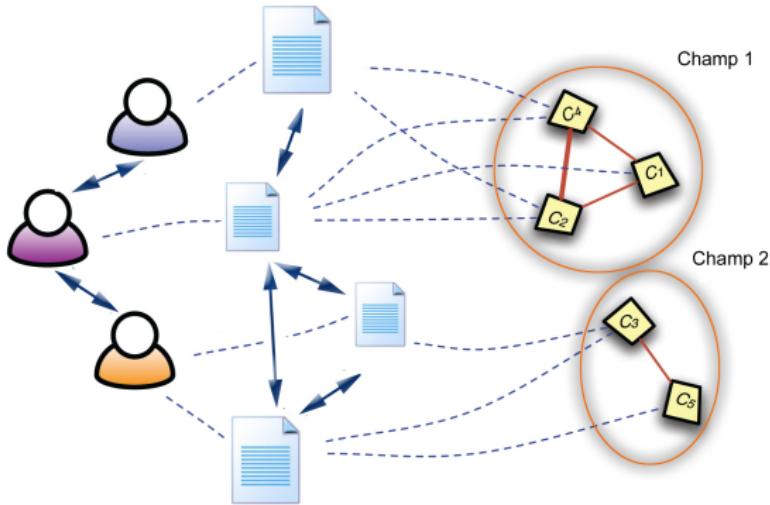
Years

- 79k/y french papers
- (> 2M peer reviewed papers/an)

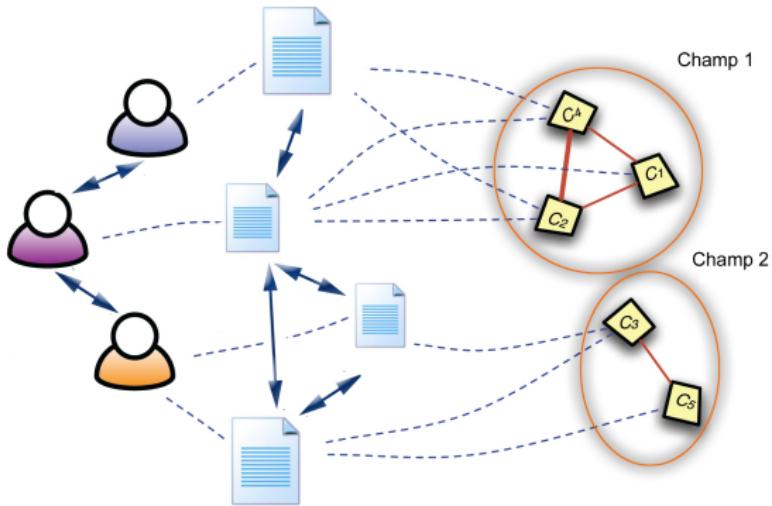


Audience

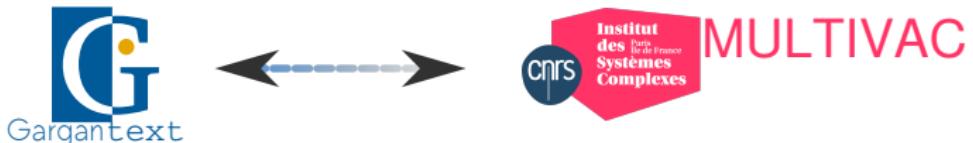
Multi-partite structures



Multi-partite structures



CNRS/ISC-PIF free and open softwares :
<http://gargantext.iscpif.fr> & <http://multivac.iscpif.fr>



HOW TO RECONSTRUCT SOCIO-SEMANTIC NETWORKS FROM DIGITAL DATA ?

How to represent knowledges spaces ?

“Knowledge is impossible to draw as a whole in a truthful manner, but only through the choice of a point of view that is both arbitrary and inevitable [...] One can create as many different systems of human knowledge as there are world maps having different projections, and each one of these systems might even have some particular advantage possessed by none of the others.”

D'Alembert (1751) *L'Encyclopédie/1ère édition/Discours préliminaire*

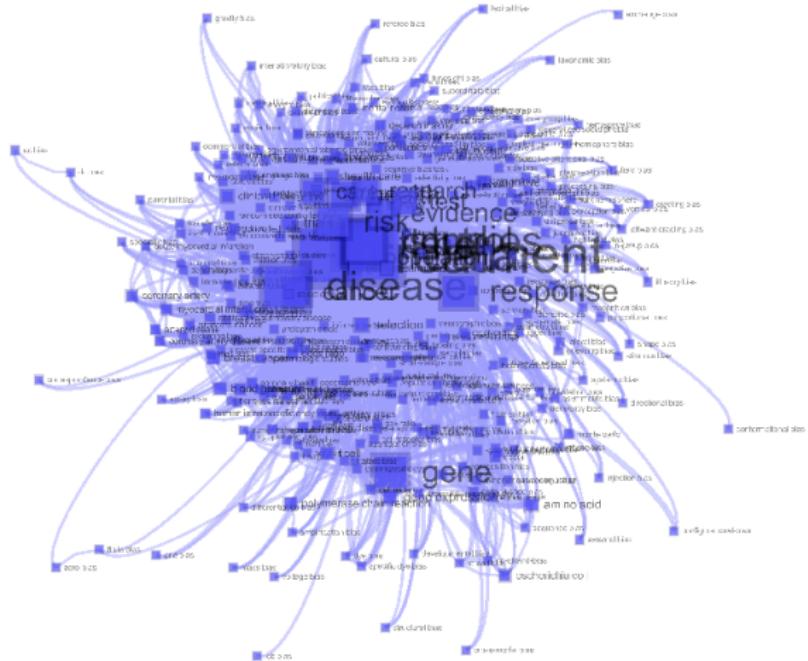
How to represent knowledges spaces ?

“Knowledge is impossible to draw as a whole in a truthful manner, but only through the choice of a point of view that is both arbitrary and inevitable [...] One can create as many different systems of human knowledge as there are world maps having different projections, and each one of these systems might even have some particular advantage possessed by none of the others.”

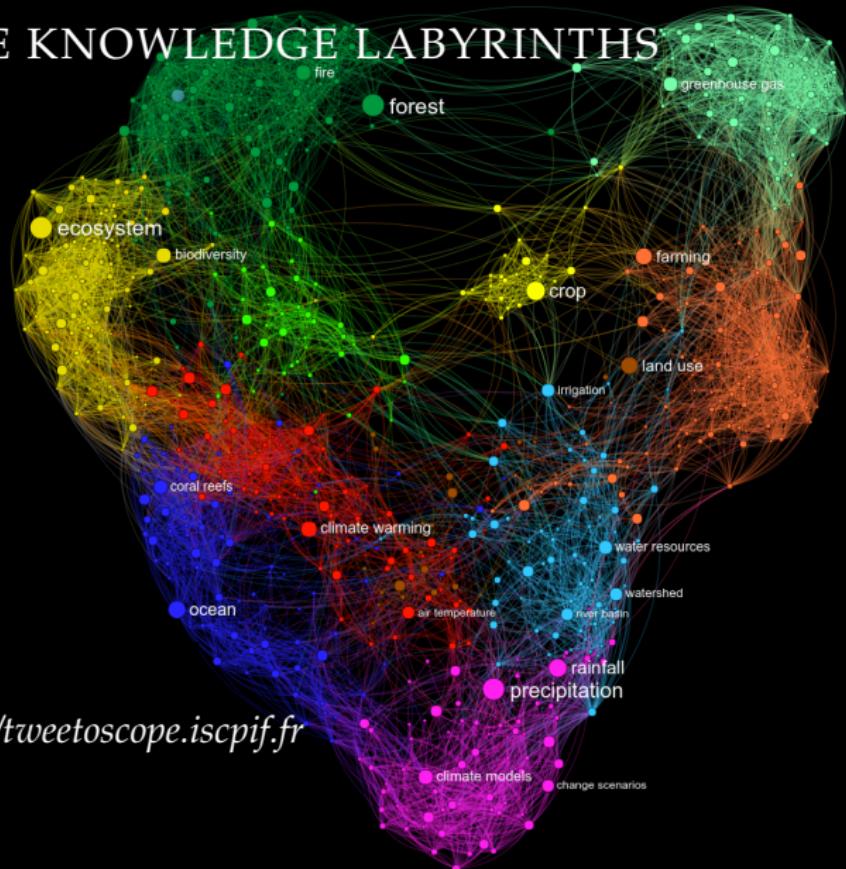
D'Alembert (1751) *L'Encyclopédie/1ère édition/Discours préliminaire*

Use complementary approaches, work with multi-views.
Here: text-mining & complex networks analysis

Seeing everything is like seeing nothing



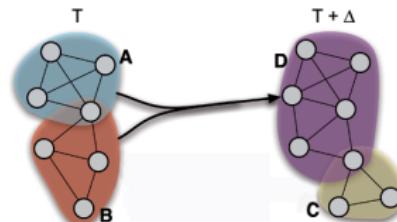
THE KNOWLEDGE LABYRINTHS



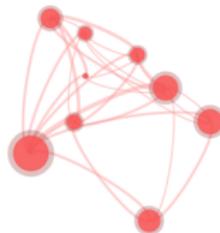
<http://tweetoscope.iscpif.fr>

Multi-scale structure of scientific fields

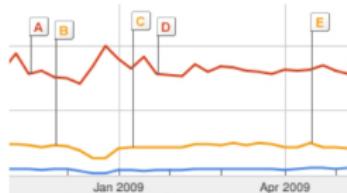
Macro Phylometry



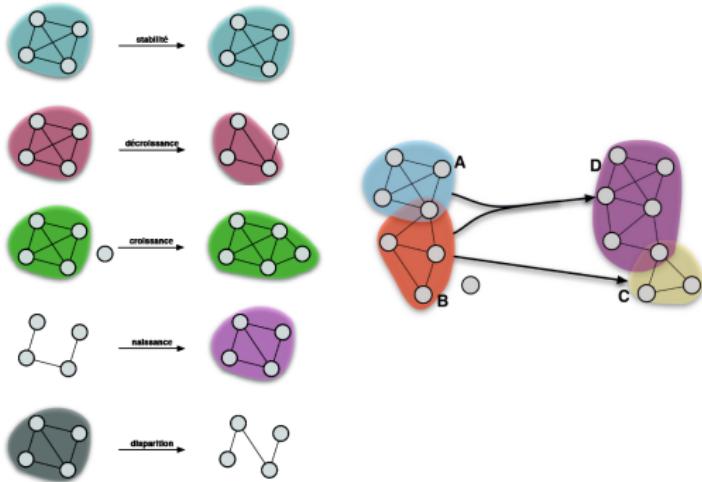
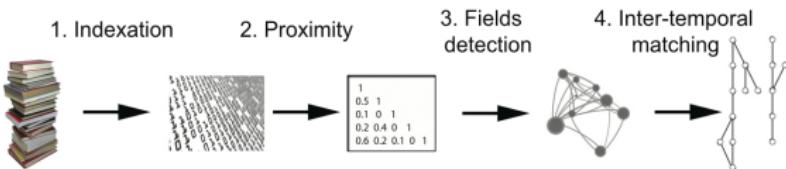
Meso clusters



Micro terms



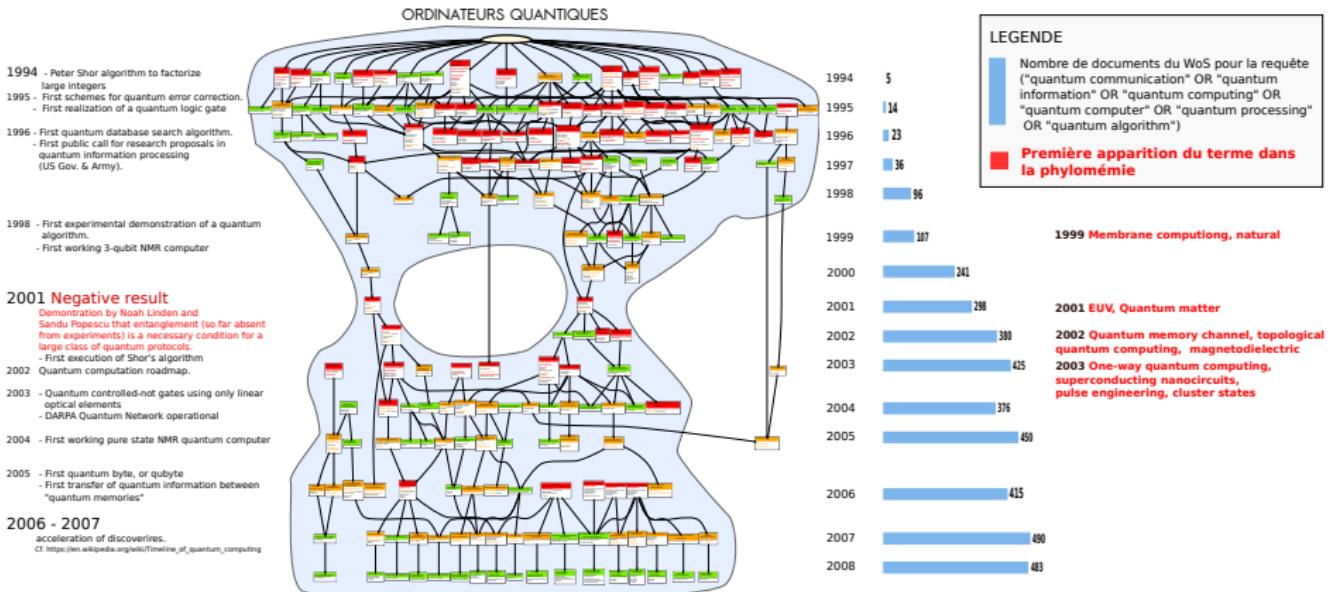
Meso-dynamics



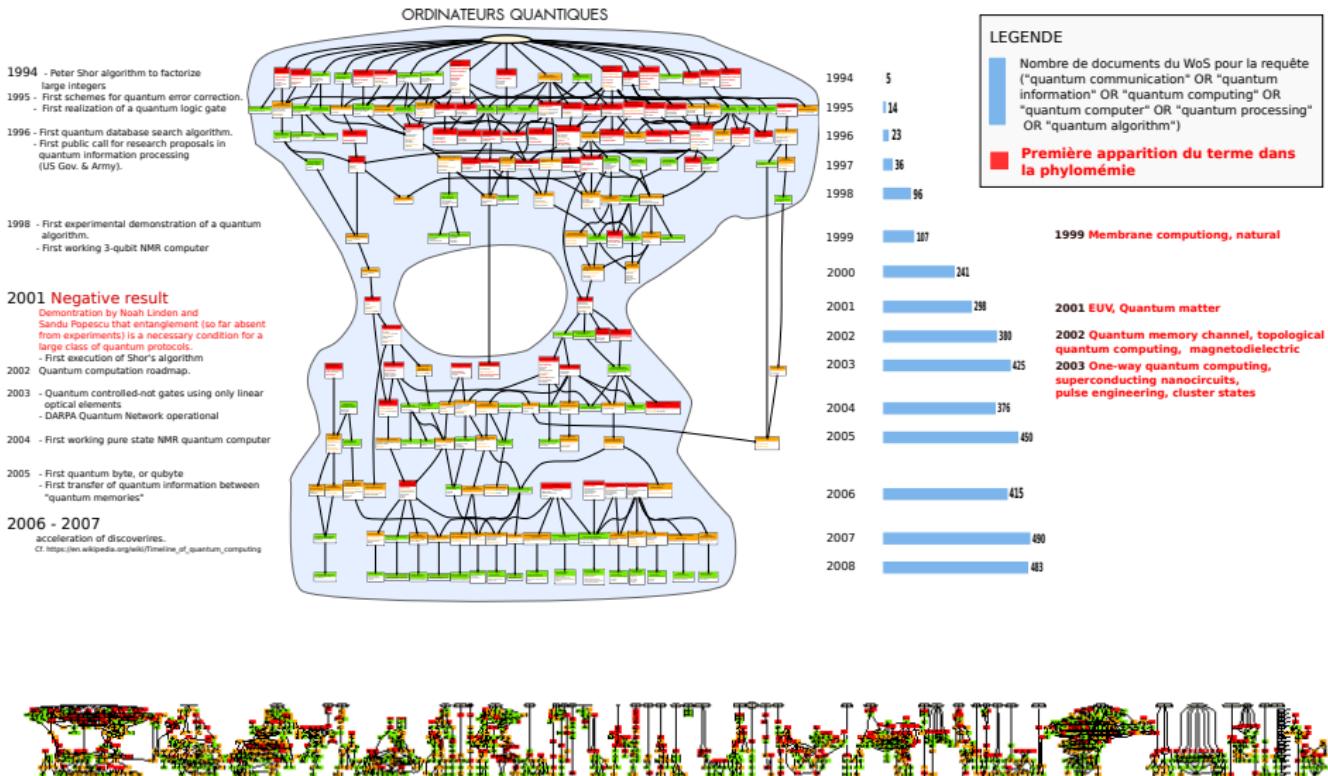
Chavalarias, D., Cointet, J.-P., 2013. Phylogenetic patterns in science evolution—the rise and fall of scientific fields. PloS one 8, e54847.

David Chavalarias, Institut des Systèmes Complexes de Paris Ile-de-France - <http://iscipif.fr>;

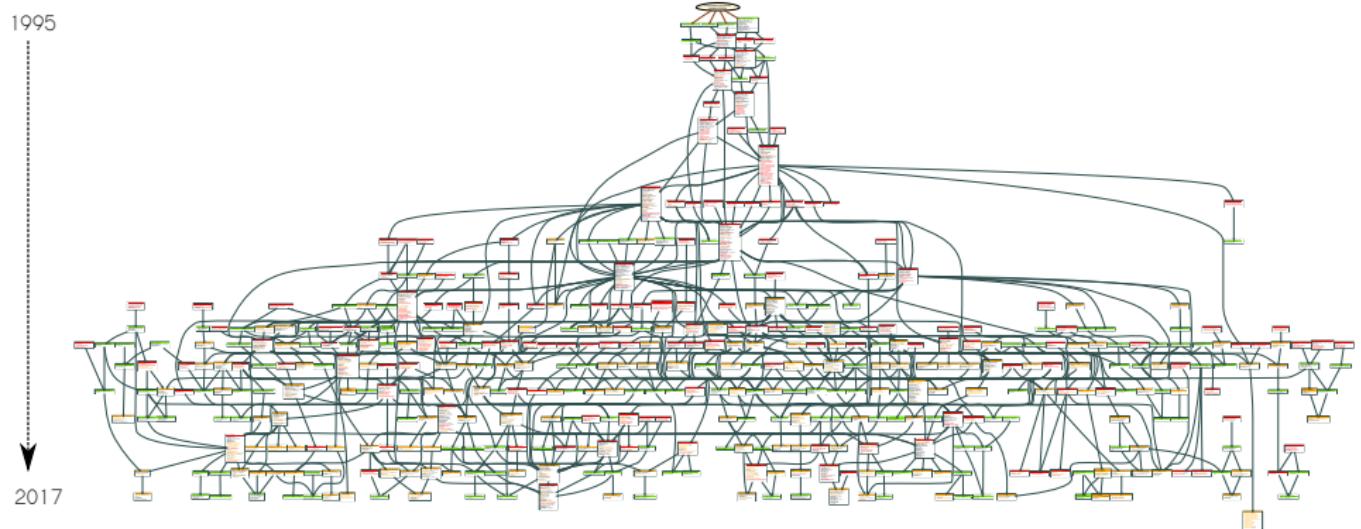
Phylomémie des sciences (Chavaliarias 2016 - HDR)



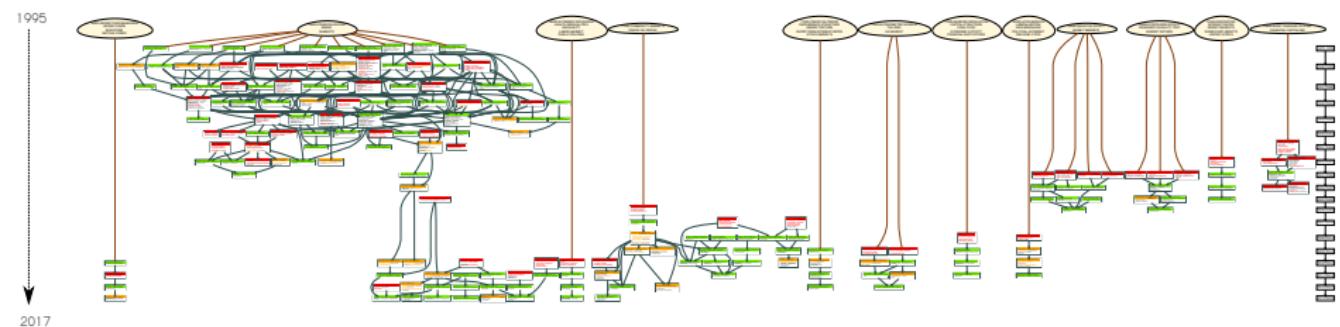
Phylomémie des sciences (Chavaliarias 2016 - HDR)



FINANCIAL CRISIS



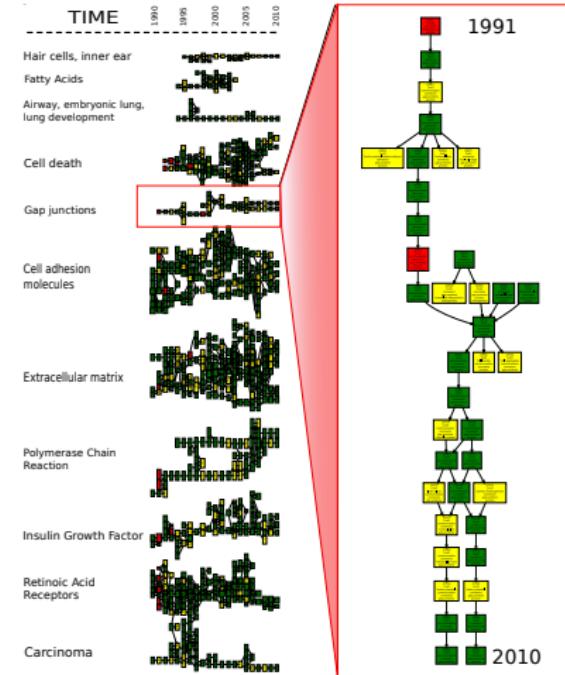
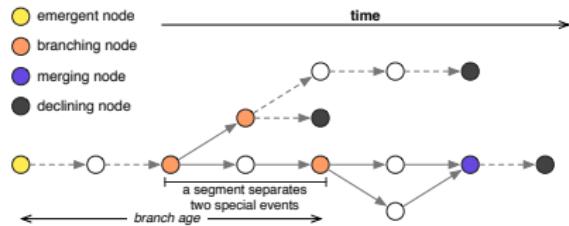
FINANCIAL CRISIS



Phylomemries : reconstruction of science dynamics

Chavalarias and Cointet 2013, PLoS One.

- ▶ Text-mining on large digital corpora (ex Embryology science > 200k papers),

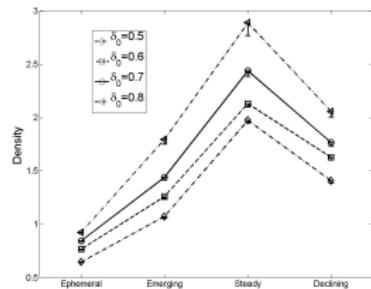


Phylomemies : reconstruction of science dynamics

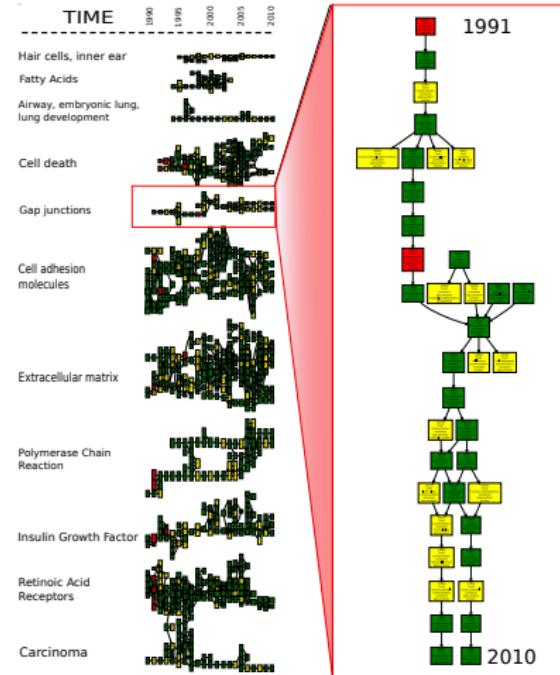
Chavalarias and Cointet 2013, PLoS One.

- ▶ Text-mining on large digital corpora (ex Embryology science > 200k papers),

Results : Strong regularities discovered in the evolution of the domains studied.



Correlations between fields consistency and position in the phylomemy.

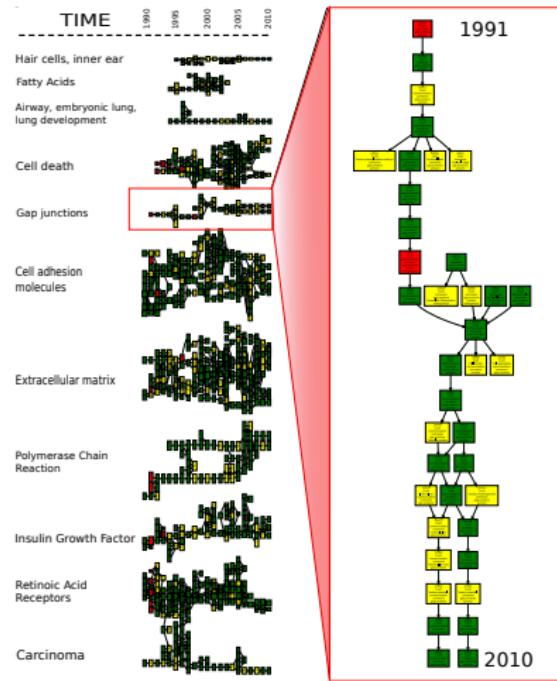
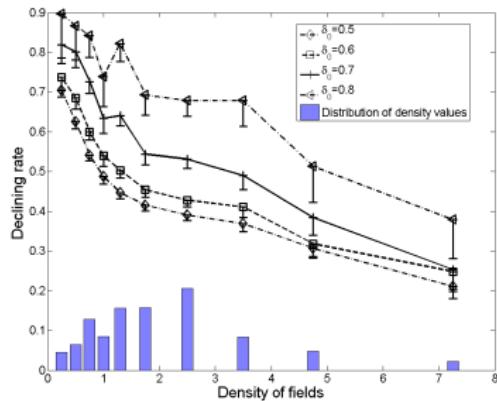


Phylomemries : reconstruction of science dynamics

Chavalarias and Cointet 2013, PLoS One.

- ▶ Text-mining on large digital corpora (ex Embryology science > 200k papers),

Empirical probability to be declining in function of the density of fields belonging to the phylogenetic network.



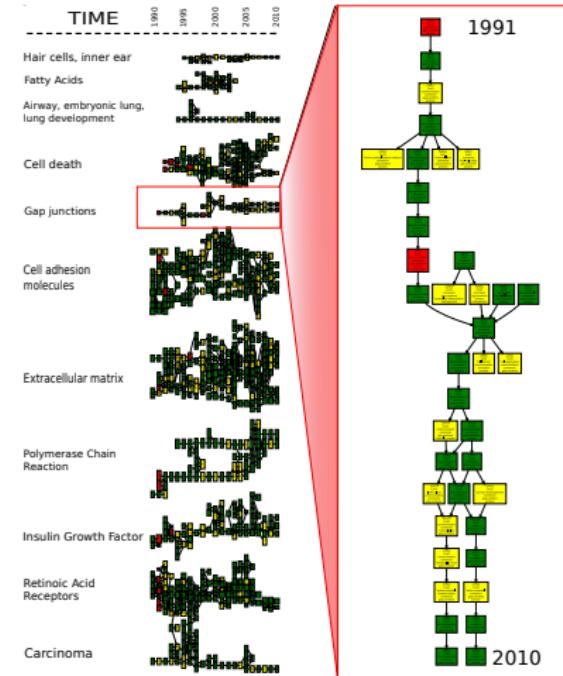
Phylomemries : reconstruction of science dynamics

Chavalarias and Cointet 2013, PLoS One.

- ▶ Text-mining on large digital corpora (ex Embryology science > 200k papers),

Questions :

- ▶ From computer science view point : can we scale this analysis to the whole science ? To other media ? How to study the multi-scale structure.
- ▶ From epistemological view point: What are the different evolutionary patterns across the different domains of science ?
- ▶ From digital humanities view point: will this pave the ways toward new ways of interacting with large digitalized corpora ?



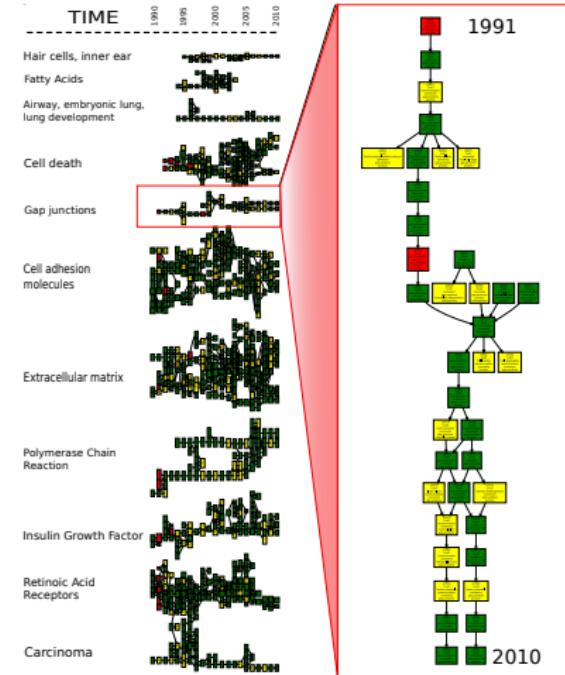
Phylomemries : reconstruction of science dynamics

Chavalarias and Cointet 2013, PLoS One.

- ▶ Text-mining on large digital corpora (ex Embryology science > 200k papers),

Questions :

- ▶ **From computer science view point :** can we scale this analysis to the whole science ? To other media ? How to study the multi-scale structure.
- ▶ **From epistemological view point:** What are the different evolutionary patterns across the different domains of science ?
- ▶ **From digital humanities view point:** will this pave the ways toward new ways of interacting with large digitalized corpora ?



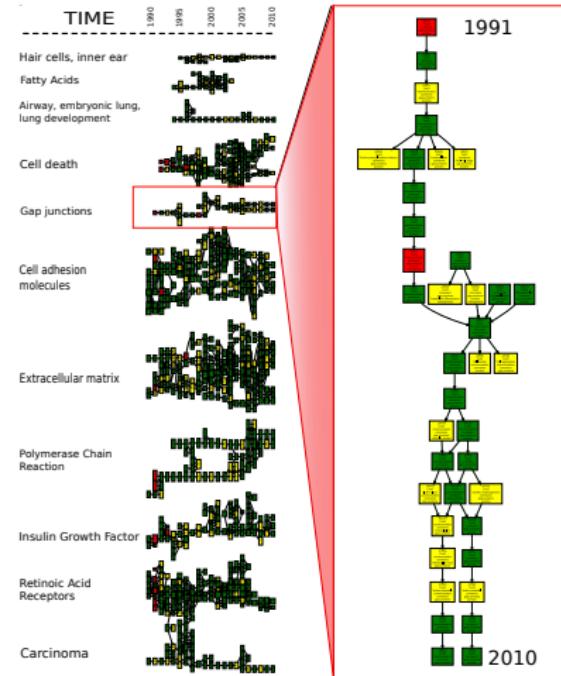
Phylomemries : reconstruction of science dynamics

Chavalarias and Cointet 2013, PLoS One.

- ▶ Text-mining on large digital corpora (ex Embryology science > 200k papers),

Questions :

- ▶ **From computer science view point :** can we scale this analysis to the whole science ? To other media ? How to study the multi-scale structure.
- ▶ **From epistemological view point:** What are the different evolutionary patterns across the different domains of science ?
- ▶ **From digital humanities view point:** will this pave the ways toward new ways of interacting with large digitalized corpora ?





#POLITOSCOPE



92 029 538

Tweëts politiques



55 465 066

Tweets media



4 695 163

Comptes twitter

Cliquez ici pour une plongée dans les données de la présidentielle ➔

ANALYSIS OF ONLINE SOCIAL STRUCTURES

with

Noé Gaumont & Maziyar Panahi

<http://politoscope.org>

DES RÉSEAUX SOCIAUX DANS NOS URNES



LES CANDIDATS ET
LEURS COMMUNAUTÉS



FAC à FACE
DU SECOND TOUR



ATTENTION DES
COMMUNAUTÉS

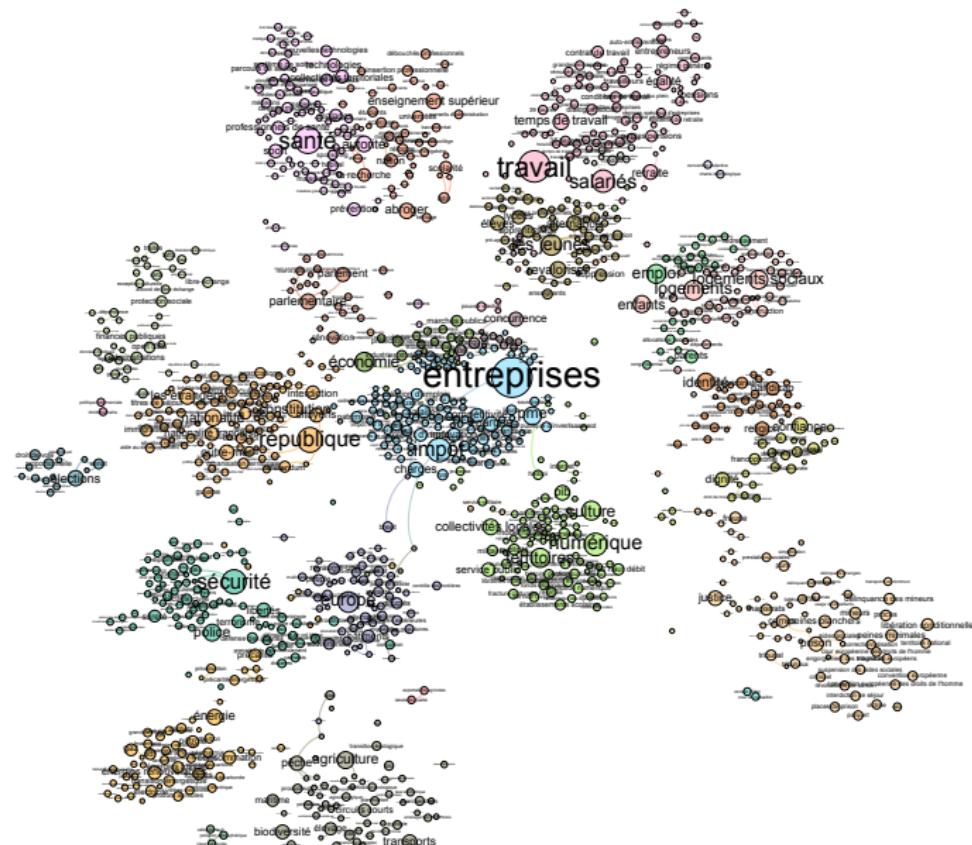
Département 2016, TBC-PIF analyse les interactions de plus de 2000 comptes Twitter lors de la campagne présidentielle. Environ 35 000 discussions, mentions, retweets, citations, ou messages directes ont été générées par les deux candidats. D'ores et déjà, nous étudions les réactions des réseaux sociaux aux résultats du scrutin. Une fois d'abord poser sur la nature des interactions et leur contenu en priorité, nous devons également étudier les réseaux sociaux et les interactions entre les deux candidats.

PLATEFORMES INTERACTIVES

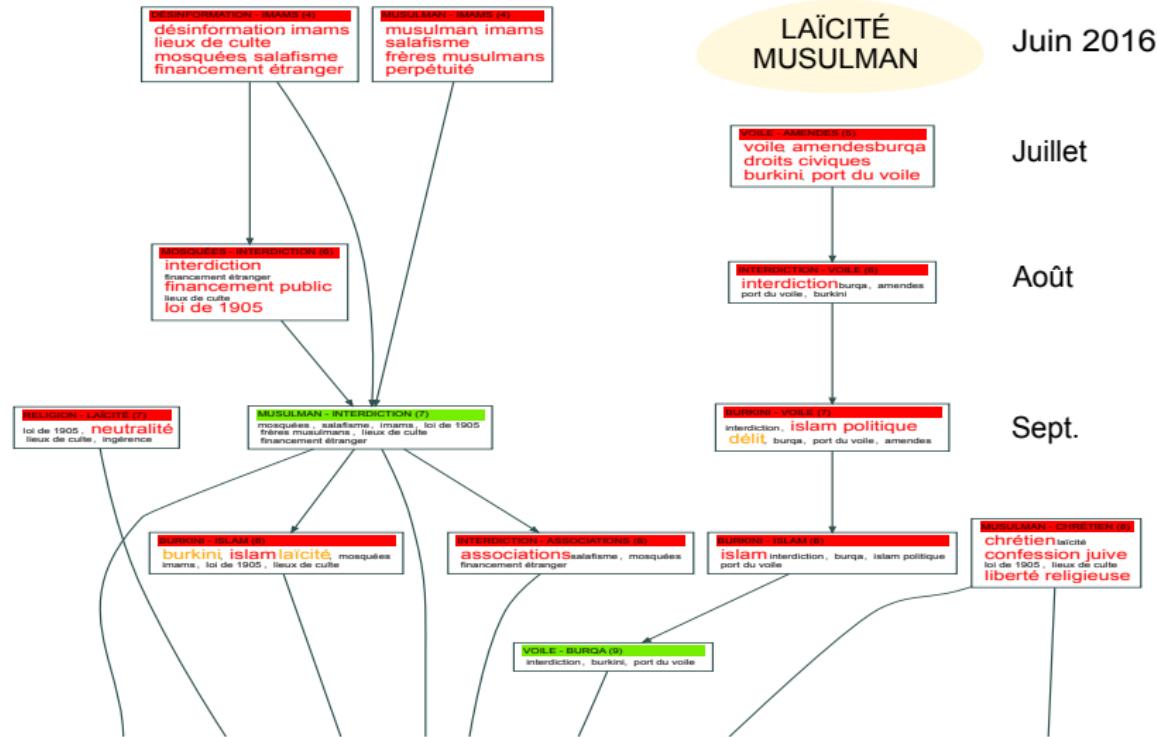
MÉTHODOLOGIE
La presse, les candidats et leurs communautés
POLITOSCOPE.ORG #PRÉSIDENTIELLE2017

The debates of the 2017 Presidential Elections

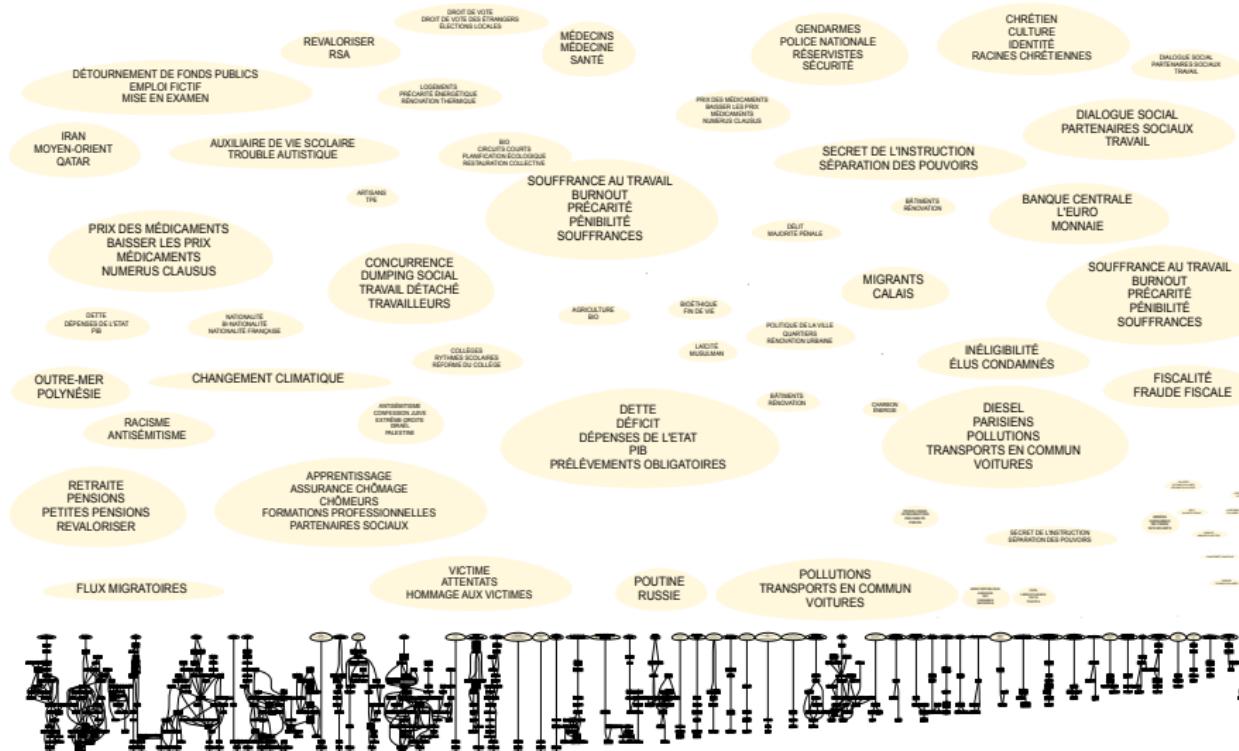
<http://gargantext.org>

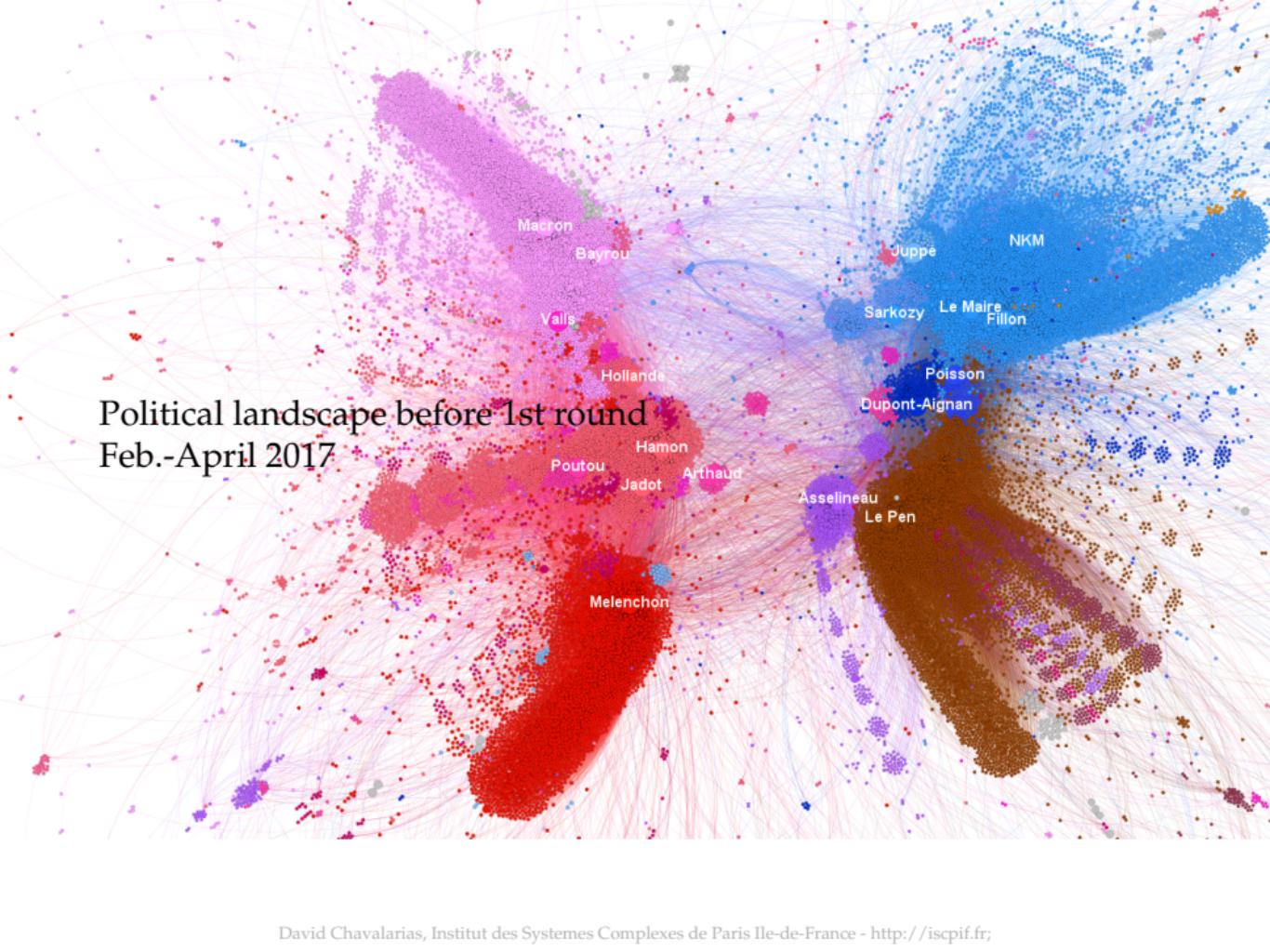


What are the debates of the 2017 Presidential Elections ?



What are the debates of the 2017 Presidential Elections ?



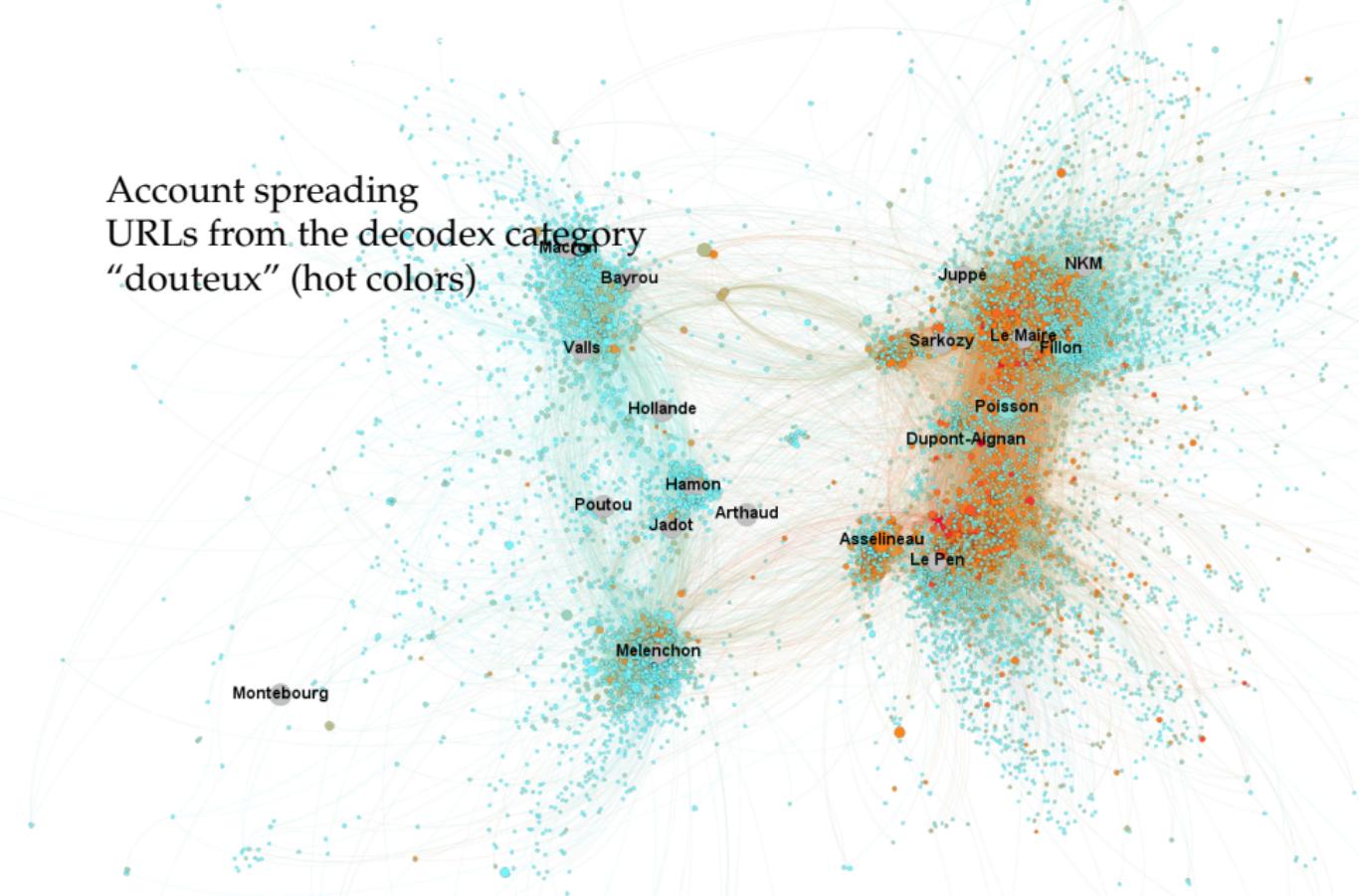


A network graph visualization showing political connections between various political figures and entities before the 2017 French presidential election. The nodes are color-coded by group, and the edges represent interactions or relationships.

Political landscape before 1st round
Feb.-April 2017

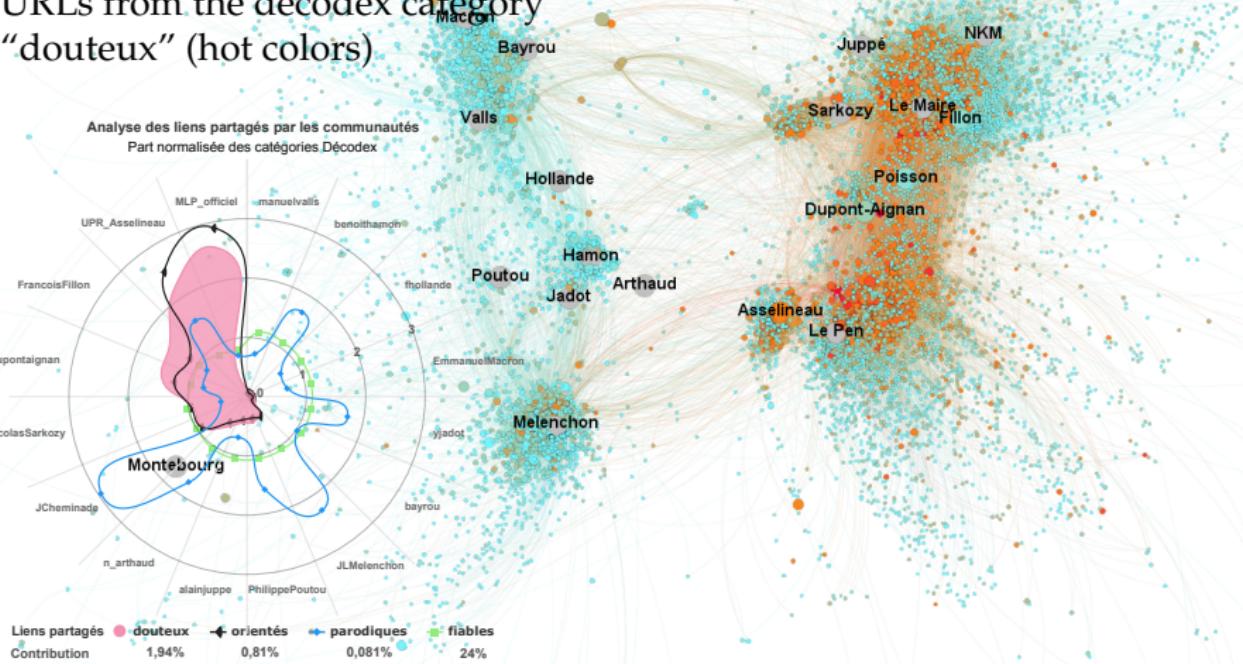
Macron
Bayrou
Valls
Hollande
Poutou
Jadot
Hamon
Arthaud
Melenchon
Asselineau
Le Pen
Poisson
Dupont-Aignan
Juppé
NKM
Sarkozy
Le Maire
Fillon

Account spreading URLs from the decodeX category “douteux” (hot colors)

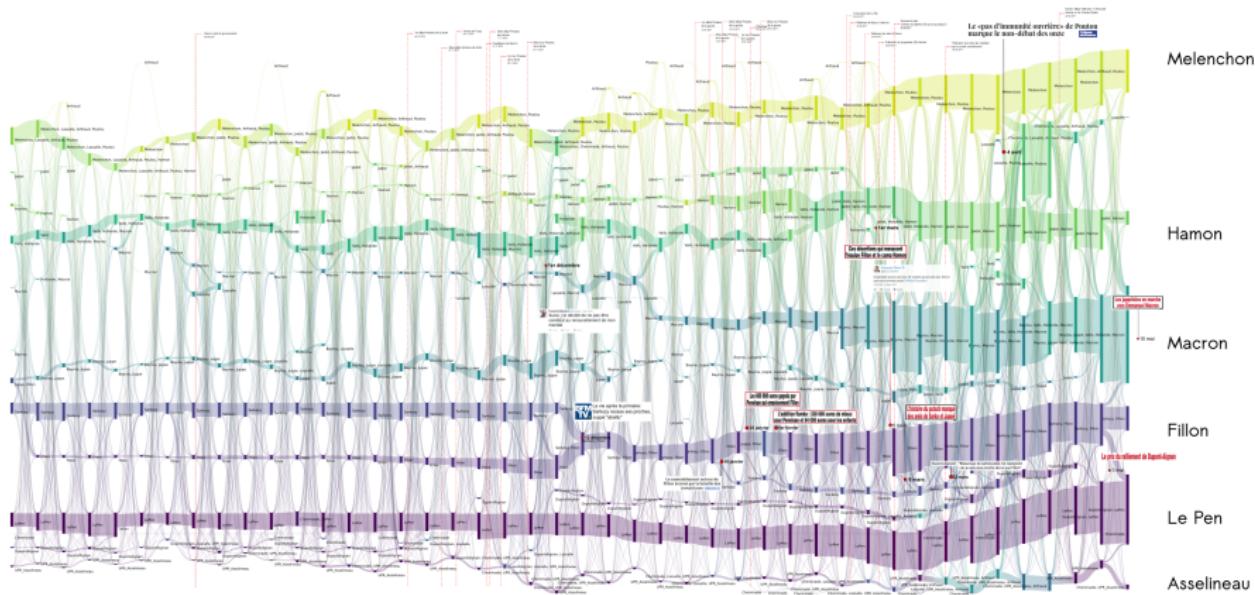


Account spreading URLs from the decodeX category “douteux” (hot colors)

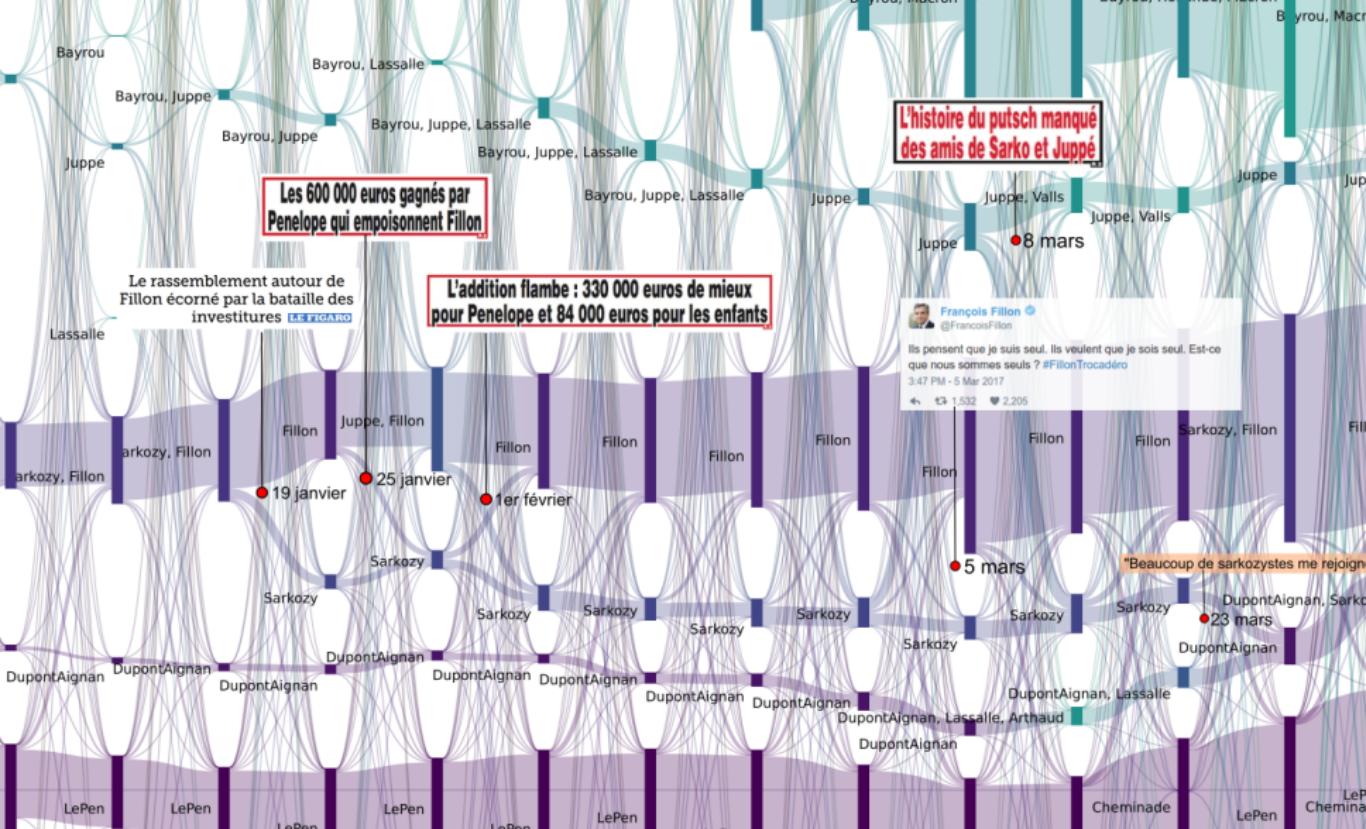
Analyse des liens partagés par les communautés
Part normalisée des catégories Décodex



Evolution of the political landscape

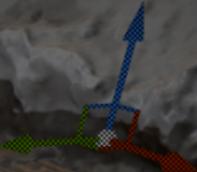


Noe Gaumont, Maziyar Panahi, David Chavalarias (forthcoming 2018) Reconstruction of the socio-semantic dynamics of political activist Twitter networks – Method and application to the 2017 French Presidential election. <hal-01575456>



STALAGMEMES

Le Rêve des Formes, Palais de Tokyo, Paris



From 7 to 77 years old
<http://stalagmemes.org>

THANKS FOR YOUR ATTENTION



David CHAVALARIAS

Web : <http://chavalarias.com>

In collaboration with : Alexandre
Delanoë, Noé Gaumont, Maziyar
Panahi.



Rendez-vous visite
iscpif.fr

How to represent knowledge ?

[...] rassembler dans le plus petit espace possible, & à placer, pour ainsi dire, le Philosophe *au-dessus de ce vaste labyrinthe dans un point de vûe fort élevé d'où il puisse appercevoir à la fois les Sciences & les Arts principaux ; voir d'un coup d'œil les objets de ses spéculations, & les opérations qu'il peut faire sur ces objets ; distinguer les branches générales des connaissances humaines, les points qui les séparent ou qui les unissent ; & entrevoir même quelquefois les routes secrètes qui les rapprochent.* C'est *une espèce de Mappemonde* qui doit montrer les principaux pays, leur position & leur dépendance mutuelle, le chemin en ligne droite qu'il y a de l'un à l'autre ; chemin souvent coupé par mille obstacles, qui ne peuvent être connus dans chaque pays que des habitans ou des voyageurs, & qui ne sauroient être montrés que dans des cartes particulières fort détaillées. Ces cartes particulières seront les différens articles de notre Encyclopédie, & l'arbre ou système figuré en sera la mappemonde.

D'Alembert (1751) *L'Encyclopédie/1ère édition/Discours préliminaire*

How to represent knowledge ?

Mais comme dans les cartes générales du globe que nous habitons, les objets sont plus ou moins rapprochés, & présentent un coup d'œil différent selon le point de vûe où l'œil est placé par le Géographe qui construit la carte, de même la forme de l'arbre encyclopédique dépendra du point de vûe où l'on se mettra pour envisager l'univers littéraire . On peut donc imaginer autant de systèmes différens de la connaissance humaine, que de Mappemondes de différentes projections ; & chacun de ces systèmes pourra même avoir, à l'exclusion des autres, quelque avantage particulier.

D'Alembert (1751) L'Encyclopédie/1ère édition/Discours préliminaire