Collecter et analyser des données du web et des réseaux sociaux pour les SHS

Benjamin Ooghe-Tabanou

SciencesPo
MÉDIALAB

Collecter et analyser des données du web et des réseaux sociaux pour les SHS

Séminaire "Méthodologie de la recherche" du CIMEOS
Université de Bourgogne – Dijon – 10 novembre 2022

Benjamin Ooghe-Tabanou (@boogheta)
Sciences Po médialab (@medialab_ScPo)

Bruno Latour, fondateur du médialab

« Google is nice, but we need something better »

The Indian Express, 2011

médialab @ Sciences Po

https://medialab.sciencespo.fr

- Laboratoire de recherche SHS
  fondé par Bruno Latour en mai 2009,
  dirigé par Dominique Cardon depuis 2017

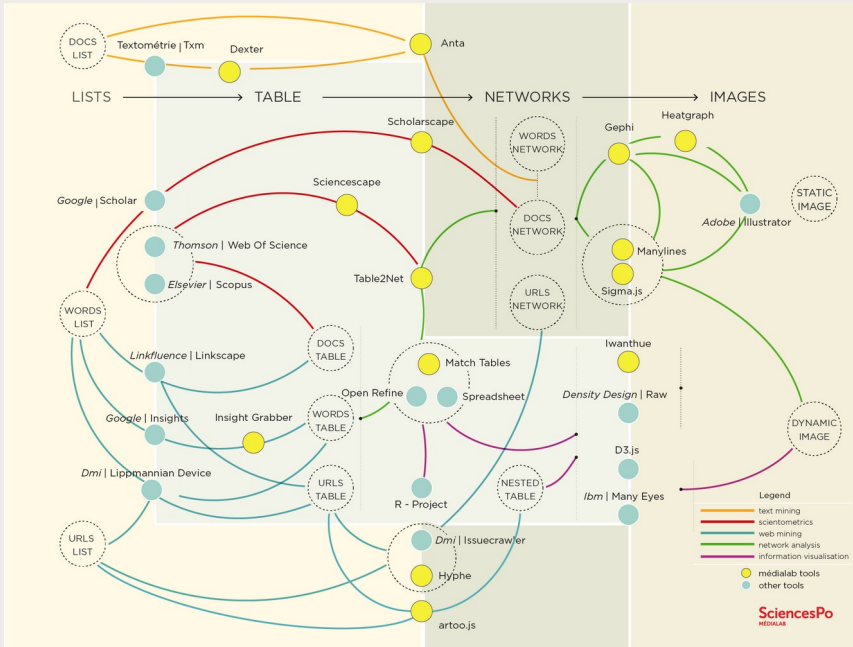
- Sciences Sociales, Ingénierie & Design
  → **interdisciplinarité**

- Articuler méthodes **quali & quanti**
  à travers une approche numérique

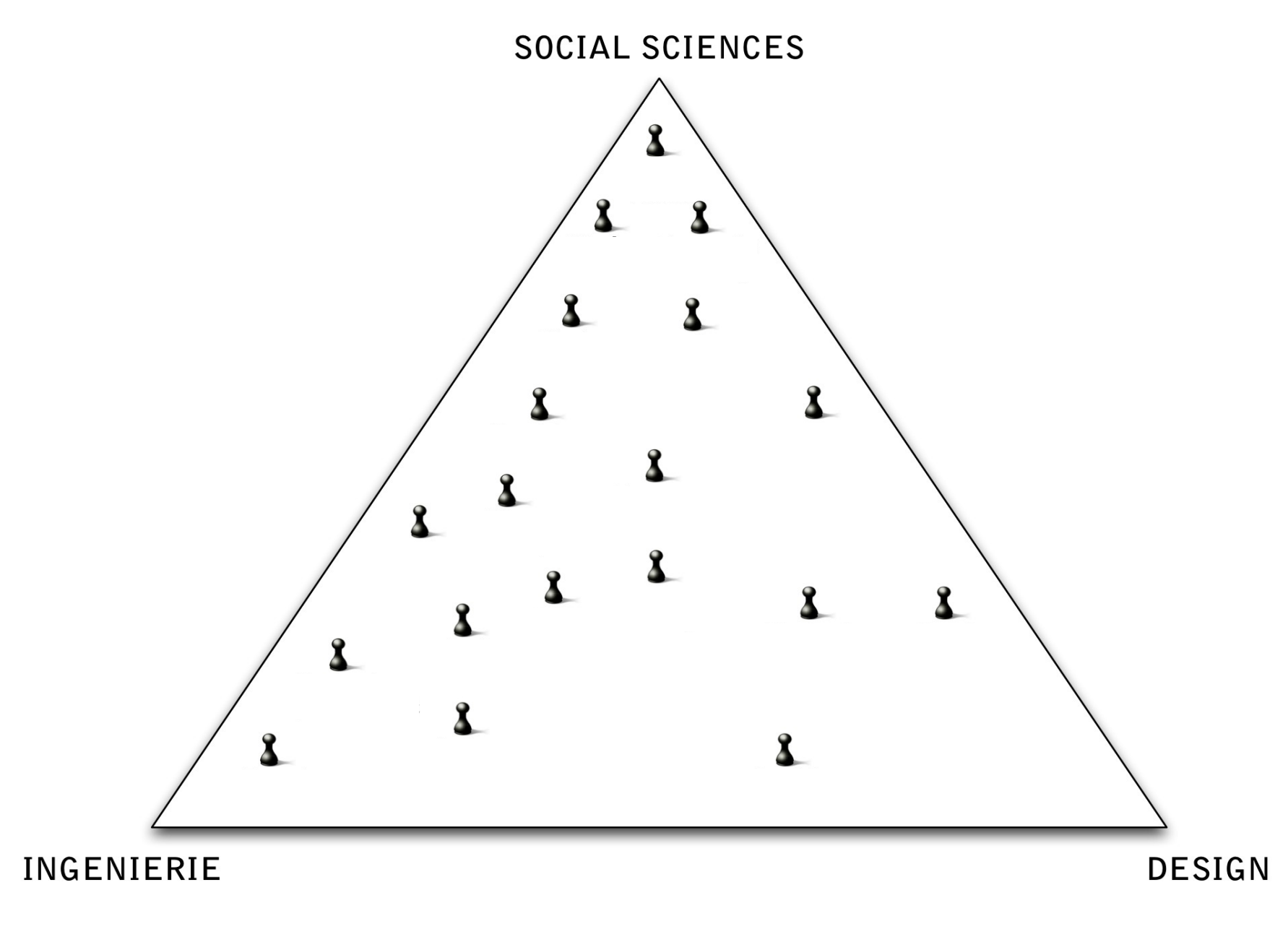
- Travailler avec les **traces numériques**

- Un écosystème d'**outils OpenSource**
  https://medialab.sciencespo.fr/outils/

- Un atelier ouvert mensuel : le METAT
  https://www.sciencespo.fr/recherche/fr/content/metat-latelier-de-methodes



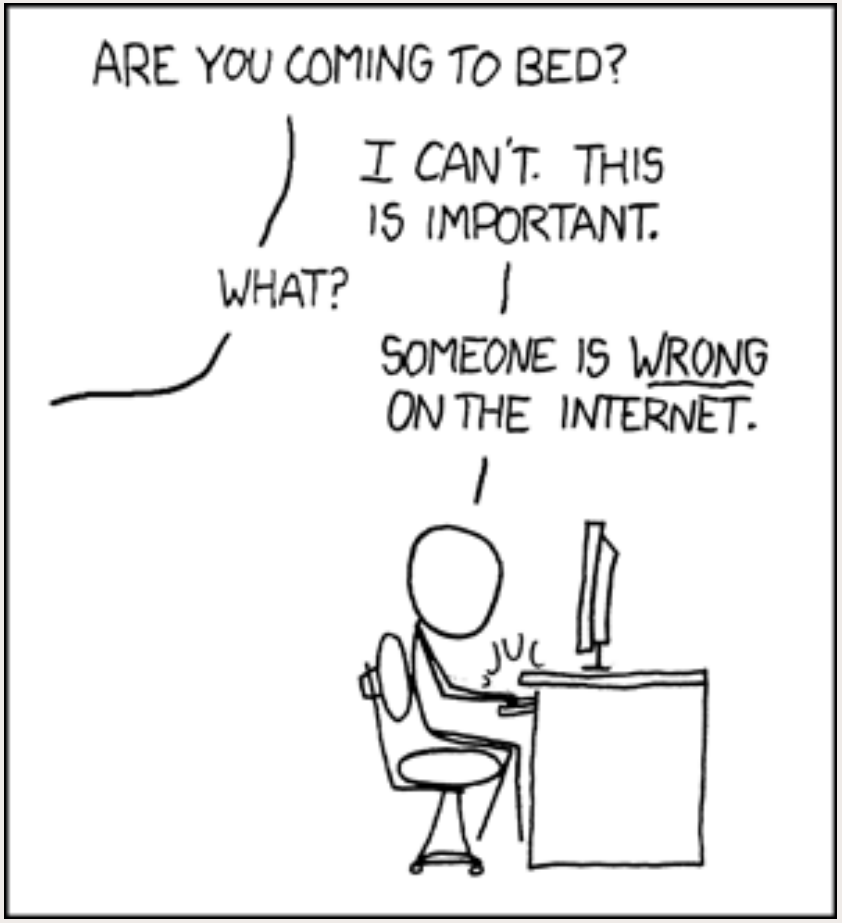
Une équipe interdisciplinaire



Exploiter le Web comme terrain d'enquêtes

Le Web : un espace de débats et de controverses

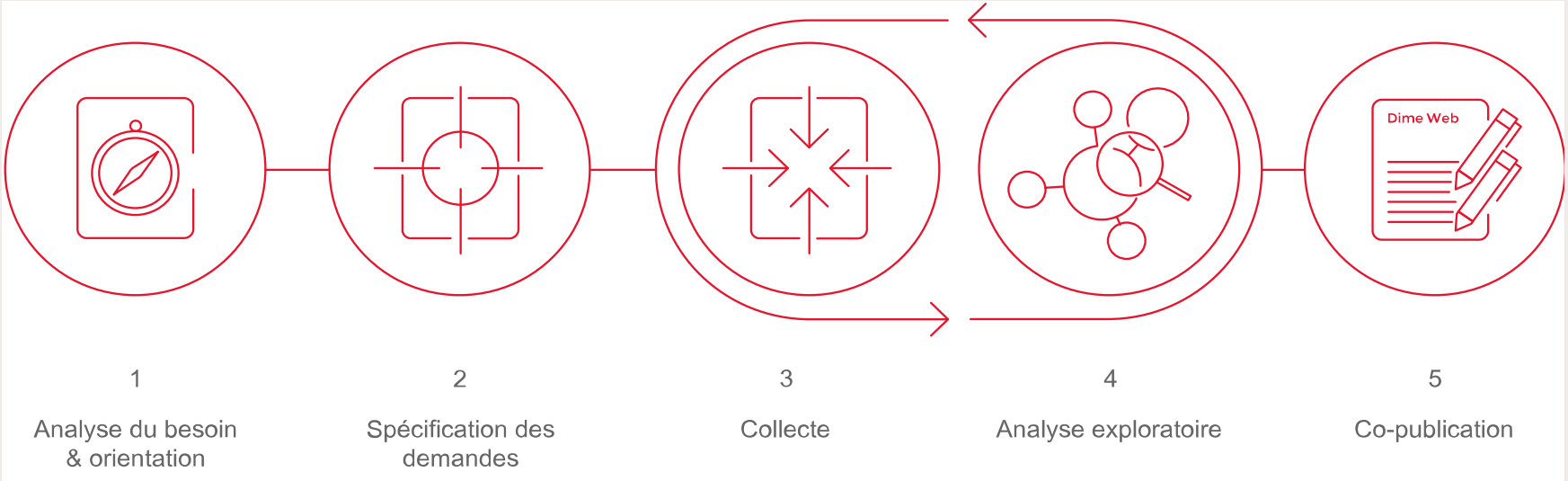
Collecter, enrichir, nettoyer, visualiser & analyser les traces numériques



Une méthodologie pour réaliser un terrain Web

Accompagnement technique et méthodologique à l'utilisation du Web comme nouveau terrain d'enquêtes

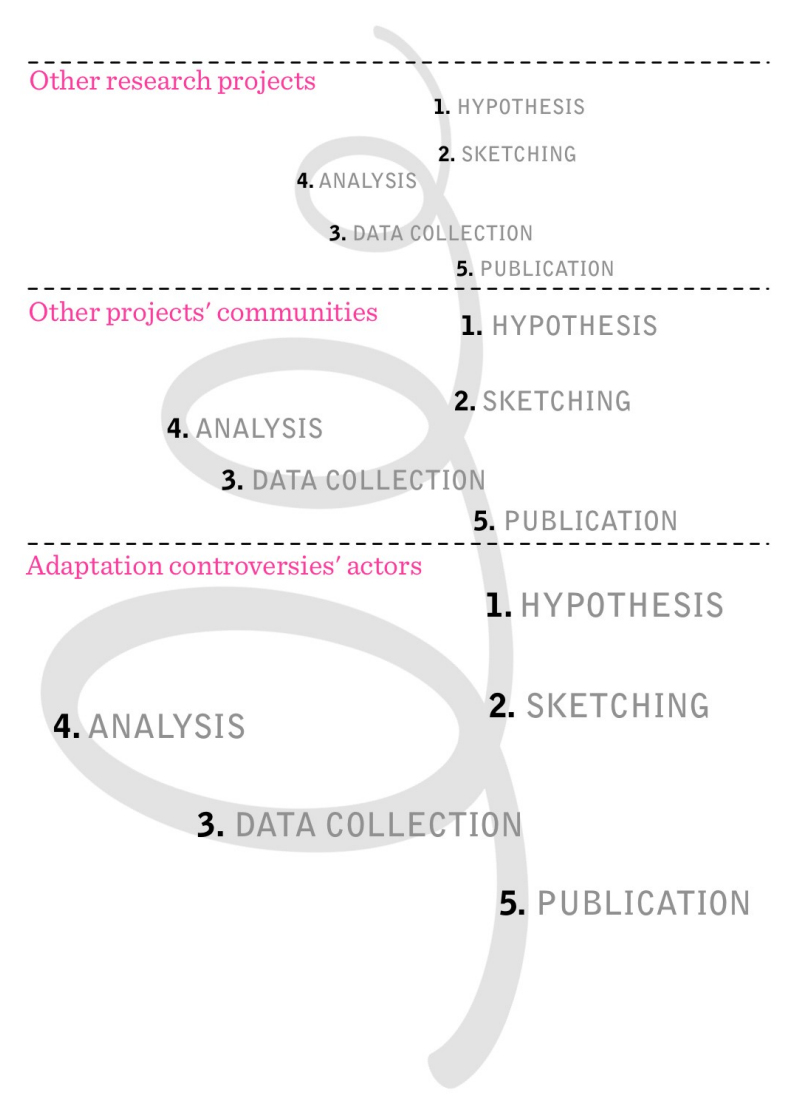
- Collecter, enrichir, nettoyer, visualiser et analyser des traces numériques
- Analyse de réseaux, archivage du web, analyse de controverses (ANT)
- Développement d'outils génériques
- Extraction ciblée de contenus
- Analyse Exploratoire de Données

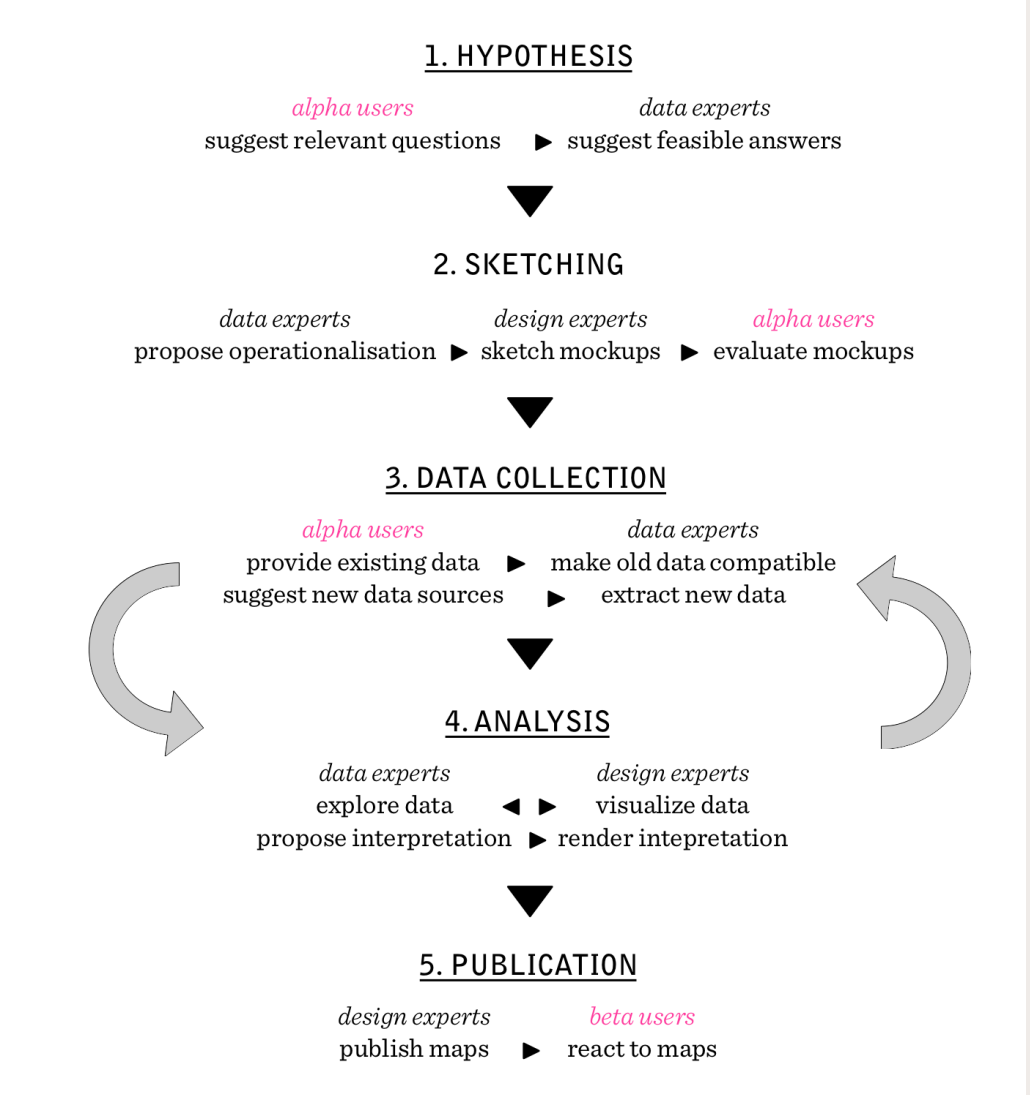


Une approche quali-quanti exploratoire itérative

Numérique ≠ Magique

→ toujours éviter l'entièrement automatique et garantir le retour aux sources





Développer un écosystème d'outils OpenSource

http://tools.medialab.sciences-po.fr

- Viser une large **Adoption :**
  - → **conception** d'outils dédiés aux besoins des chercheurs
  - → **design** d'interfaces centrées sur l'utilisateur
  - → **publication** d'outils web utilisables directement en ligne

- Assurer un maximum de **Réutilisabilité :**
  - → développement « **opportuniste** » de fonctionnalités
  - → diffusion en Logiciel Libre **Open Source**
    (téléchargeable, installable, vérifiable & modifiable)

- **Documentation académique** et pratique des outils & méthodes
  (publications scientifiques, tutoriels, formations...)

Organiser des « datasprints »

Ateliers collaboratifs de travail exploratoire et analytique

→ centrés sur un ou quelques jeux de données

→ sur plusieurs jours, en petits groupes

→ rassemblant une diversité de profils
(ingénieurs, designers, académiques, journalistes...)

→ confrontant les auteurs ou experts à leurs données

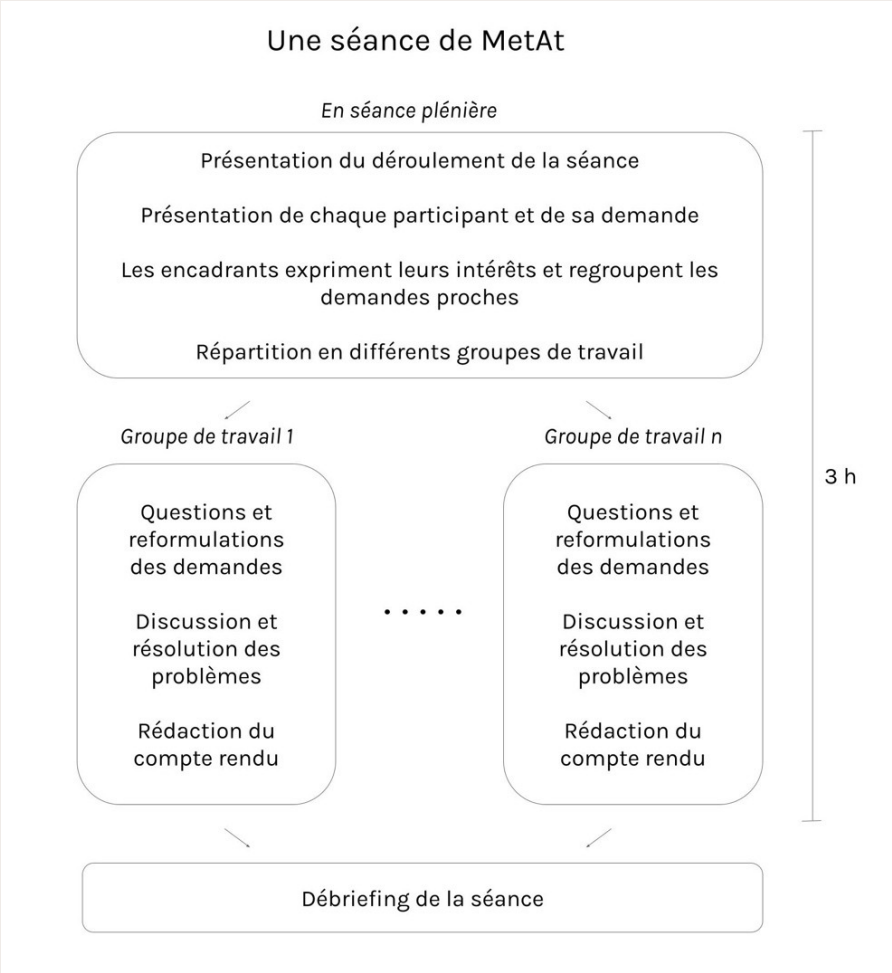
Le MetAt, atelier de méthodes numériques

https://www.sciencespo.fr/recherche/fr/content/metat-latelier-de-methodes

- Demandes d'accompagnement :

  - discussion et conseil méthodologique
  - collecte & nettoyage de données
  - visualisation exploratoire
  - formation aux outils
  - …

- Un mardi après-midi par mois
- Ouvert à tous, sur inscription préalable
- Initialement « atelier du médialab » :
  → canaliser les sollicitations
- Élargi à la communauté des ingénieurs de recherche de Sciences Po en 2017
- Contribue à l'autoformation continue



Diego Antolinos-Basso, Audrey Baneyx, Héloïse Théro, Benjamin Ooghe-Tabanou and Paul Girard, **"L'atelier de méthodes de Sciences Po : apprendre, aider, rassembler"**, Humanités numériques, 5 | 2022, Online since 01 June 2022, connection on 09 November 2022.
http://journals.openedition.org/revuehn/2799 DOI: https://doi.org/10.4000/revuehn.2799

SHS et « Big Data » ?

Big Data = trop gros pour être manipulé par un ordinateur (> To)
des données personnelles le plus souvent

→ extrêmement rare en Sciences Humaines

Différents modes de collecte de données web

2 approches bien distinctes aux cibles et résultats différents

CRAWLING **Vs.** **SCRAPING**

fouille systématique
(sources multiples hétérogènes)

contenus textuels & hyperliens

traitement
du langage

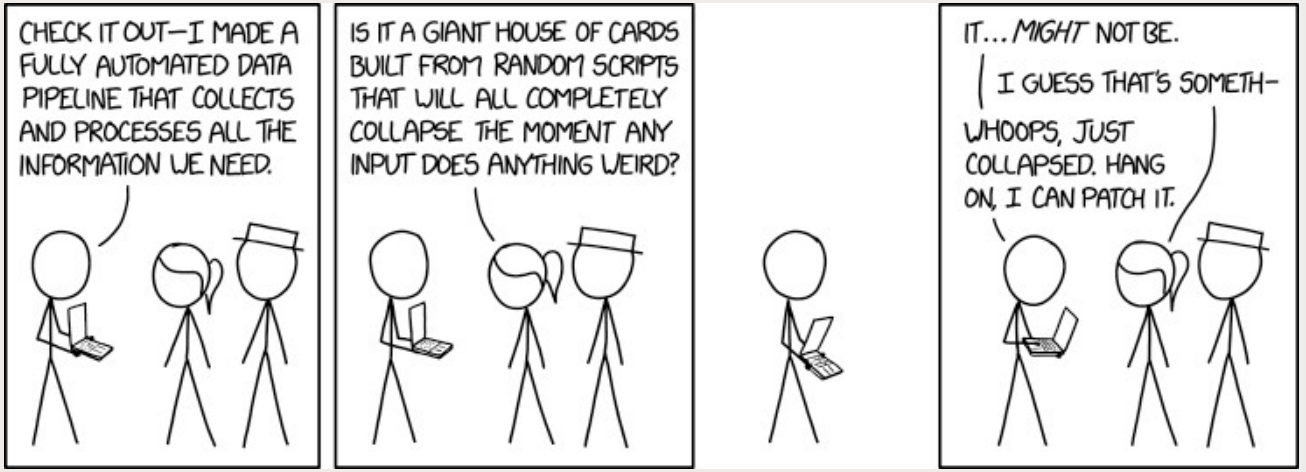
analyse de réseau
(effets de communauté)

extraction ciblée
(source unique ou ensemble cohérent)

données structurées

méthodes quantitatives,
statistiques…

Problèmes : des données « sales » et un important coût en maintenance

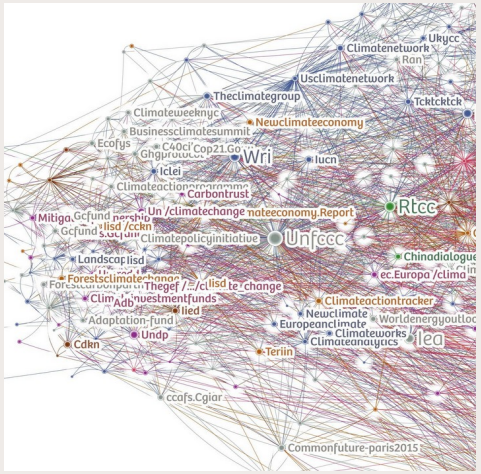


Hyphe : un crawler orienté recherche

http://hyphe.medialab.sciences-po.fr/demo/

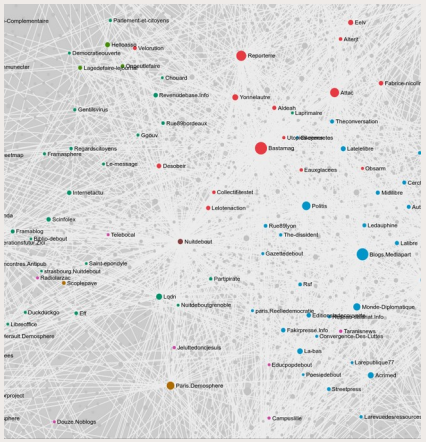
- Les liens hypertextes : nouveaux révélateurs de relations entre acteurs d'une thématique
- Créer un corpus documentaire
  - « acteurs web » & contenus textuels respectifs
  - liens hypertextes entre ces acteurs
- Études exploratoires ou de controverses dans tous les domaines

hyphe



http://medialab.github.io/double-dating-data/

COP 21
Vie privée
Extrême droite
Tissu associatif
Produits laitiers
Cellules souches
Administrations culturelles
…

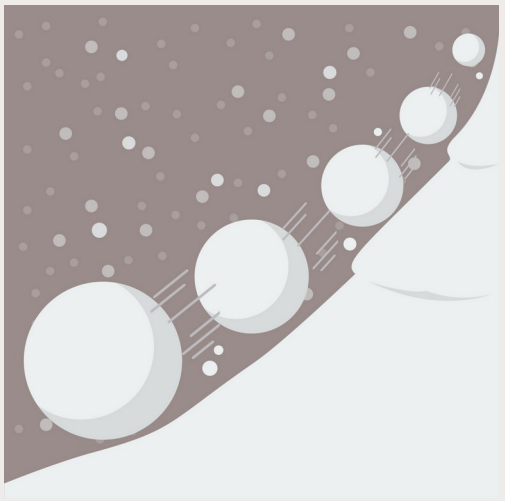


http://utopies-concretes.org/

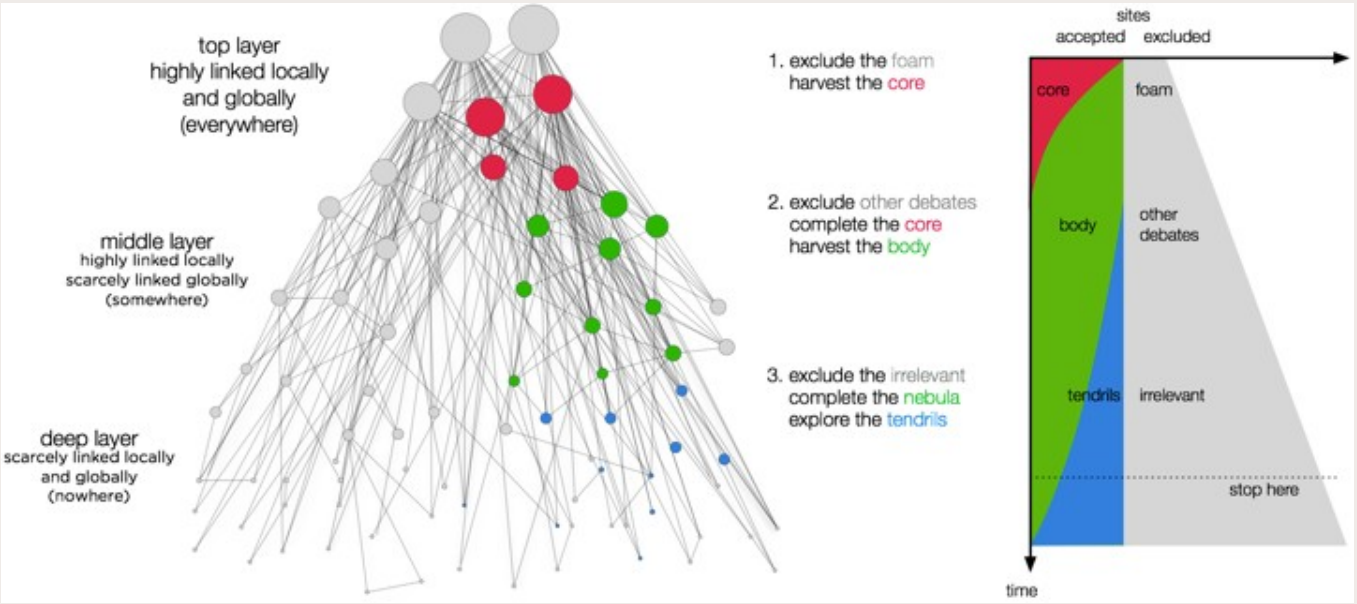
OOGHE-TABANOU, Benjamin, JACOMY, Mathieu, GIRARD, Paul & PLIQUE, Guillaume, "**Hyperlink is not dead!**", In Proceedings of the 2nd International Conference on Web Studies (WS.2 2018). ACM, New York, NY, USA, 12-18. DOI: https://doi.org/10.1145/3240431.3240434

Hyphe : une stratégie de crawling contrôlé

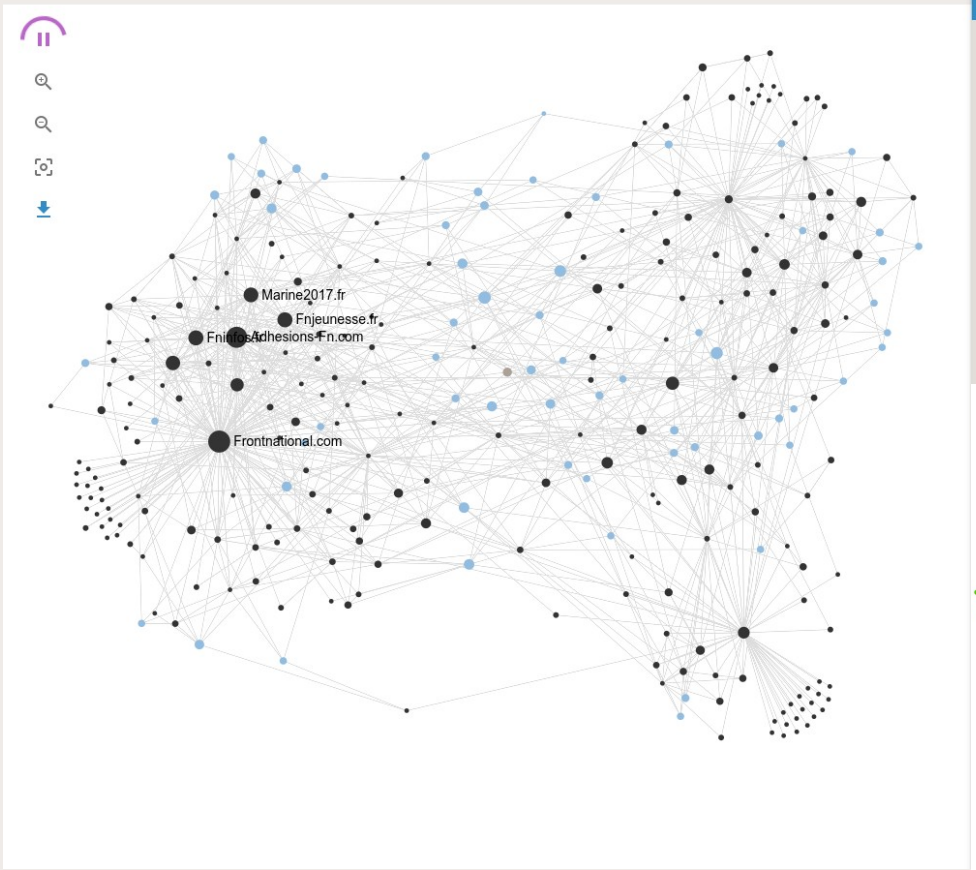
- Crawlers classiques : snowballing
  → Surreprésentation des couches hautes (Google, YouTube, Wikipedia…)
  → Dérive thématique rapide

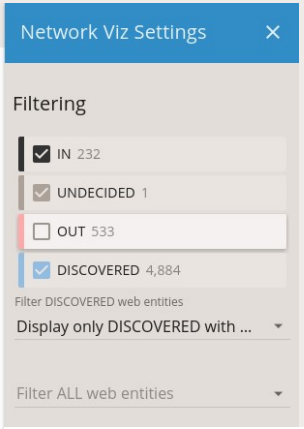


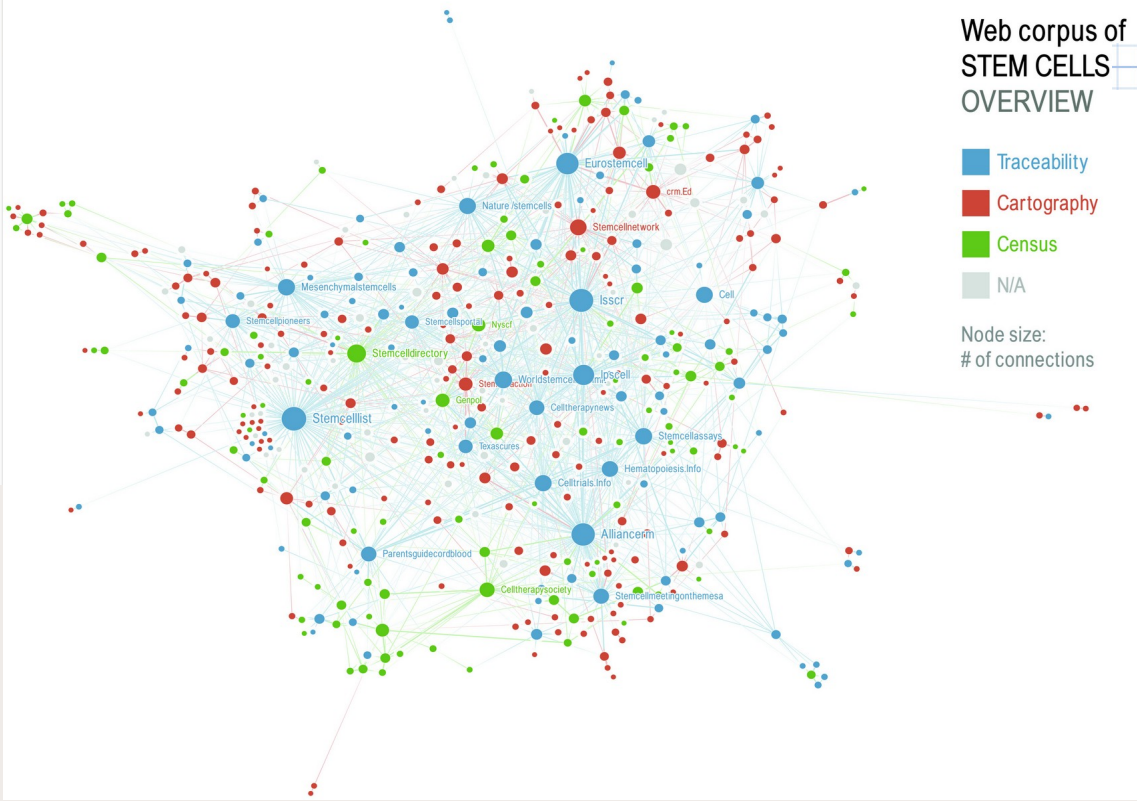
- Hyphe : crawling semi-automatique
  - Fouille systématique des pages des WebEntités choisies uniquement
  - Choix humain des autres WebEntités à crawler grâce au degré de citation



Analyse de réseau : clusters, oppositions & affinités







Social Representations of Stem Cells, Virginie Tournay, CEVIPOF, 2016

SciencesPo
MÉDIALAB

Cartographier le web autour d'un thème/débat

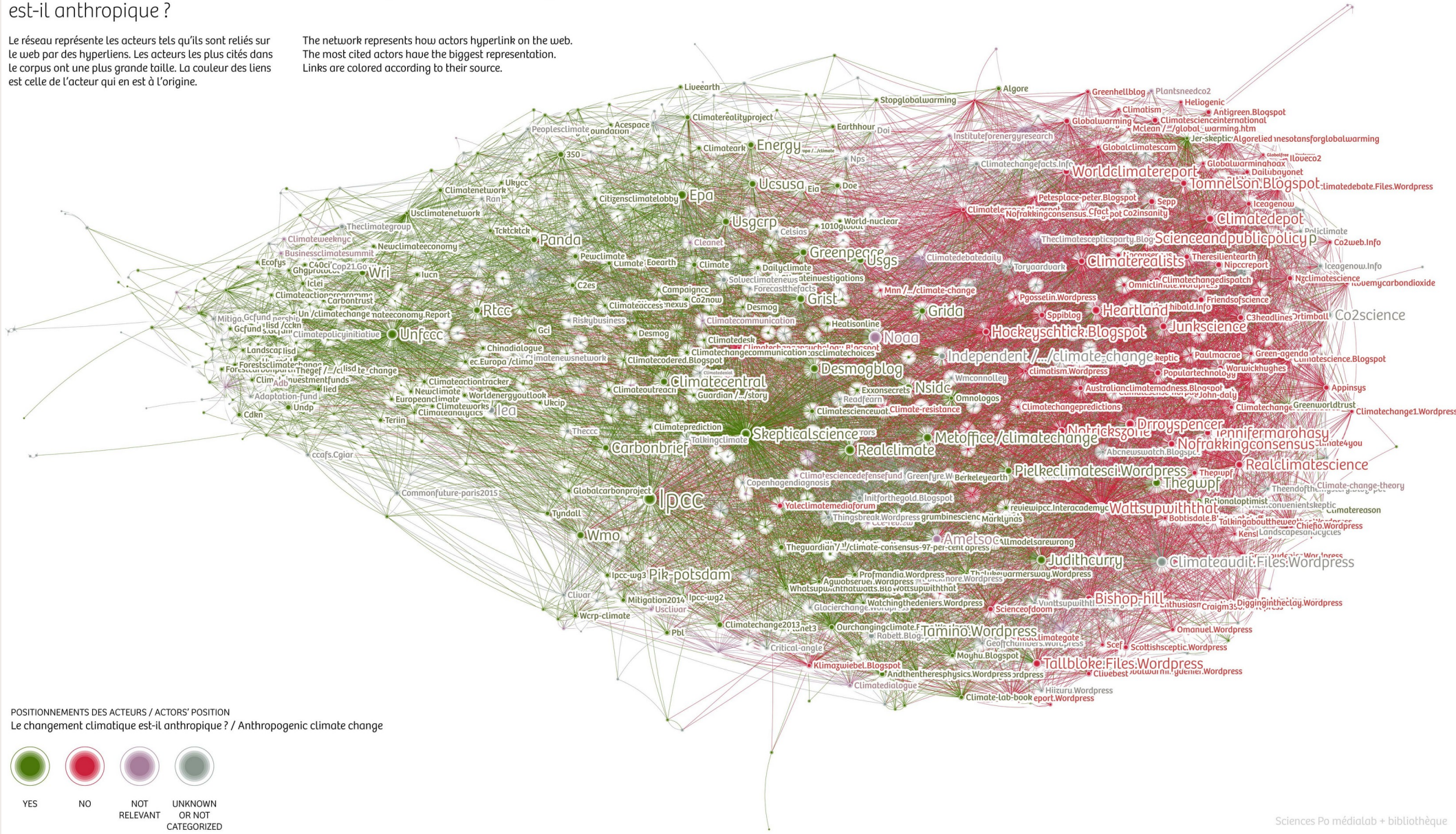
Corpus web sur le changement climatique : le changement climatique est-il anthropique ?

Le réseau représente les acteurs tels qu'ils sont reliés sur le web par des hyperliens. Les acteurs les plus cités dans le corpus ont une plus grande taille. La couleur des liens est celle de l'acteur qui en est à l'origine.

Web corpus on climate change: anthropogenic climate change

The network represents how actors hyperlink on the web. The most cited actors have the biggest representation. Links are colored according to their source.

https://medialab.github.io/double-dating-data/#/



POSITIONNEMENTS DES ACTEURS / ACTORS' POSITION
Le changement climatique est-il anthropique ? / Anthropogenic climate change

YES NO NOT RELEVANT UNKNOWN OR NOT CATEGORIZED

SciencesPo
MÉDIALAB

Cartographier le web par typologie d'acteurs

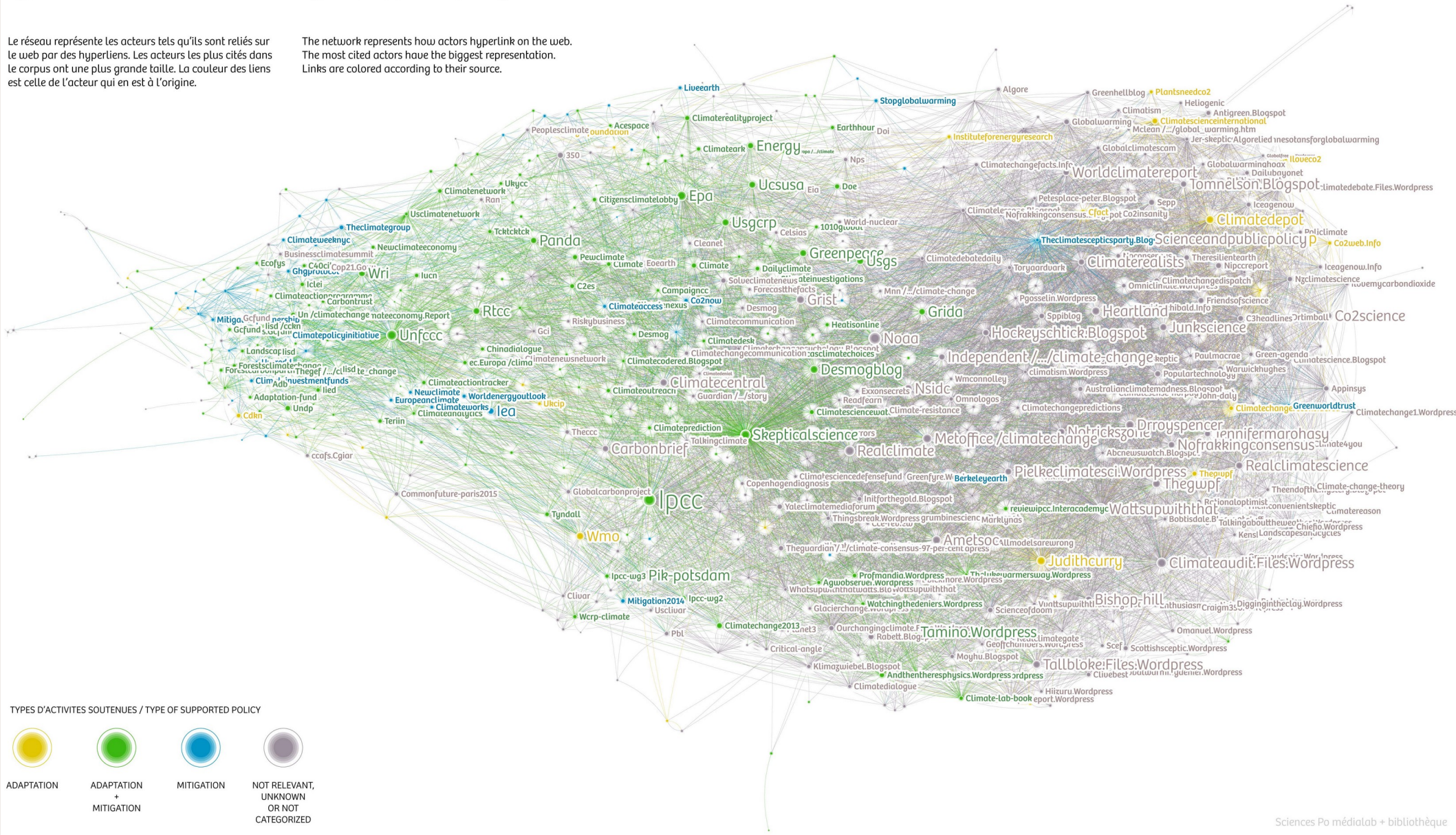
Corpus web sur le changement climatique : types d'activités soutenues

Web corpus on climate change: type of supported policy

https://medialab.github.io/double-dating-data/#/

Le réseau représente les acteurs tels qu'ils sont reliés sur le web par des hyperliens. Les acteurs les plus cités dans le corpus ont une plus grande taille. La couleur des liens est celle de l'acteur qui en est à l'origine.

The network represents how actors hyperlink on the web. The most cited actors have the biggest representation. Links are colored according to their source.



TYPES D'ACTIVITES SOUTENUES / TYPE OF SUPPORTED POLICY

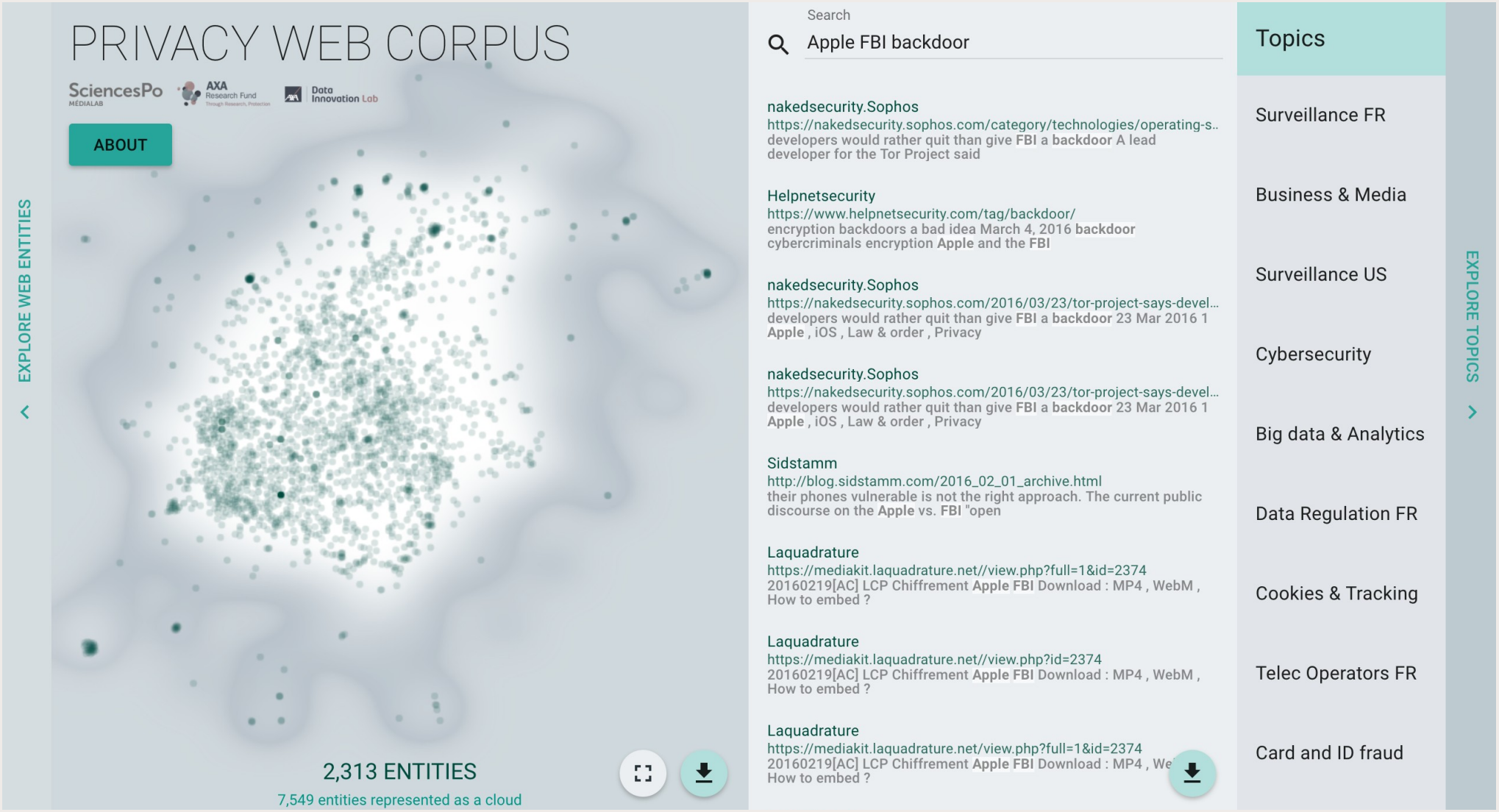
ADAPTATION

ADAPTATION + MITIGATION

MITIGATION

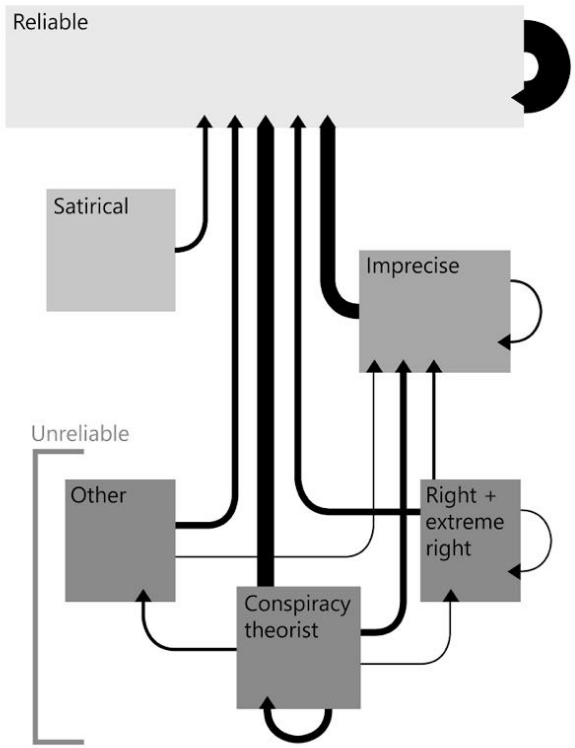
NOT RELEVANT, UNKNOWN OR NOT CATEGORIZED

Analyse de fond : traiter les contenus texte



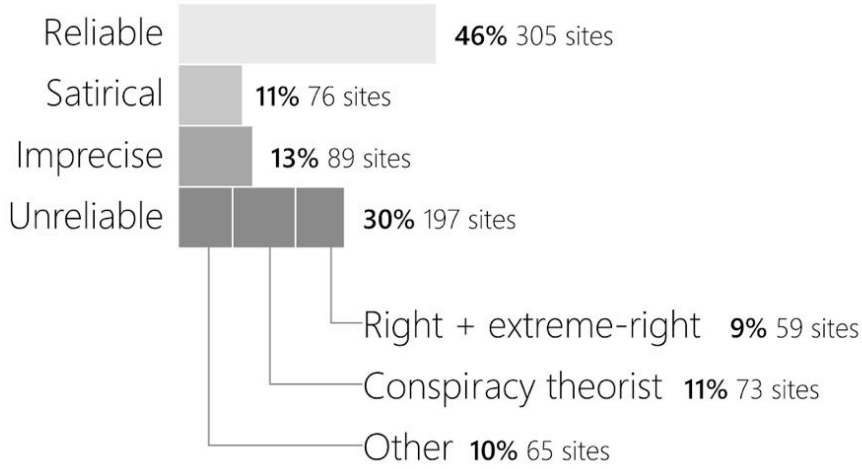
http://tools.medialab.sciences-po.fr/privacy/

Exploiter la directionnalité des hyperliens

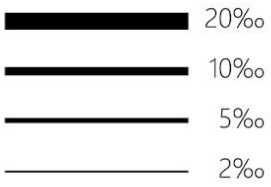


Most hyperlinks stem from the unreliable and aim at the reliable resources

Each bloc's surface is proportional to the count of websites. The color code si the same as the "Décodex".



The thickness si proportional to the normalized link density*



*Accounts for the different group sizes.

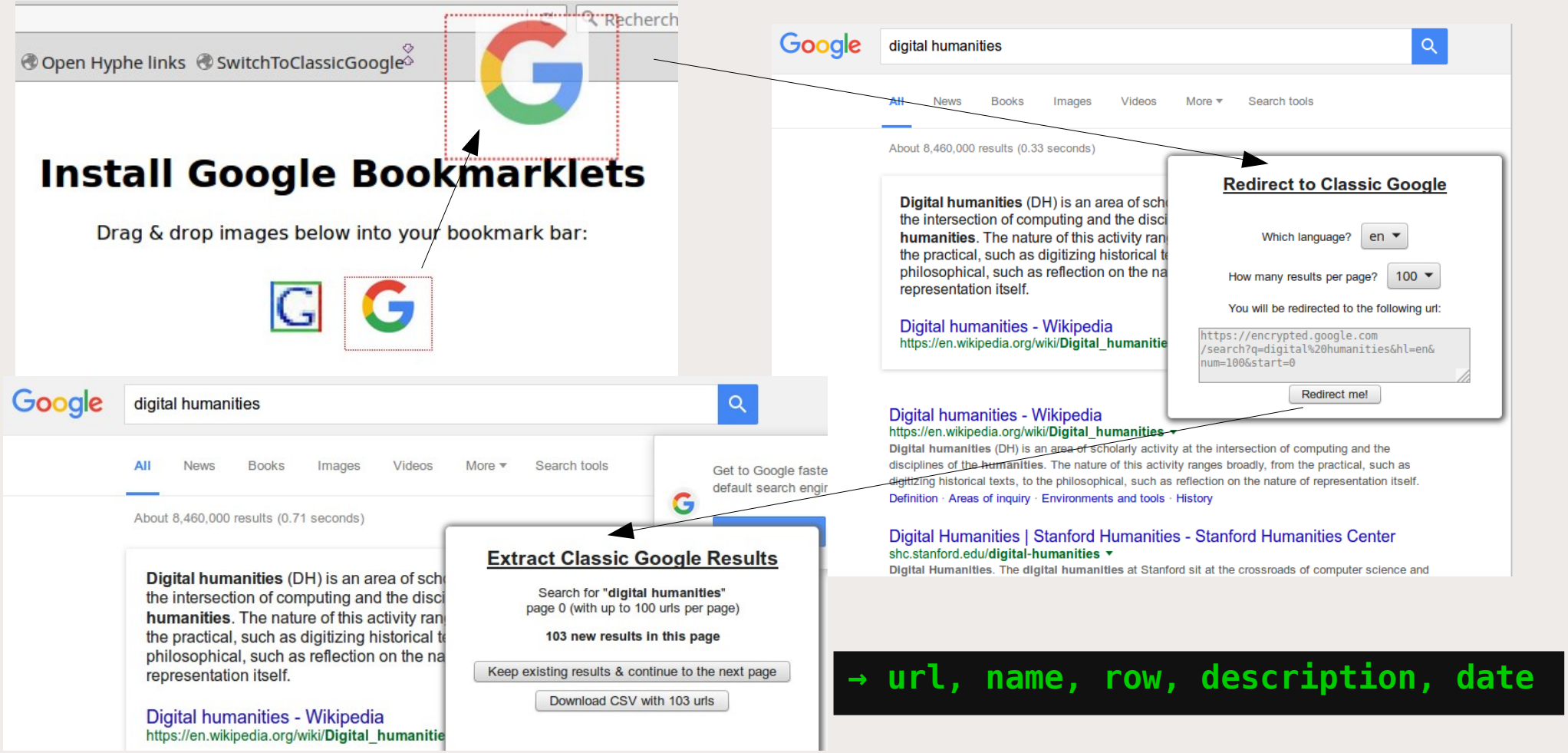
Densities below 2‰ have been omitted.

Venturini, Tommaso & Jacomy, Mathieu & Bounegru, Liliana & Gray, Jonathan. (2018). **Visual Network Exploration for Data Journalists.**
https://www.researchgate.net/publication/320225750_Visual_Network_Exploration_for_Data_Journalists

Exemple de scraping ciblé : Google Bookmarklets

https://medialab.github.io/google-bookmarklets/

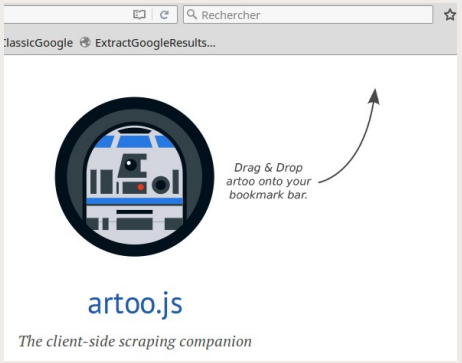
Des petits boutons installables simplement dans les favoris du navigateur pour exporter simplement en tableur des résultats d'une recherche Google



→ url, name, row, description, date

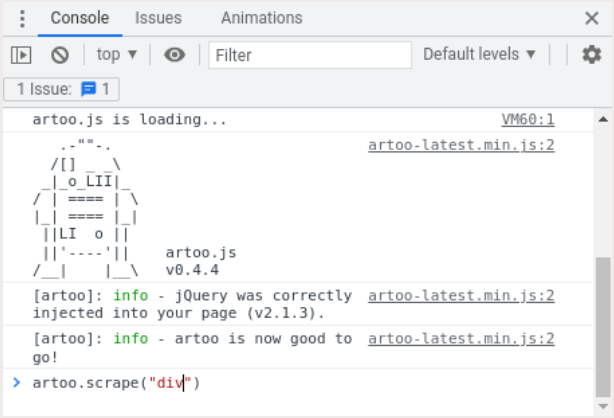
artoo.js : extraire des données du web (avancé)

https://medialab.github.io/artoo/



- Un bookmarklet à ajouter dans la barre de favoris du navigateur

- Une librairie JavaScript de fonctions utiles pour le scraping depuis la console du navigateur (F12)

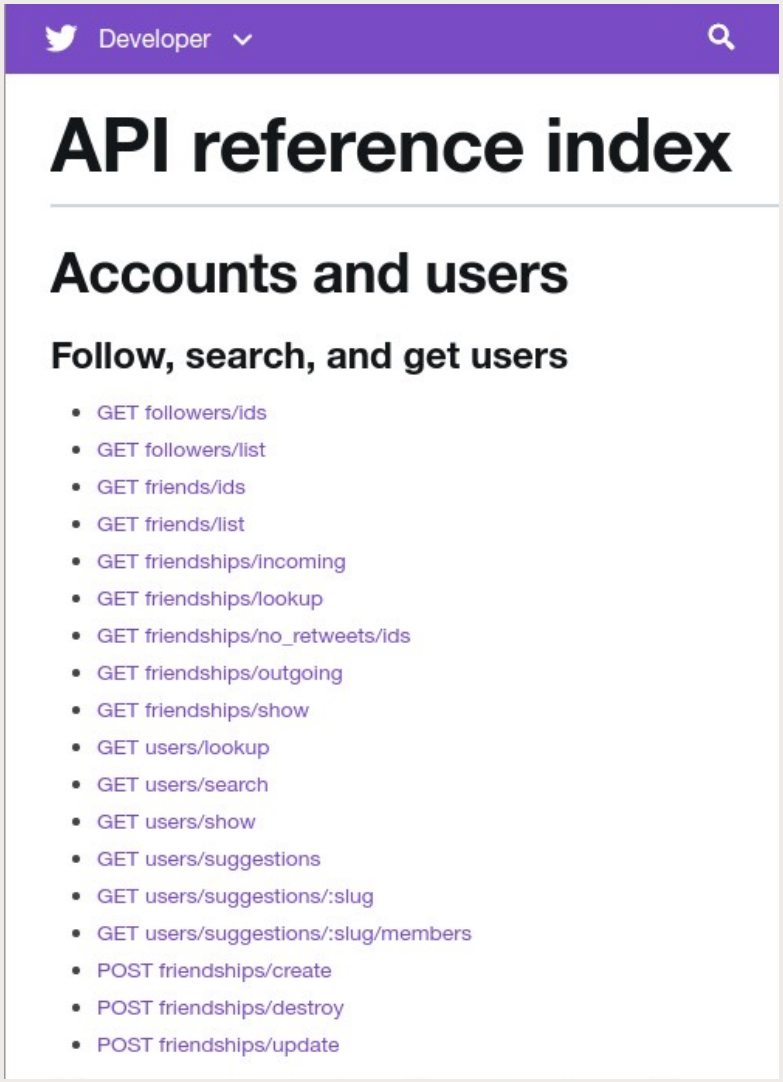


→ exemple : https://en.wikipedia.org/wiki/List_of_countries_by_carbon_dioxide_emissions

```
> var data = artoo.scrapeTable( ".wikitable", {headers: 'th'} );
 undefined
> data.length;
 49
> data[0];
 Object {Country: " World", CO2 emissions (kt) in 2014[2]: "35,669,000", " % CO2
Emissions by Country": "100%", Emission per capita (t) in 2014[3]: "5.0"}
> artoo.saveCsv(data, "CO2-world-emissions.csv");
 undefined
```

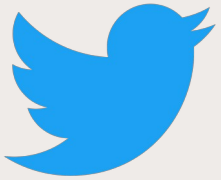
Accès contrôlé via les APIs des plateformes

- **« Application Programmation Interface »**
- **Avantages :**
  - données structurées « propres »
  - accès à de gros volume
  - relative complétude
  - accès à des informations d'usage
- **Problèmes :**
  - accès à de gros volumes
  - limitation des appels
  - « boîtes noires »
  - risques de refermeture
  - dépendance à la vision des plateformes



https://developer.twitter.com/en/docs/api-reference-index

De nombreuses métadonnées à exploiter




```
TWEET_FIELDS = [
  "id",                          # digital ID
  "time",                        # UNIX timestamp of creation
  "created_at",                  # ISO datetime of creation
  "from_user_name",              # author's user text ID (@user)
  "text",                        # message's text content
  "filter_level",                # internal TCAT field, ignorable
  "possibly_sensitive",          # whether a link present in the message might contain sensitive content according to Twitter
  "withheld_copyright",          # whether the tweet might be censored by Twitter following copyright requests, ignorable
  "withheld_scope",              # whether the content withheld is the "status" or a "user", ignorable
  "withheld_countries",          # list of ISO country codes in which the message is withheld, separated by |, ignorable
  "truncated",                   # whether the tweet is bigger than 140 characters, obsolete
  "retweet_count",               # number of retweets of the message (at collection time)
  "favorite_count",              # number of likes of the message (at collection time)
  "reply_count",                 # number of answers to the message, dropped by Twitter (since Oct 17, now charged), unreliable and ignorable
  "lang",                        # language of the message automatically identified by Twitter's algorithms (equals "und" when no language could be detected)
  "to_user_name",                # text ID of the user the message is answering to
  "to_user_id",                  # digital ID of the user the message is answering to
  "in_reply_to_status_id",       # digital ID of the tweet the message is answering to
  "source",                      # medium used by the user to post the message
  "source_name",                 # name of the medium used to post the message
  "source_url",                  # link to the medium used to post the message
  "location",                    # location declared in the user's profile (at collection time)
  "lat",                         # latitude of messages geolocalized
  "lng",                         # longitude of messages geolocalized
  "from_user_id",                # author's user digital ID
  "from_user_realname",          # author's detailed textual name (at collection time)
  "from_user_verified",          # whether the author's account is certified
  "from_user_description",       # description given in the author's profile (at collection time)
  "from_user_url",               # link to a website given in the author's profile (at collection time)
  "from_user_profile_image_url", # link to the image avatar of the author's profile (at collection time)
  "from_user_utcoffset",         # time offset due to the user's timezone, dropped by Twitter (since May 18), ignorable
  "from_user_timezone",          # timezone declared in the user's profile, dropped by Twitter (since May 18), ignorable
  "from_user_lang",              # language declared in the user's profile (at collection time)
  "from_user_tweetcount",        # number of tweets sent by the user (at collection time)
  "from_user_followercount",     # number of users following the author (at collection time)
  "from_user_friendcount",       # number of users the author is following (at collection time)
  "from_user_favourites_count",  # number of likes the author has expressed (at collection time)
  "from_user_listed",            # number of users lists the author has been included in (at collection time)
  "from_user_withheld_scope",    # whether the user content is withheld, ignorable
  "from_user_withheld_countries", # list of ISO country codes in which the user content is withheld, separated by |, ignorable
  "from_user_created_at",        # ISO datetime of creation of the author's account
  "collected_via_thread",        # whether the tweet was retrieved only as part of a thread including a tweet matching the desired query
  "retweeted_id",                # digital ID of the retweeted message
  "retweeted_user_name",         # text ID of the user who authored the retweeted message
  "retweeted_user_id",           # digital ID of the user who authoring the retweeted message
  "quoted_id",                   # digital ID of the retweeted message
  "quoted_user_name",            # text ID of the user who authored the retweeted message
  "quoted_user_id",              # digital ID of the user who authoring the retweeted message
  "links",                       # list of links included in the text content, with redirections resolved, separated by |
  "medias_urls",                 # list of links to images/videos embedded, separated by |
  "medias_files",                # list of filenames of images/videos embedded and downloaded, separated by |, ignorable when medias collections isn't enabled
  "mentioned_user_names",        # list of text IDs of users mentionned, separated by |
  "mentioned_user_ids",          # list of digital IDs of users mentionned, separated by |
  "hashtags"                     # list of hashtags used, lowercased, separated by |
]
```

Risques de refermeture des APIs

- Twitter API v2.0 « *designed for academics needs* »
  … or not…

	scraping web	search API v1.1	search API v2 standard	search API v2 académique
time coverture	all time	8 days back	8 days back	all time
# tweets limit	~ 50 M / month	~ 180 M / month	0.2 M / month	10 M / month
retweets included	no	yes	yes	yes
extra metadata	yes	no	yes	yes

task	API version	results / query	app API / 15 min	user API / 15 min	queries / heure	results / heure	max queries / jour	queries / jour	max tweets / mois	results / jour
user followers/friends	1.1	5000	15	15	120	600000		2880		14400000
users followers/friends (max si 10000 / user)	1.1	0,5	15	15	120	60		2880		1440
users metas	1.1	100	300	900	4800	480000		115200		11520000
tweets by ids	1.1	100	300	900	4800	480000		115200		11520000
tweets by ids	2.0	100	300	900	4800	480000		115200	500000	500000
user tweets	1.1	200	1500	900	9600	1920000	100000	100000		20000000
users tweets (max si 3250 tweets / use	1.1	0,05882352941	1500	900	9600	564	100000	100000		5882
tweet retweets	1.1	100	300	75	1500	150000		36000		3600000
tweets by search	1.1	100	450	180	2520	252000		60480		6048000
tweets by search	2.0	100	450	180	2520	252000		60480	500000	500000
tweets by search	2.0 académique	500	300	0	1200	600000	86400	28800	10000000	10000000

Creuser le web avec Minet

https://github.com/medialab/minet

- Pour utilisateurs semi-avancés
  (no-code, mais ligne de commande)
- Généricizer nos pratiques et expertises de webmining
- Autonomiser chercheurs et doctorants sur la collecte de données
- Extraire des contenus textes, des liens, des images, etc.
  à partir de listes d'URLs ou mots-clés
- <u>SCRAPING :</u>
  - Facebook posts, pages & groups
  - Twitter search
  - Instagram en cours d'ajout `(v0.65.0!)`
  - Telegram en cours d'ajout
  - TikTok un jour ?
- <u>APIs :</u>
  - Crowdtangle (métriques Facebook)
  - Twitter friends/followers/search
  - YouTube videos & channels
  - Reddit à venir

```
# Yomgui at mbp-de-plique-1.home in ~/code/minet on git:master ✖ [15:51:23]
→ ./ftest/fetch.sh
```

Gazouilloire : extraction systématique de tweets

https://github.com/medialab/gazouilloire

- Collecter en direct en continu (et jusqu'à 7 jours en arrière)
  - des tweets par mots-clés, urls, utilisateurs, localisation, langue…
  - les conversations et médias associés
  - des profils d'utilisateurs

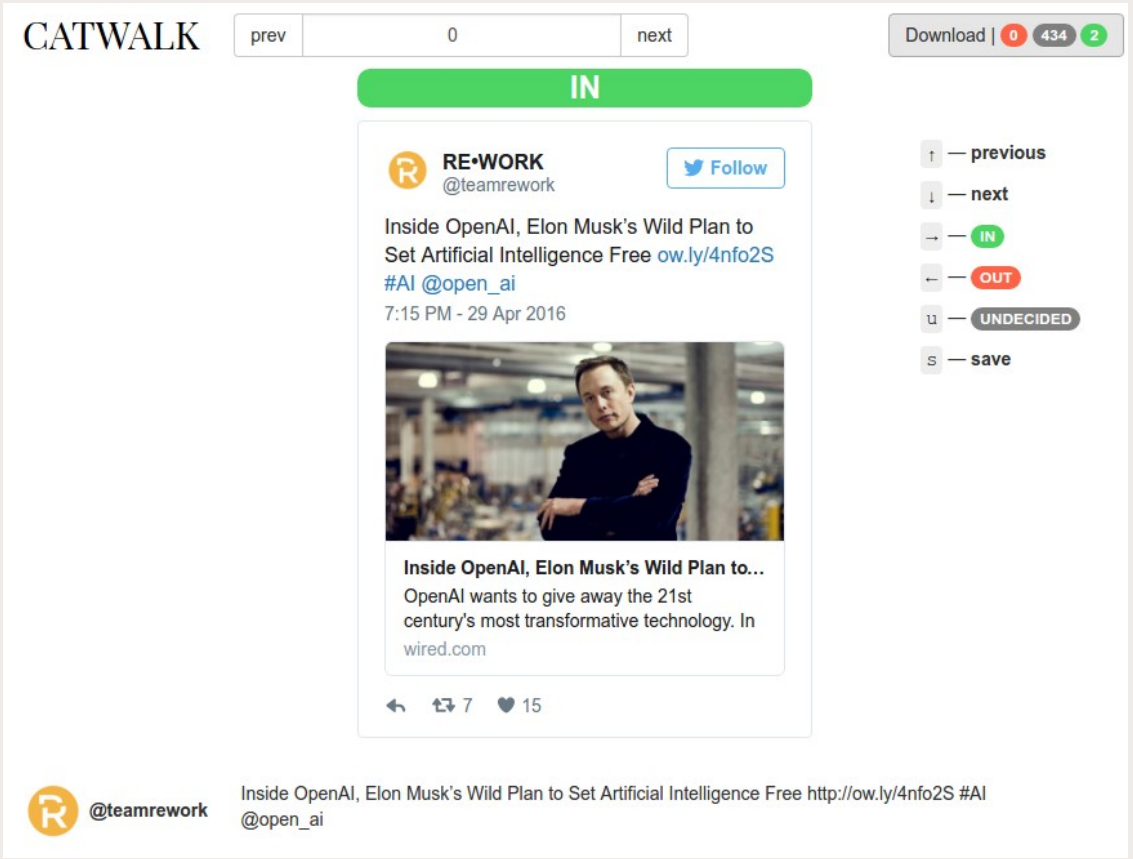
```
{
    "twitter": {
        "user": "Gazou_medialab2",
        "key": "[redacted]",
        "secret": "[redacted]",
        "oauth_token": "[redacted]",
        "oauth_secret": "[redacted]"
    },
    "mongo": {
        "host": "localhost",
        "port": 27017,
        "db": "tweets-naturpradi"
    },
    "keywords": [
        "écologique Paris",
        "végétation Paris",
        "verger Paris",
        "grenelle environnement Paris",
        "locavore Paris"
    ],
    "time_limited_keywords": {
    },
    "geolocalisation": null,
    "geolocalisation_type": "admin",
    "resolve_redirected_links": true,
    "grab_conversations": true,
    "download_medias": true,
    "medias_directory": "/store/tweets/naturpradi/media/",
    "timezone": "Europe/Paris",
    "debug": true
}
```

```
[2016-11-22 15:23:34.056196] DEBUG: Starting search queries with 328 remaining calls for the next 655 seconds
[2016-11-22 15:23:34.259849] DEBUG: [search] +1 tweets (agriculture%20Paris OR agricultures%20Paris OR agroforesterie%20Paris)
[2016-11-22 15:23:35.807085] DEBUG: Saved 1 tweets in MongoDB
[2016-11-22 15:23:37.358533] DEBUG: [search] +1 tweets (espaces%20verts%20Paris OR ferme%20Paris OR fermes%20Paris)
[2016-11-22 15:23:37.810930] DEBUG: Saved 1 tweets in MongoDB
[2016-11-22 15:23:45.049743] DEBUG: [stream] +1 tweet
[2016-11-22 15:23:45.821150] DEBUG: Saved 1 tweets in MongoDB
[2016-11-22 15:24:51.598045] DEBUG: [stream] +1 tweet
[2016-11-22 15:24:51.893009] DEBUG: Saved 1 tweets in MongoDB
[2016-11-22 15:24:52.401661] DEBUG: [medias] +1 files
[2016-11-22 15:24:58.073013] DEBUG: Starting search queries with 286 remaining calls for the next 571 seconds
[2016-11-22 15:25:00.383614] DEBUG: [stream] +1 tweet
[2016-11-22 15:25:01.905385] DEBUG: Saved 1 tweets in MongoDB
[2016-11-22 15:26:18.060840] DEBUG: Starting search queries with 246 remaining calls for the next 491 seconds
[2016-11-22 15:26:19.922864] DEBUG: [search] +1 tweets (compost%20Paris OR composts%20Paris OR compostage%20Paris)
[2016-11-22 15:26:19.989779] DEBUG: Saved 1 tweets in MongoDB
```

CatWalk : sélection qualitative de tweets

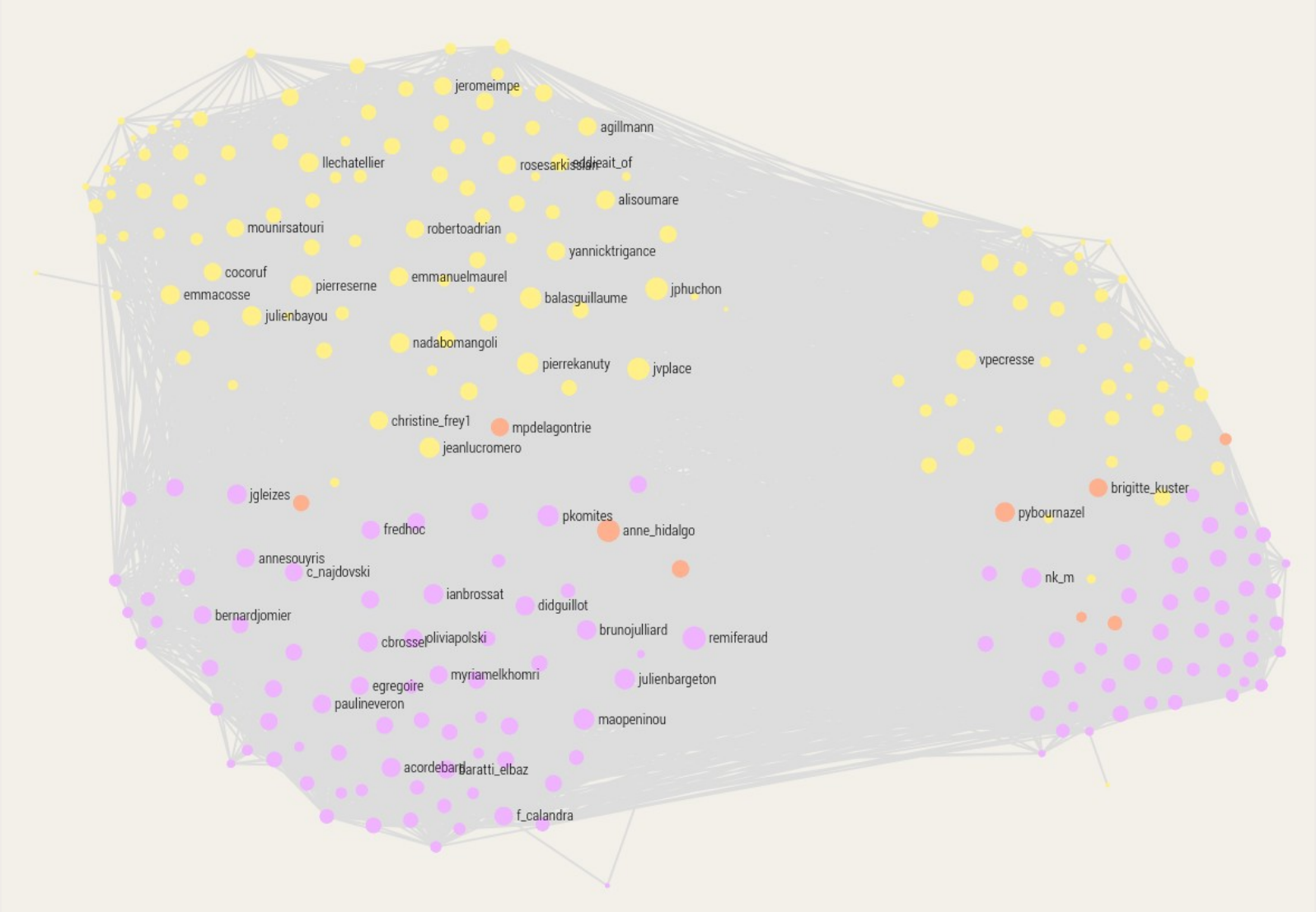
https://medialab.github.io/catwalk/

Passer en revue rapidement « *à la Tinder* » tous les tweets d'un CSV pour décider de les inclure / exclure d'un corpus



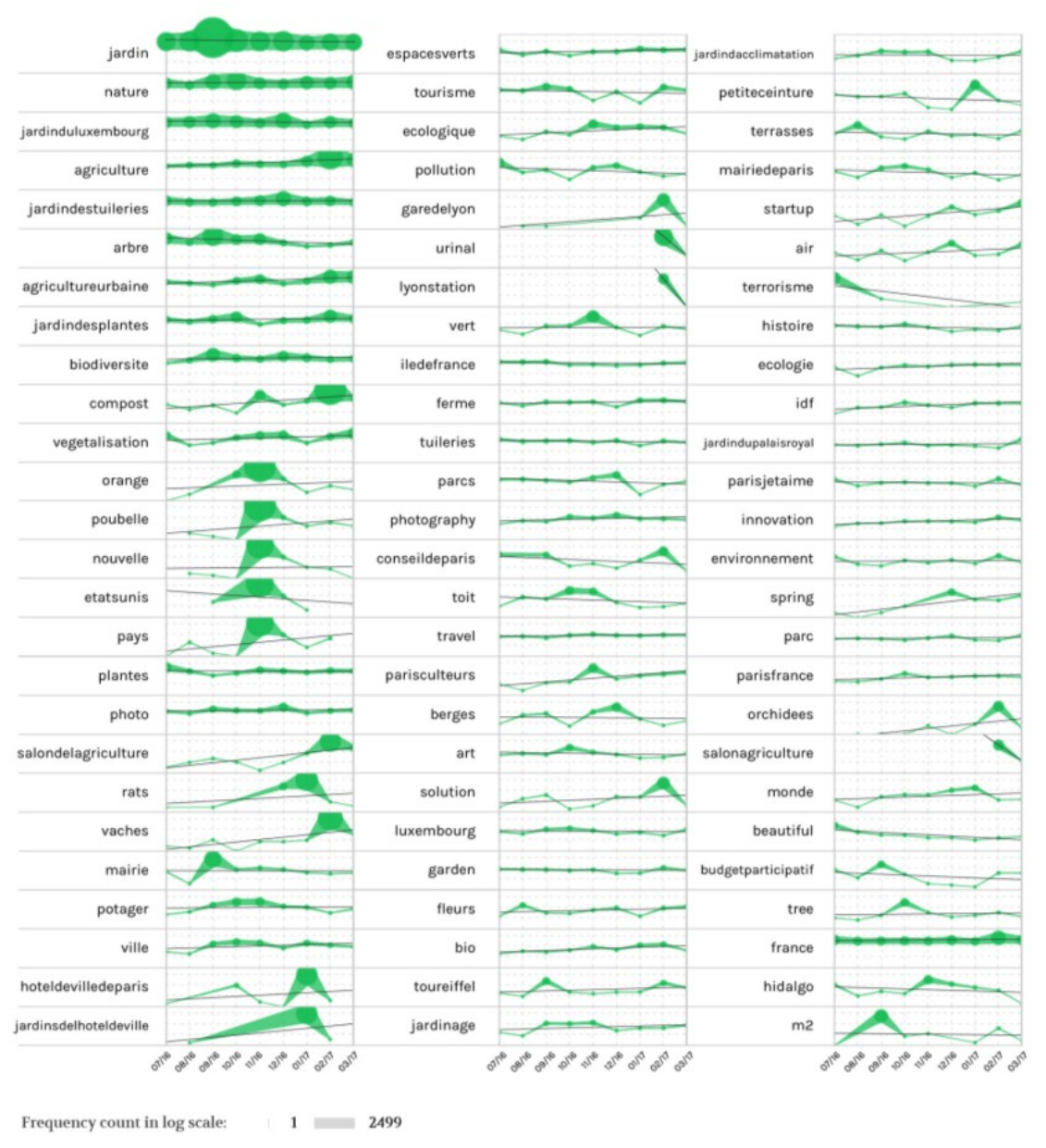
→ V2 en approche !

Explorer des réseaux de followers



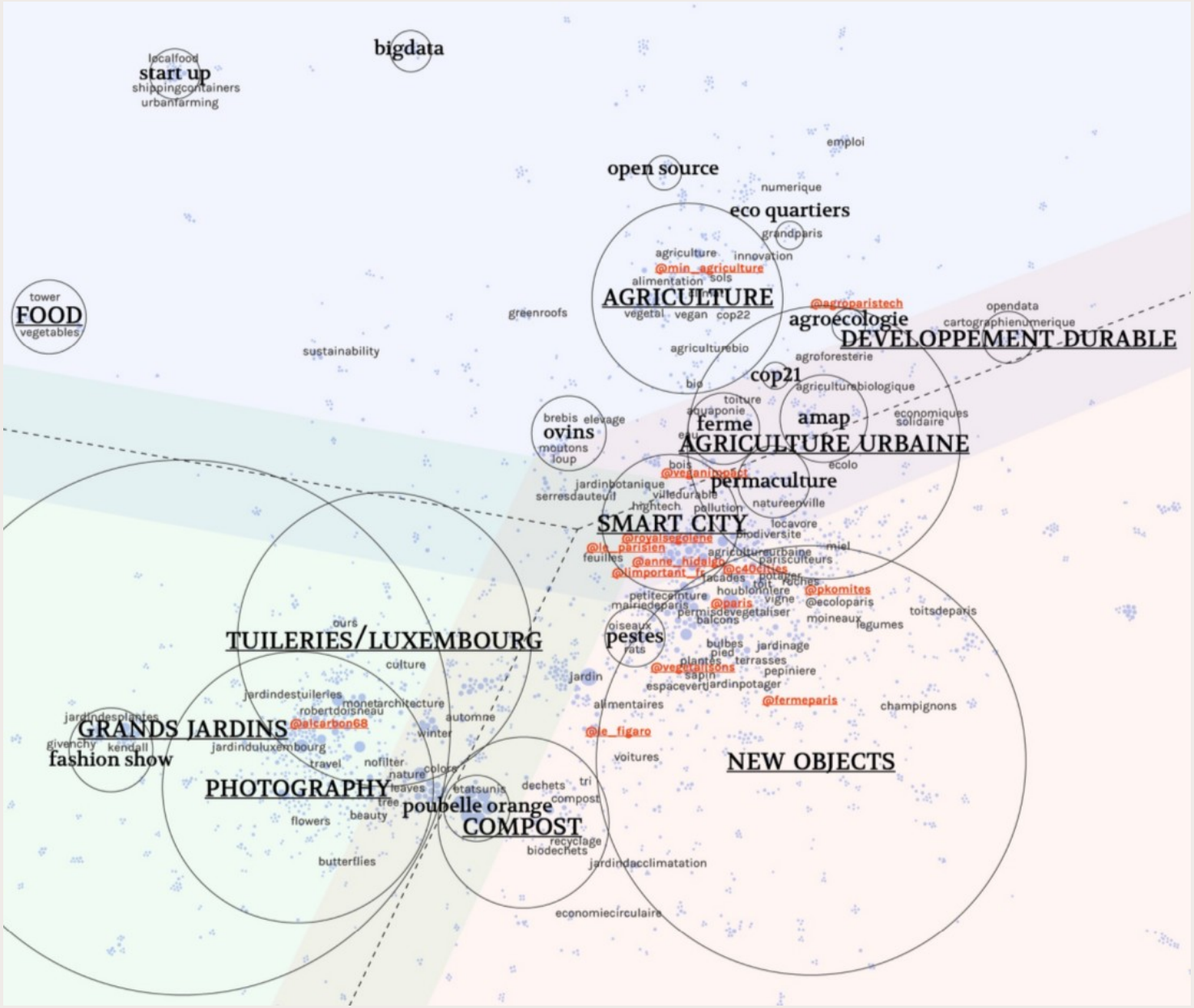
Liens de proximité Twitter entre les élus du Conseil Régional d'Île-de-France et du Conseil de Paris

Explorer les dynamiques temporelles



SciencesPo
MÉDIALAB

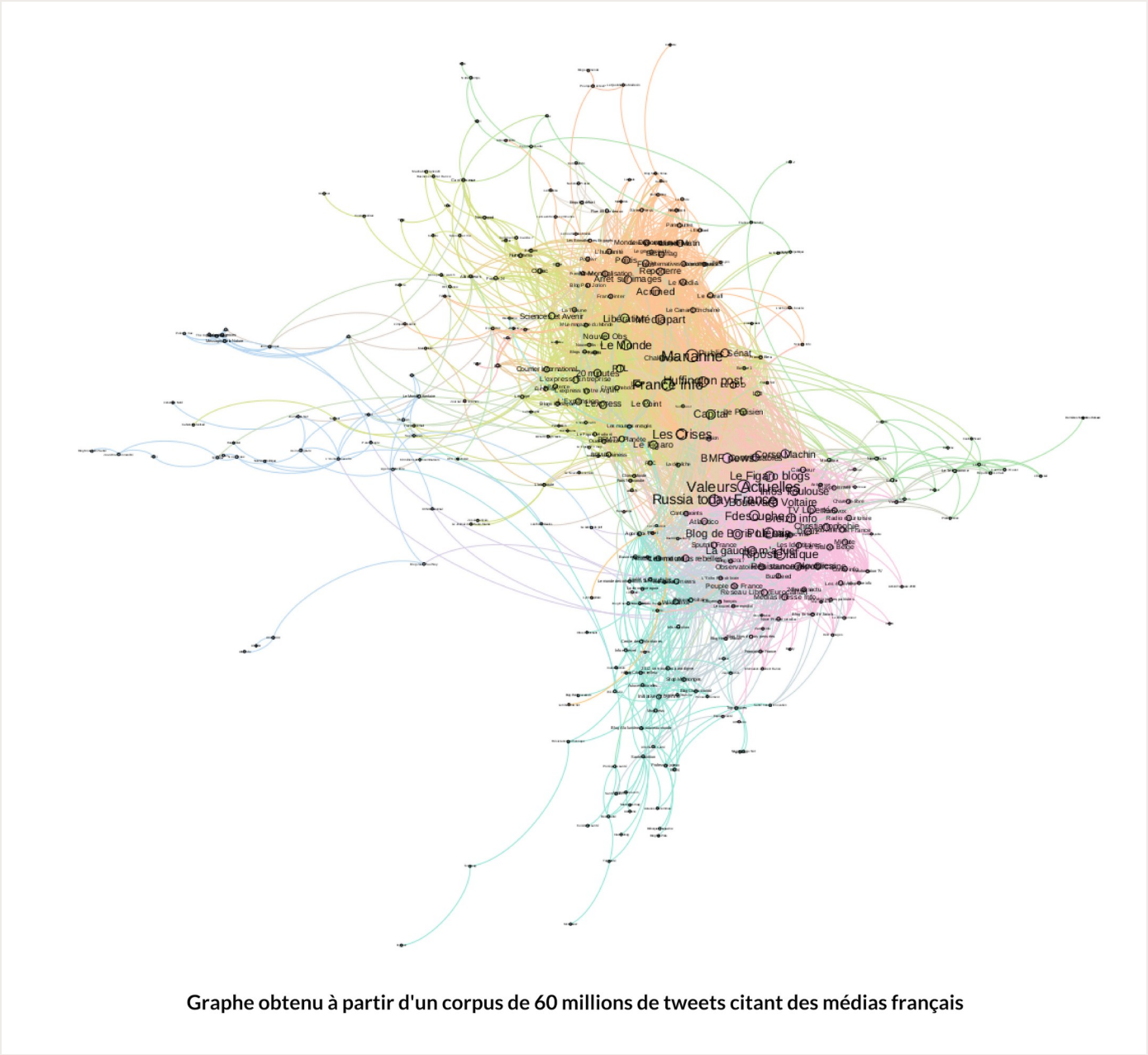
Explorer des réseaux sémantiques



Explorer l'espace visuel d'un corpus

RICCI, Donato, COLOMBO, Gabriele, MEUNIER, Axel, et al. **Designing Digital Methods to monitor and inform Urban Policy. The case of Paris and its Urban Nature initiative.** In: 3rd International Conference on Public Policy (ICPP3)-Panel T10P6 Session 1 Digital Methods for Public Policy. https://re.public.polimi.it/bitstream/11311/1038509/1/IPPA_Ricci-Colombo-Meunier-Brilli.pdf

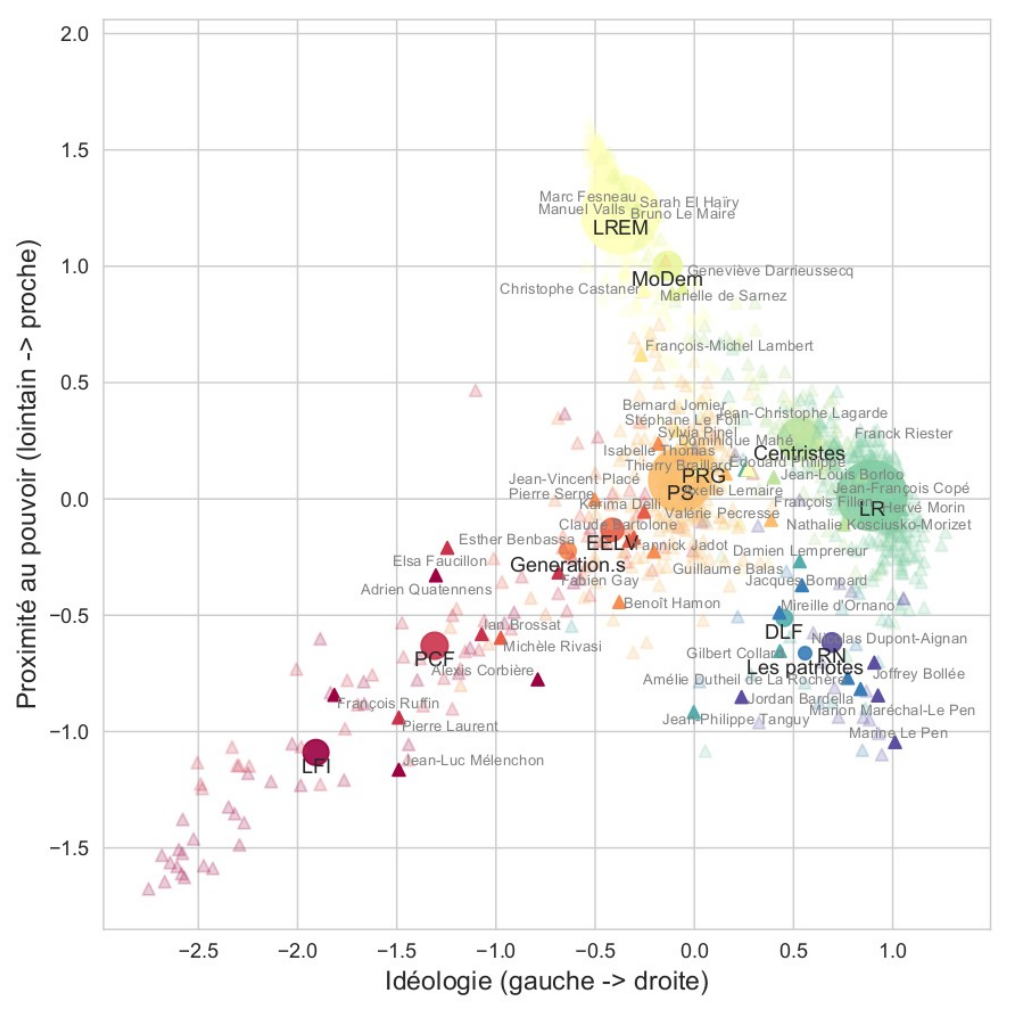
Explorer des réseaux de co-citation de sites web



Graphe obtenu à partir d'un corpus de 60 millions de tweets citant des médias français

Étudier la propagation de l'info sur l'espace politique

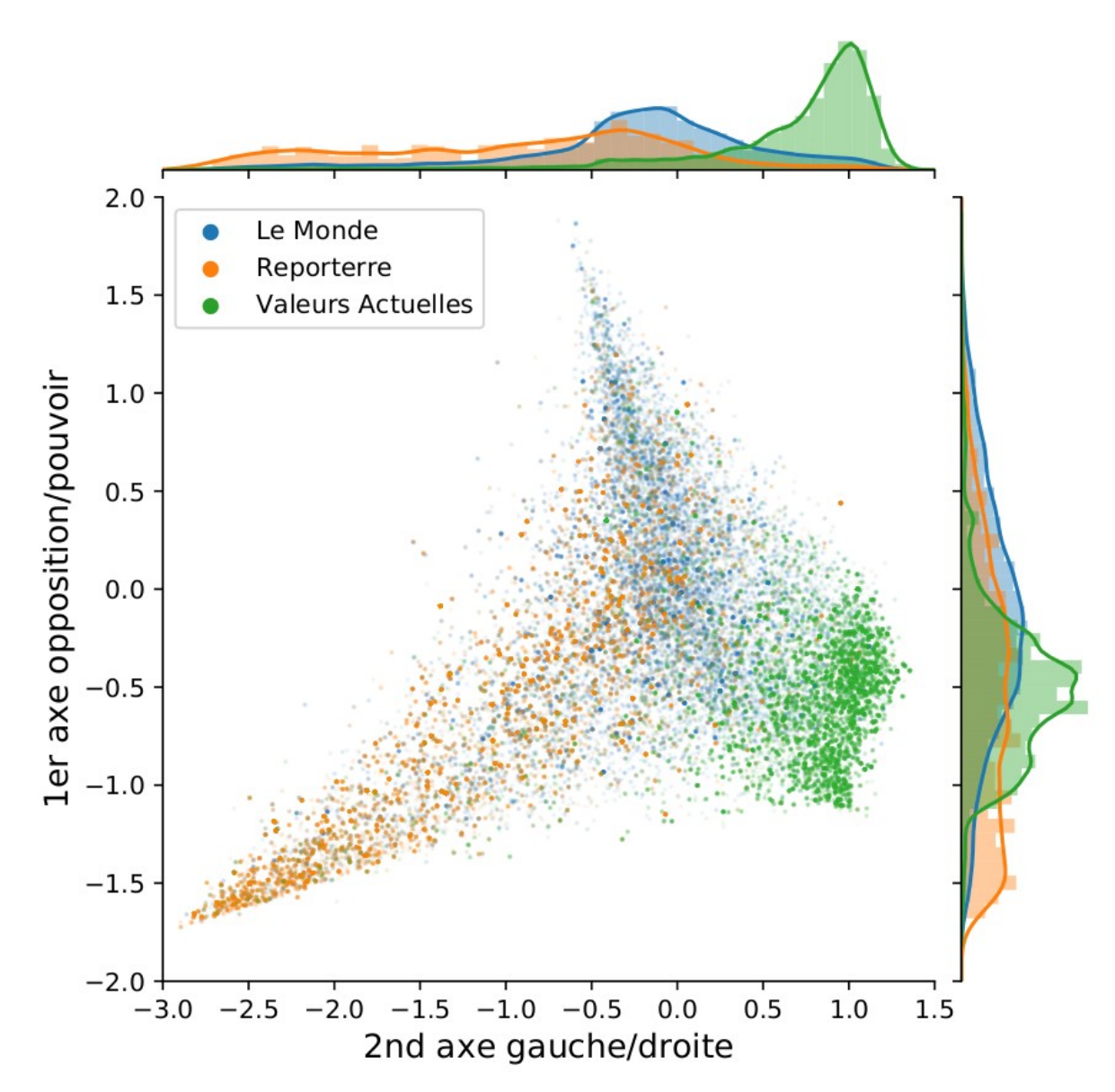
Cartographie de l'espace politique français des utilisateurs de Twitter à partir des followers de parlementaires (méthode de Barbera)



Jean-Philippe Cointet, Pedro Ramaciotti Morales, Dominique Cardon, Caterina Froio, Benjamin Ooghe, et al.. **De quelle(s) couleur(s) sont les Gilets jaunes ? Plonger des posts Facebook dans un espace idéologique latent.** Statistique et Société, Société française de statistique, 2021. https://hal.archives-ouvertes.fr/hal-03700083

Étudier la propagation de l'info sur l'espace politique

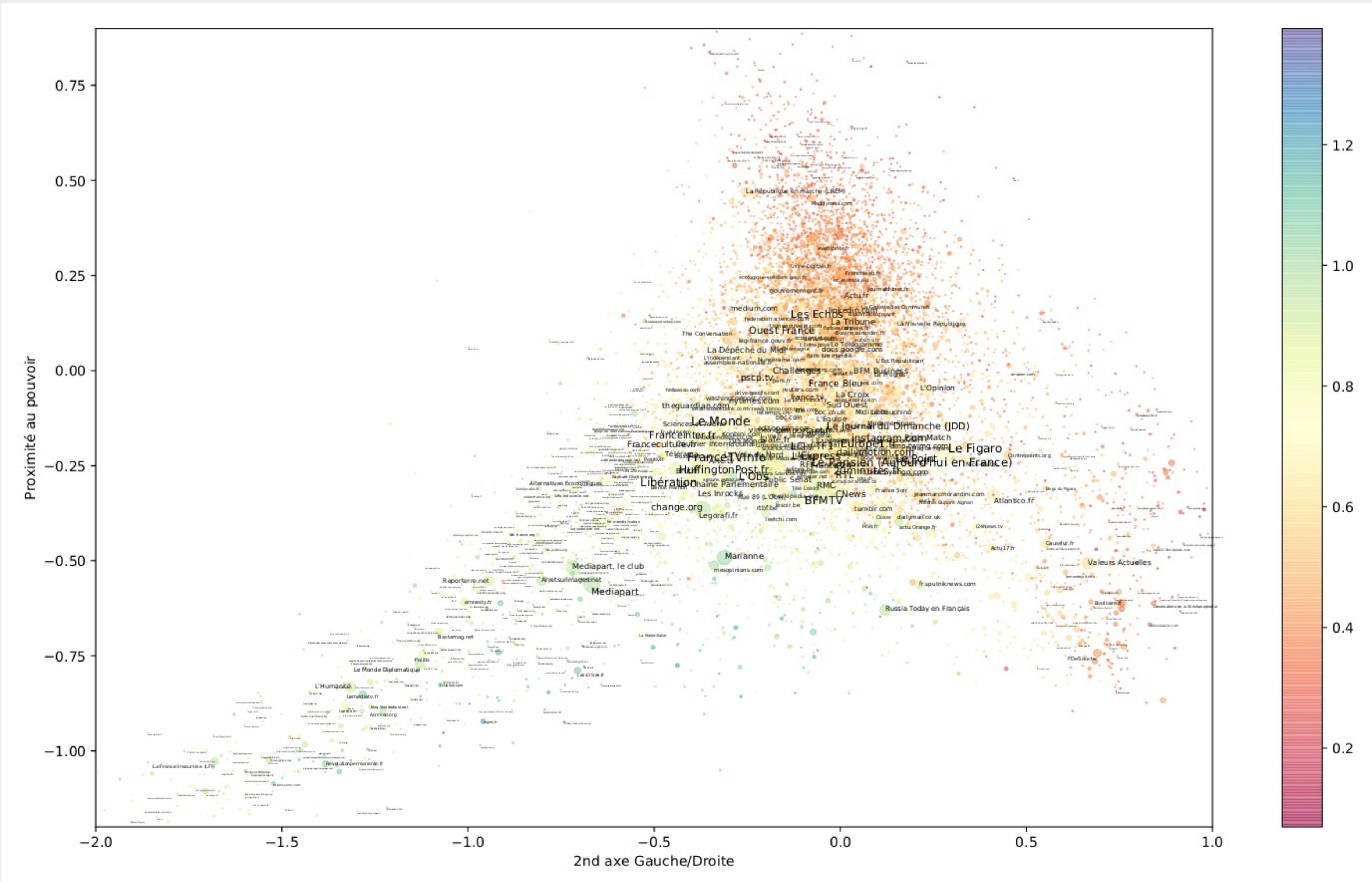
Projection des articles de médias partagés par les utilisateurs de Twitter sur la cartographie de l'espace politique



SciencesPo MÉDIALAB

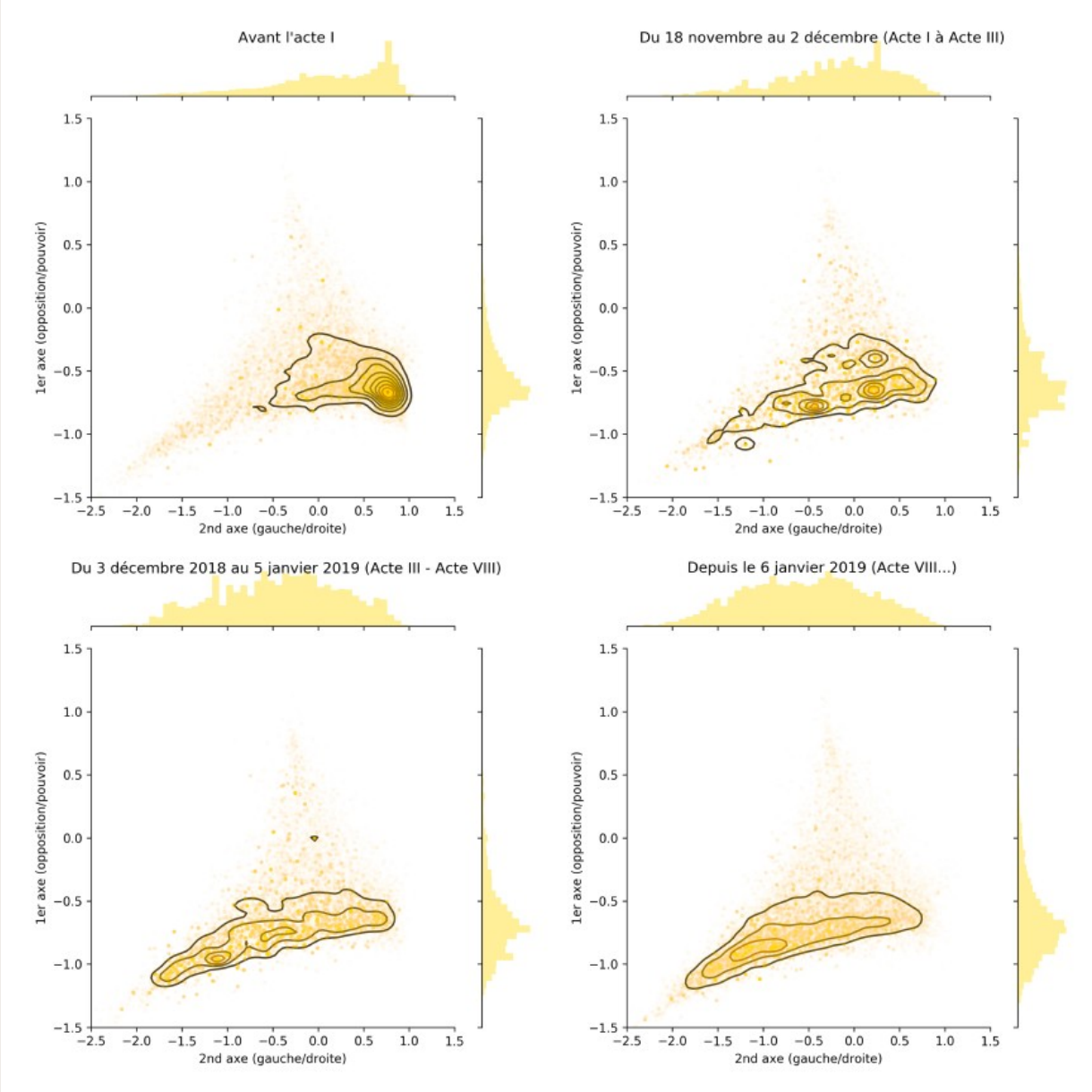
Étudier la propagation de l'info sur l'espace politique

Application à l'ensemble des médias partagés pour positionner ceux-ci sur la cartographie de l'espace politique



Étudier la propagation de l'info sur l'espace politique

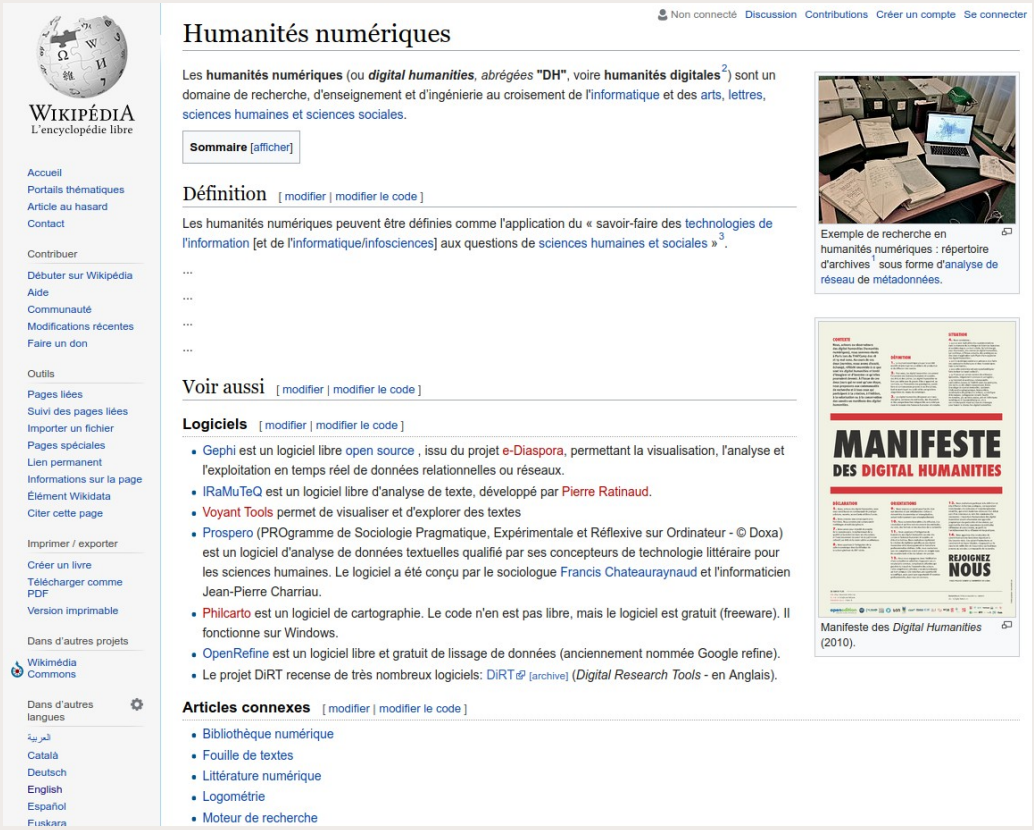
Analyse dynamique des médias partagés sur les groupes Facebook de Gilets Jaunes au fil du mouvement



SeeAlsology : exploration sémantique rapide

http://tools.medialab.sciences-po.fr/seealsology/

Construire & explorer un réseau sémantique de liens entre concepts identifiés via les sections « Voir aussi » de Wikipedia



Wikipédia
L'encyclopédie libre

Accueil
Portails thématiques
Article au hasard
Contact

Contribuer
Débuter sur Wikipédia
Aide
Communauté
Modifications récentes
Faire un don

Outils
Pages liées
Suivi des pages liées
Importer un fichier
Pages spéciales
Lien permanent
Informations sur la page
Élément Wikidata
Citer cette page

Imprimer / exporter
Créer un livre
Télécharger comme PDF
Version imprimable

Dans d'autres projets
Wikimédia Commons

Dans d'autres langues
العربية
Català
Deutsch
English
Español
Euskara

Non connecté Discussion Contributions Créer un compte Se connecter

Humanités numériques

Les **humanités numériques** (ou ***digital humanities***, *abrégées* **"DH"**, voire **humanités digitales**[2]) sont un domaine de recherche, d'enseignement et d'ingénierie au croisement de l'informatique et des arts, lettres, sciences humaines et sciences sociales.

Sommaire [afficher]

Définition [modifier | modifier le code]

Les humanités numériques peuvent être définies comme l'application du « savoir-faire des technologies de l'information [et de l'informatique/infosciences] aux questions de sciences humaines et sociales »[3].

...
...
...
...

Voir aussi [modifier | modifier le code]

Logiciels [modifier | modifier le code]

- Gephi est un logiciel libre open source , issu du projet e-Diaspora, permettant la visualisation, l'analyse et l'exploitation en temps réel de données relationnelles ou réseaux.
- IRaMuTeQ est un logiciel libre d'analyse de texte, développé par Pierre Ratinaud.
- Voyant Tools permet de visualiser et d'explorer des textes
- Prospero (PROgramme de Sociologie Pragmatique, Expérimentale et Réflexive sur Ordinateur - © Doxa) est un logiciel d'analyse de données textuelles qualifié par ses concepteurs de technologie littéraire pour les sciences humaines. Le logiciel a été conçu par le sociologue Francis Chateauraynaud et l'informaticien Jean-Pierre Charriau.
- Philcarto est un logiciel de cartographie. Le code n'en est pas libre, mais le logiciel est gratuit (freeware). Il fonctionne sur Windows.
- OpenRefine est un logiciel libre et gratuit de lissage de données (anciennement nommée Google refine).
- Le projet DiRT recense de très nombreux logiciels: DiRT [archive] (*Digital Research Tools* - en Anglais).

Articles connexes [modifier | modifier le code]

- Bibliothèque numérique
- Fouille de textes
- Littérature numérique
- Logométrie
- Moteur de recherche

Exemple de recherche en humanités numériques : répertoire d'archives[1] sous forme d'analyse de réseau de métadonnées.

Manifeste des *Digital Humanities* (2010).

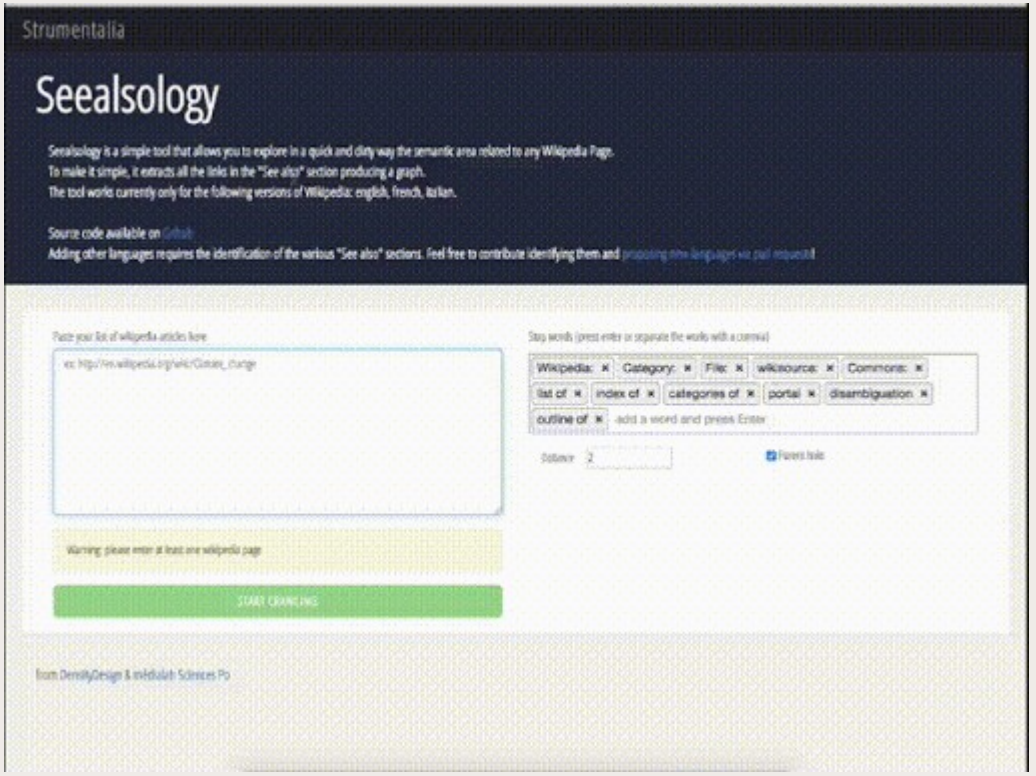
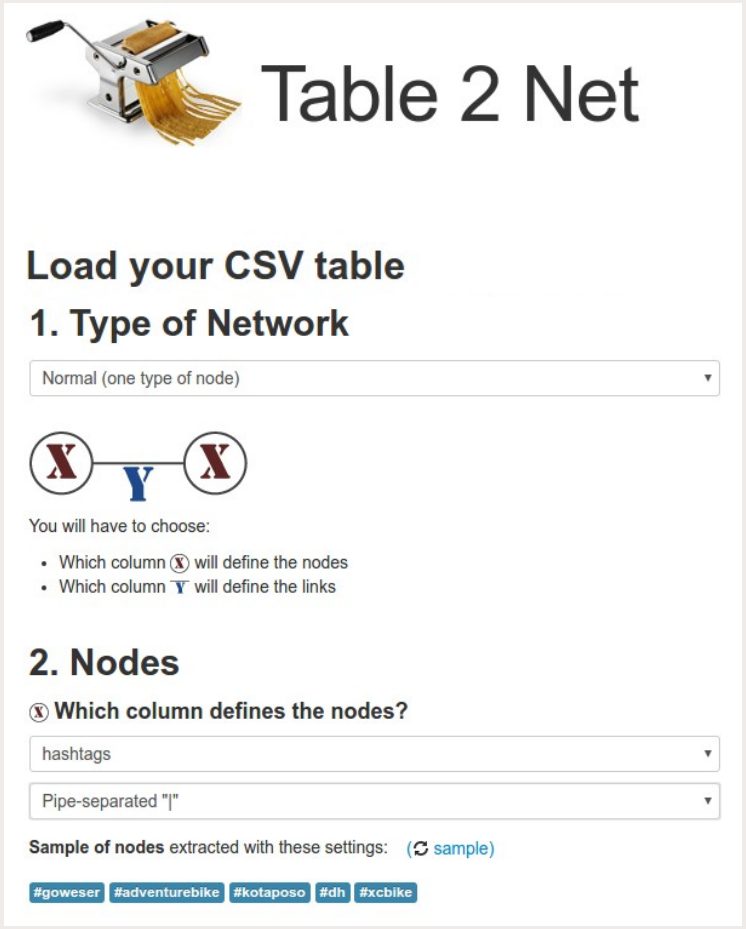
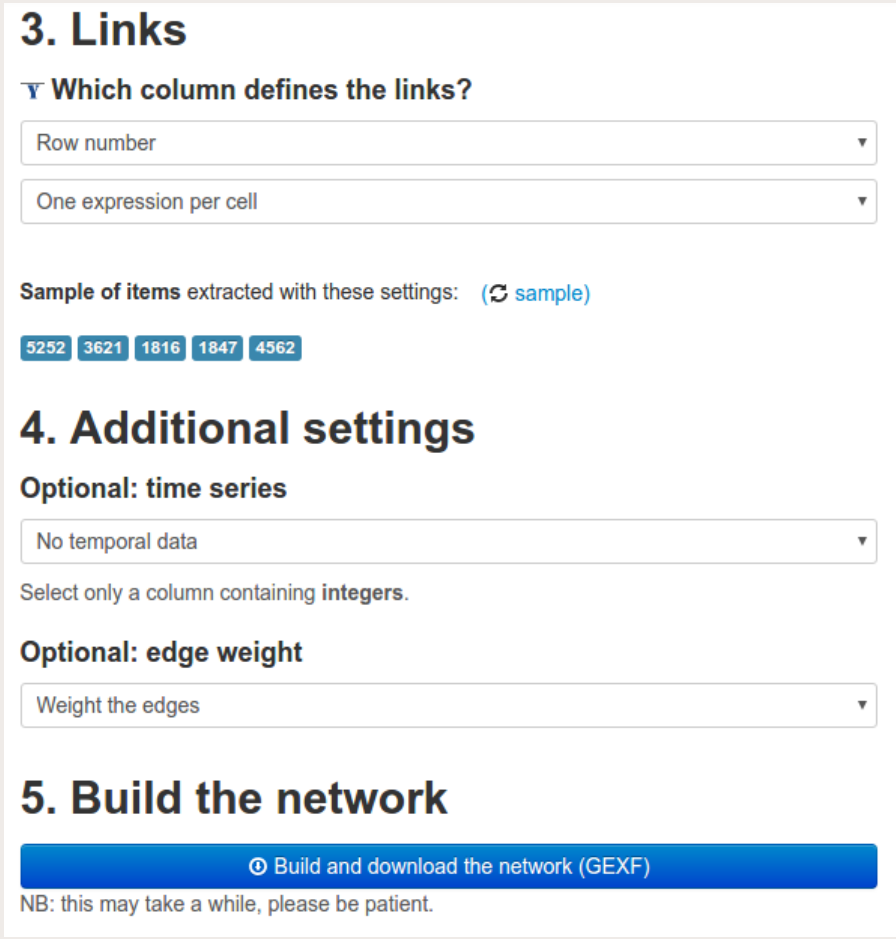


Table2Net : construire un réseau à partir d'un CSV

http://tools.medialab.sciences-po.fr/table2net/

Générer un réseau de liens entre éléments à partir des données d'un fichier tableur

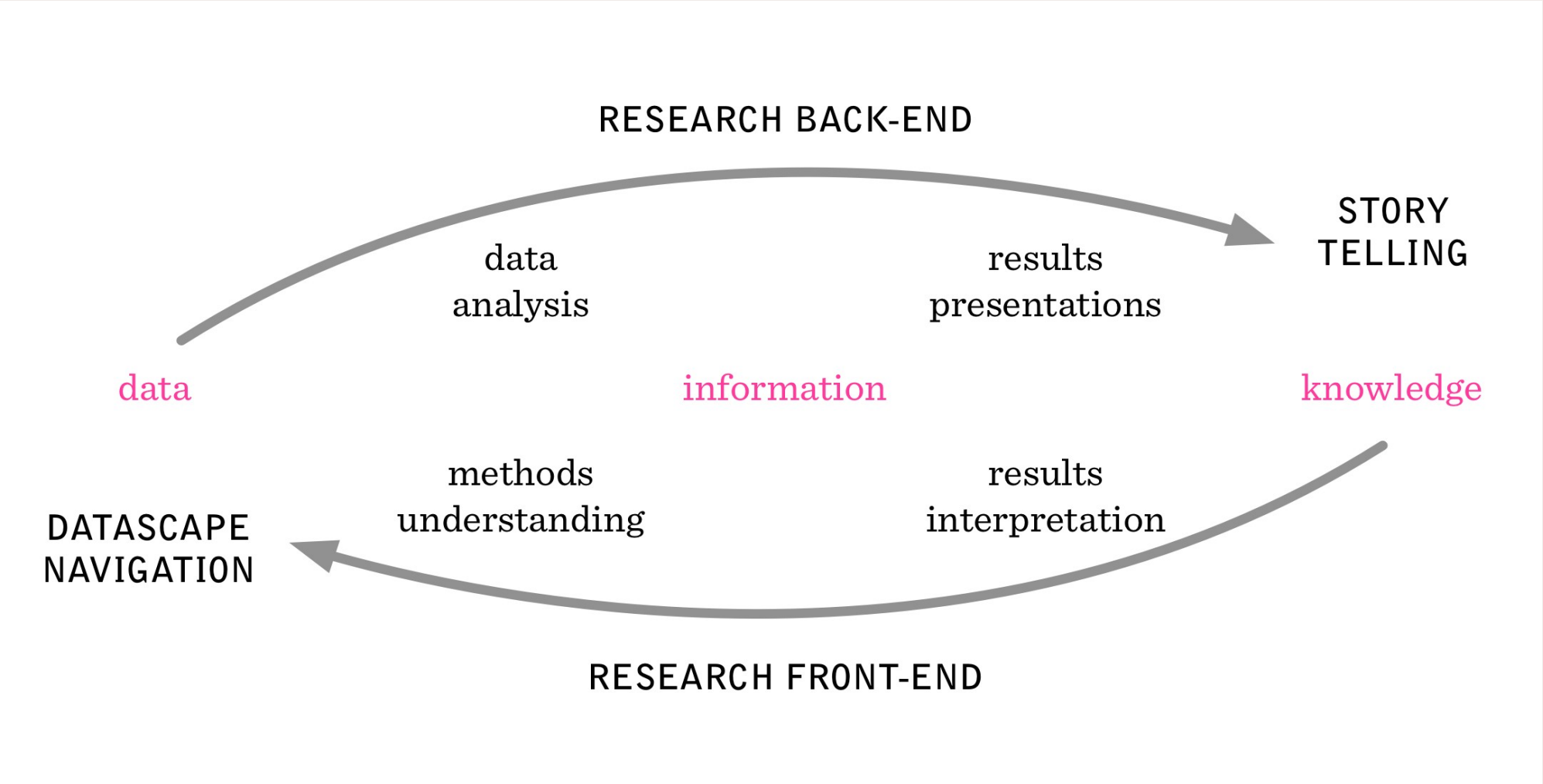




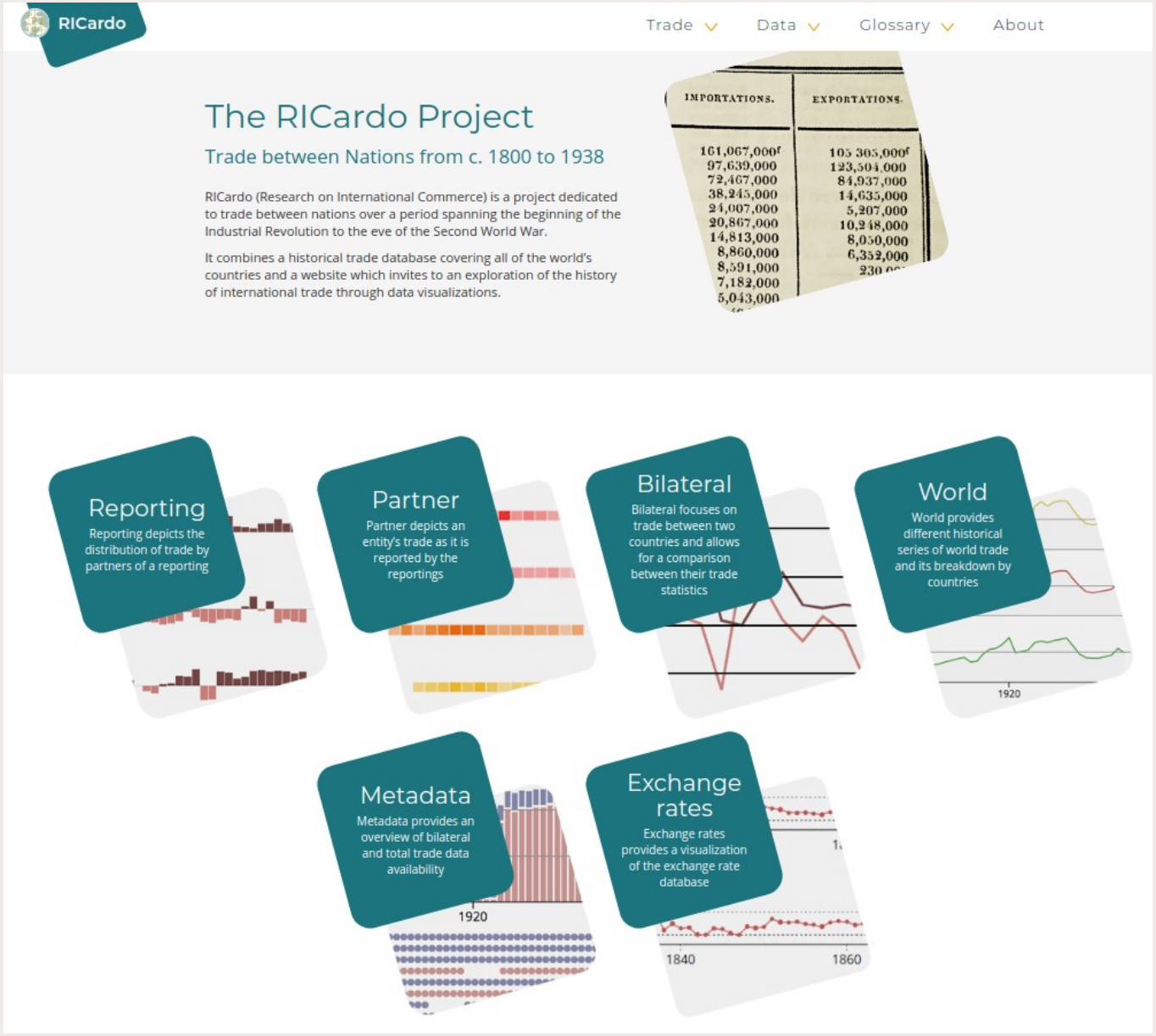
→ **Visualisation de réseaux avec Gephi / Nansi / MiniVan**

Les « datascapes » : zoomer/dézoomer

Construire des interfaces interactives d'exploration d'un jeu de données à différentes granularité



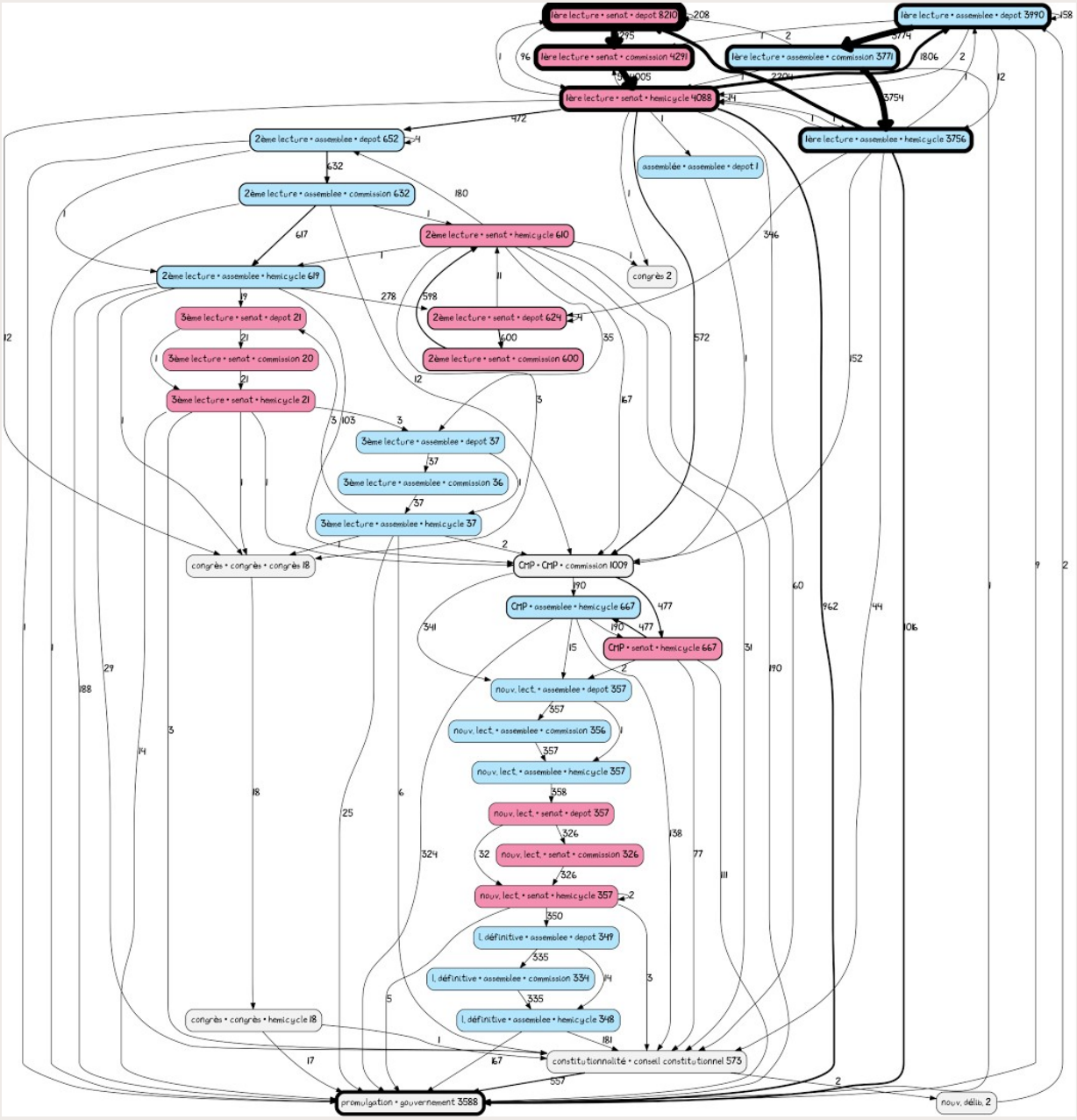
RICardo : visualiser les échanges au 19ème siècle



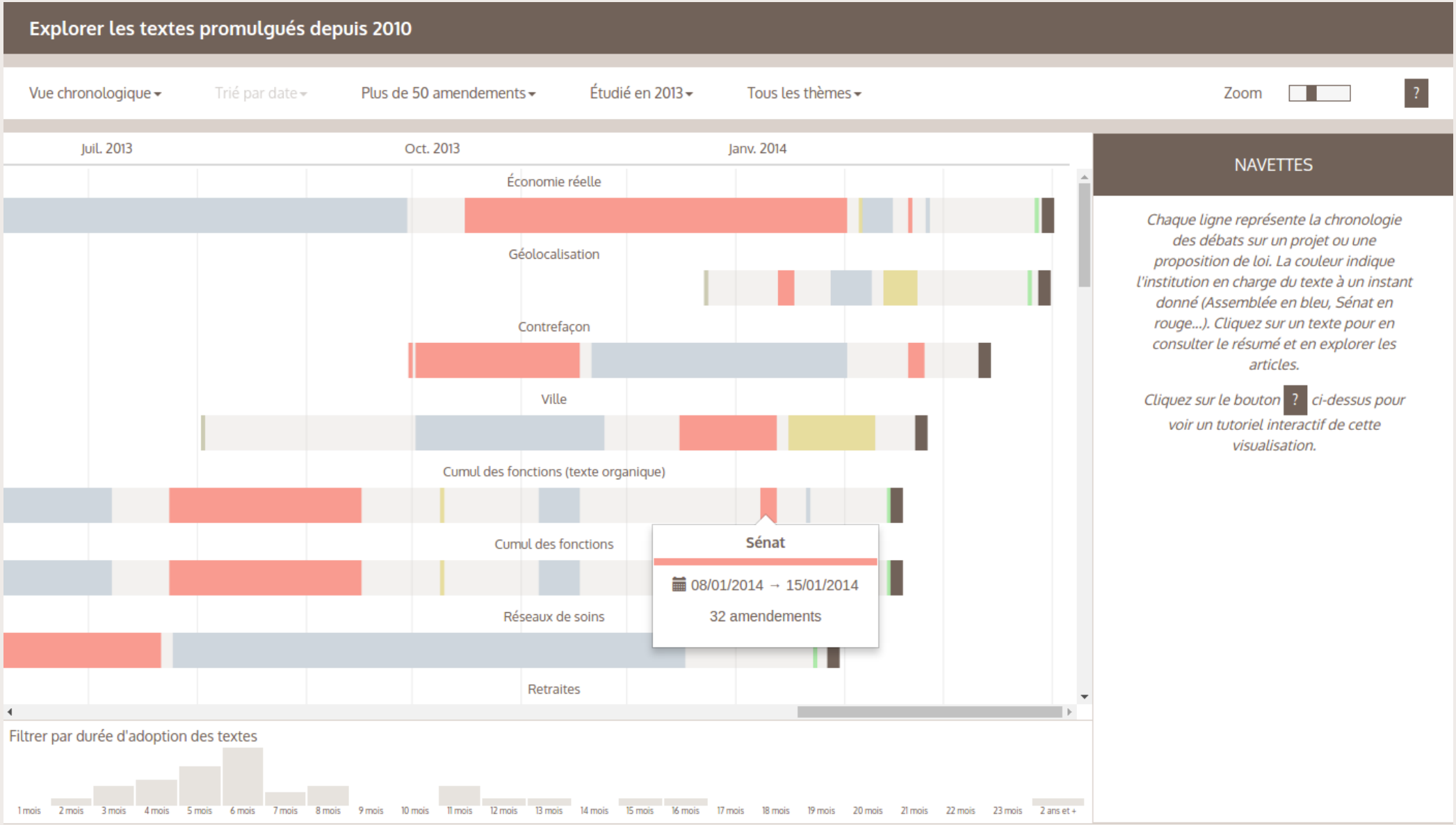
http://ricardo.medialab.sciences-po.fr

LaFabriqueDeLaLoi : explorer la complexité législative

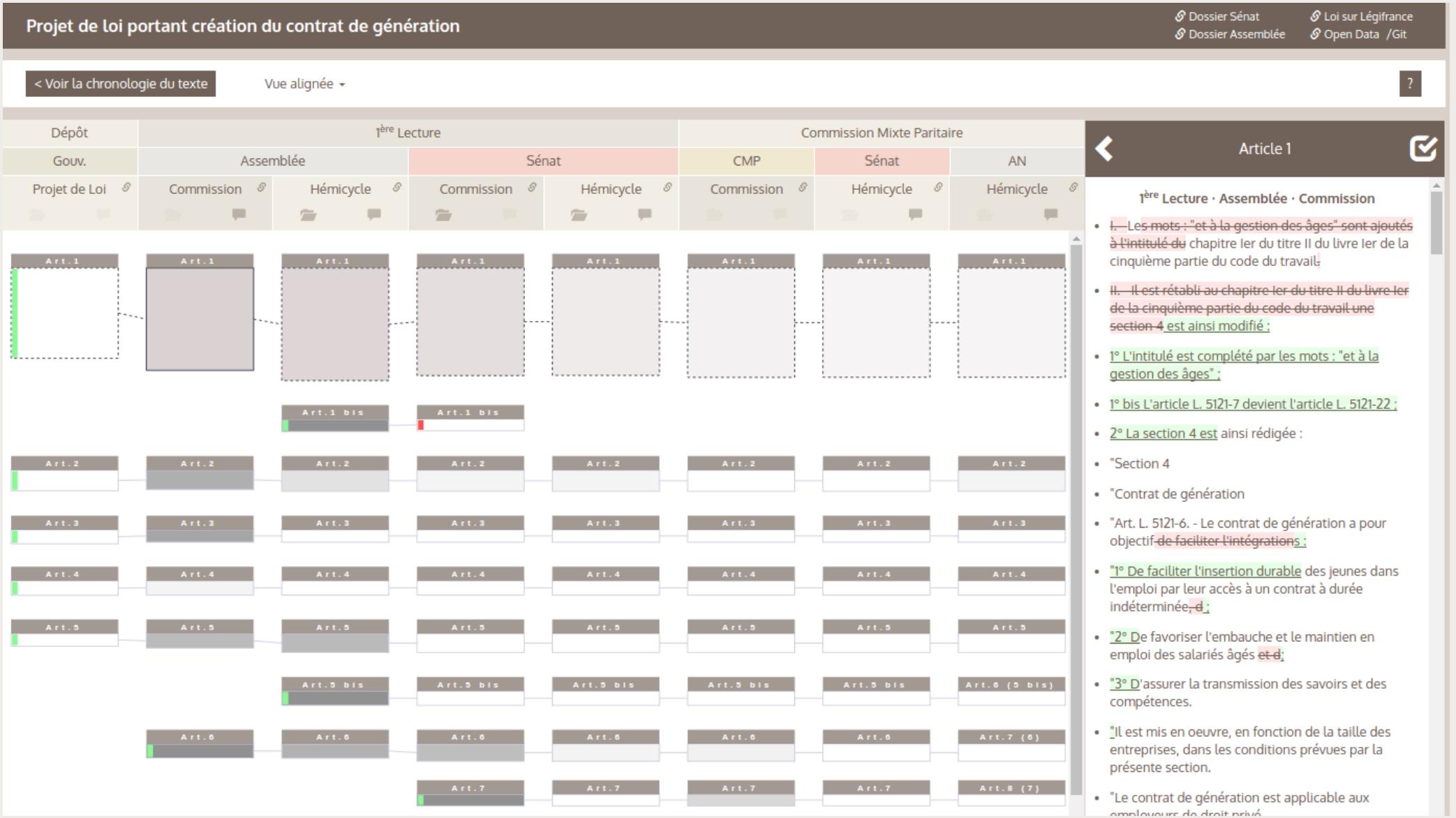
LA FABRIQUE DE LA LOI.FR



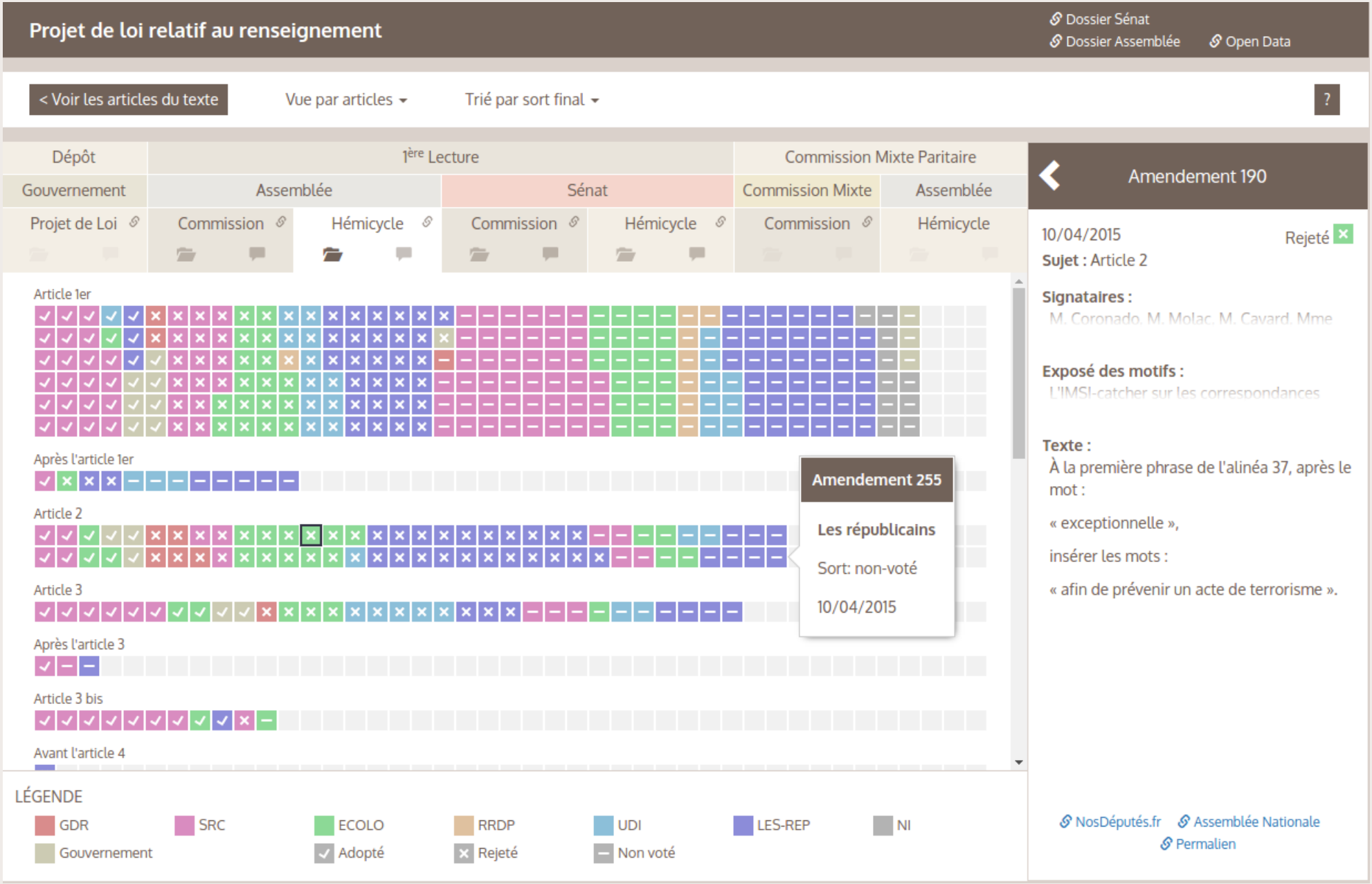
LFDLL : plus de 1000 lois promulguées ou en cours



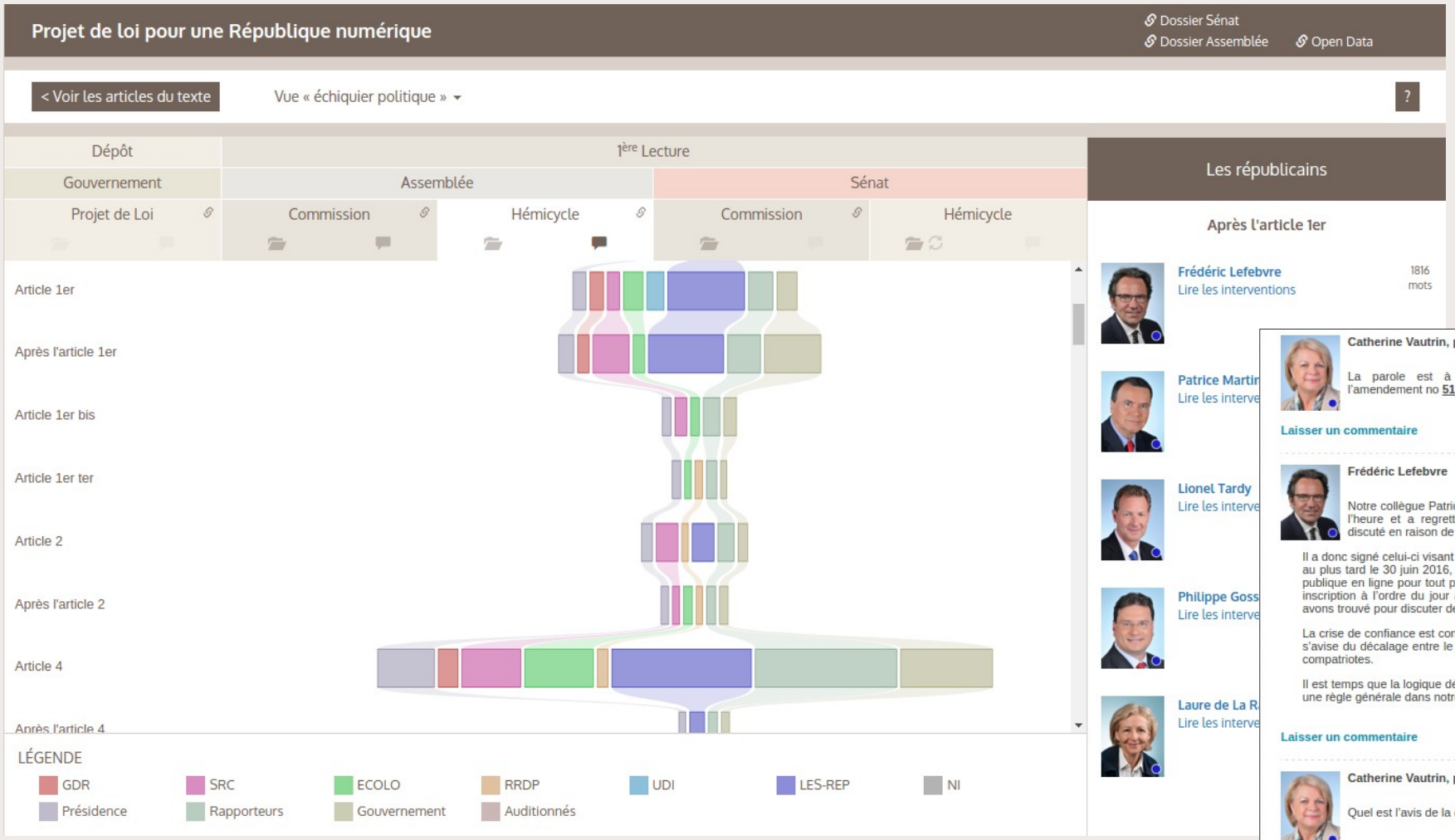
LFDLL : retracer l'évolution du texte à chaque étape



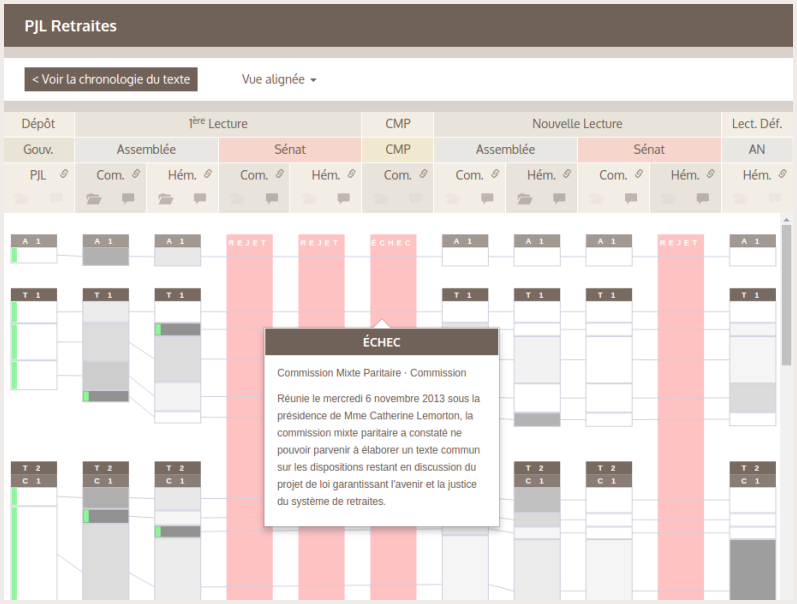
LFDLL : retrouver les amendements discutés et votés

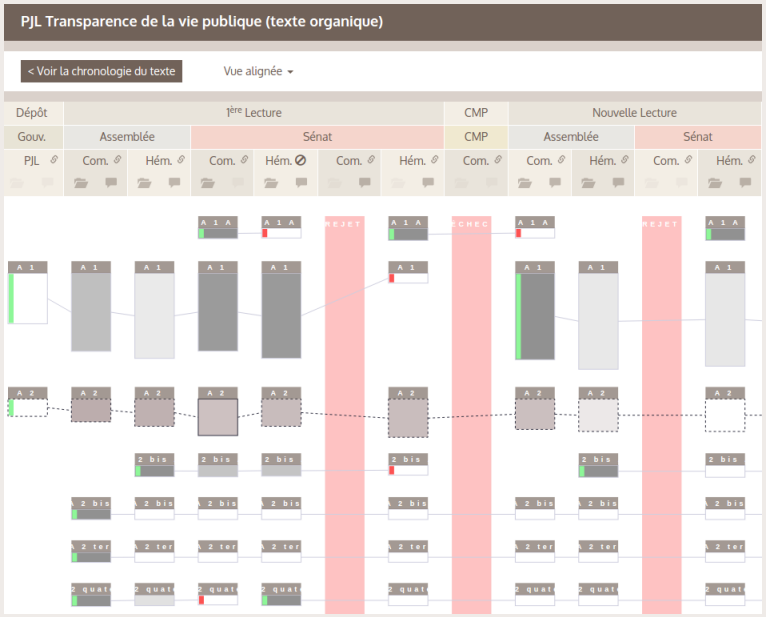


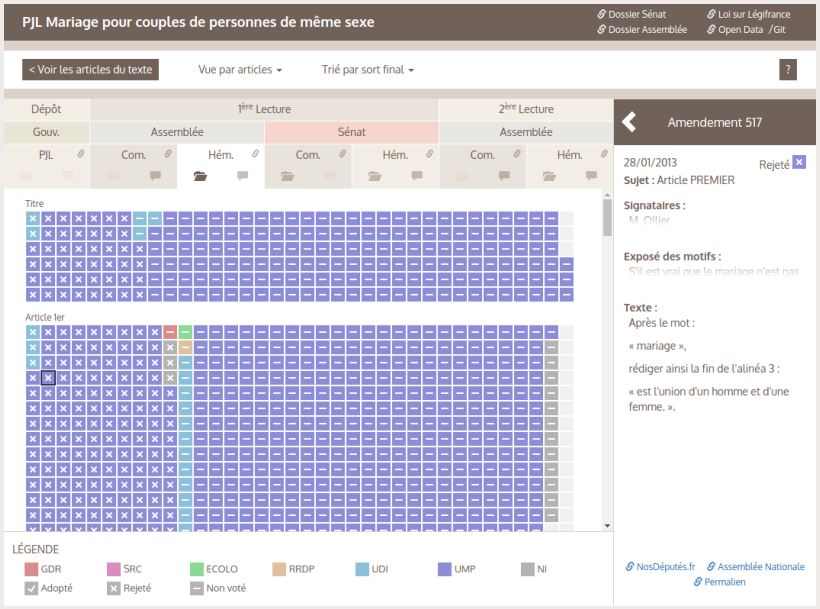
LFDLL : remonter aux débats parlementaires

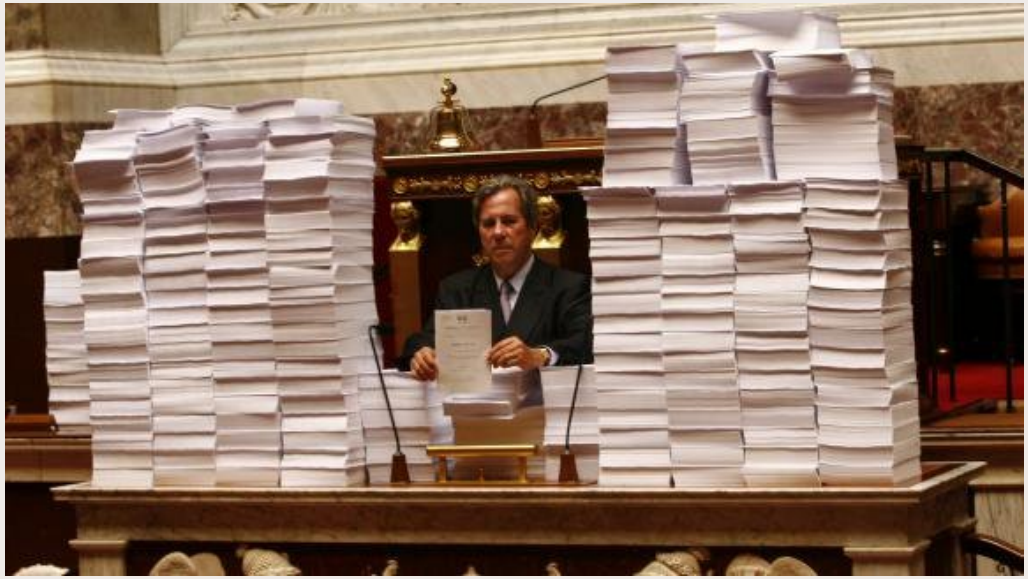


LFDLL : révéler les techniques d'obstruction

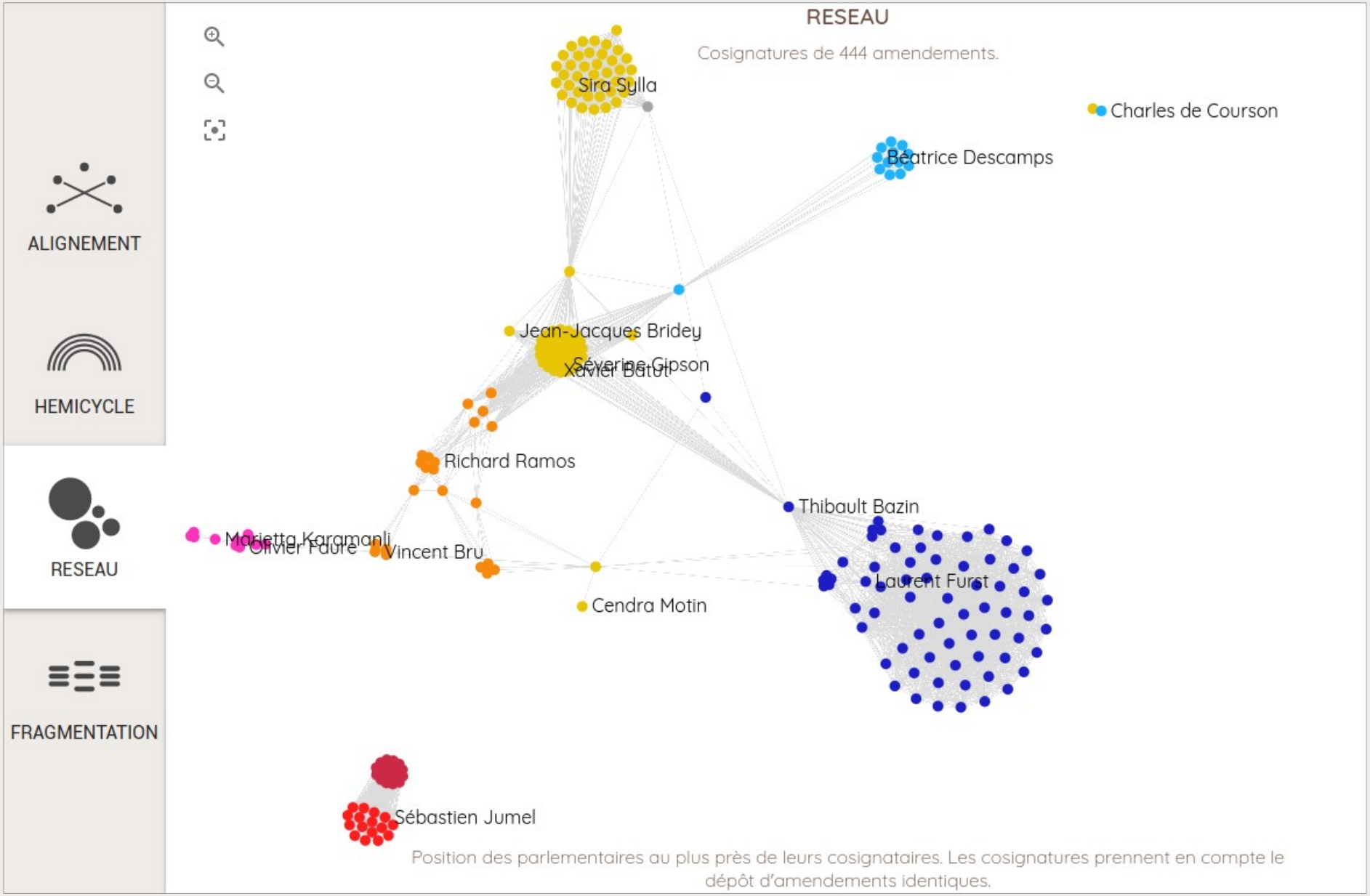








LFDLL : l'alignement et la fragmentation des partis



À vos questions !

https://medialab.sciencespo.fr

benjamin.ooghe@sciencespo.fr

@boogheta@piaille.fr @boogheta @medialab_ScPo