



HAL
open science

Hyperlink is not dead!

Benjamin Ooghe-Tabanou, Mathieu Jacomy, Paul Girard, Guillaume Plique

► **To cite this version:**

Benjamin Ooghe-Tabanou, Mathieu Jacomy, Paul Girard, Guillaume Plique. Hyperlink is not dead!. WS.2 2018 International conference on Web Studies, Paris, France - October 03 - 05, 2018, Oct 2018, Paris, France. hal-03903954

HAL Id: hal-03903954

<https://sciencespo.hal.science/hal-03903954>

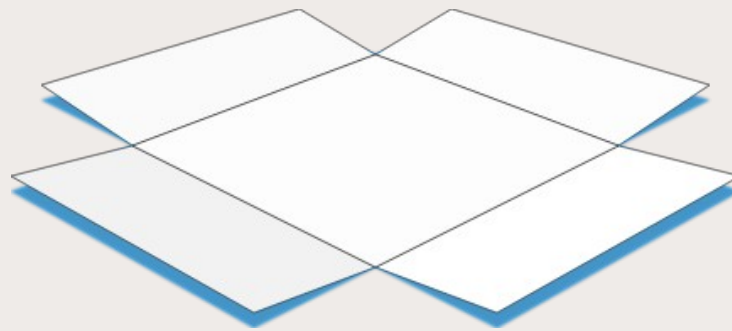
Submitted on 16 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



Equipex DIME-SHS
ANR-10-EQPX-19-01

Hyperlink is not dead!

Digital Tools & Uses Congress
MSH Paris Nord - October 4th, 2018

Benjamin Ooghe-Tabanou – Mathieu Jacomy – Paul Girard – Guillaume Plique
Sciences Po médialab (@medialab_ScPo)

SciencesPo
MÉDIALAB



Hyperlink is not dead!

Benjamin Ooghe-Tabanou
Sciences Po, médialab
Paris, France
benjamin.ooghe@sciencespo.fr

Paul Girard
Sciences Po, médialab
Paris, France
paul.girard@science.spo.fr

Mathieu Jacomy
Sciences Po, médialab
Paris, France
mathieu.jacomy@gmail.com

Guillaume Plique
Sciences Po, médialab
Paris, France
guillaume.plique@sciencespo.fr

ABSTRACT

The emergence and success of web platforms nurtured a trend within social studies: "Hyperlink is dead!". Capturing their users into mobile applications and specialised web interface to propose them a specific user experience (and business model), the platforms indeed created new information silos in the open World Wide Web space. The simplified availability of user behavioural data through these platforms APIs reinforced this idea in academic communities by providing scholars with an easy way to collect rich user centric data for their research. After discussing the methodological aspects of the web divide between platforms and classical websites, we will argue that although it becomes more and more invisible, the hyperlink, modern incarnation of intertextual links between documents, is still a central and structural element of the web. Hyperlinks remain an invaluable resource to turn the web into a research field in spite of the complexity to collect, manipulate and curate them. We will illustrate those methodological challenges by describing the choices we made in designing Hyphe, a tool dedicated to the creation of web corpora tailored for mining hypertexts.

CCS CONCEPTS

• Information systems → Web mining; Web applications; Internet communications tools;

KEYWORDS

Hyperlink, hypertext, web mining, crawler, corpus, curation, network analysis.

ACM Reference Format:

Benjamin Ooghe-Tabanou, Mathieu Jacomy, Paul Girard, and Guillaume Plique. 2018. Hyperlink is not dead!. In *International conference on Web Studies (WS'2 2018)*, October 3–5, 2018, Paris, France. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3240431.3240434>

1 ARE HYPERTEXT STUDIES OUTDATED?

The World Wide Web's original design as a vast open documentary space built around the concept of hypertext made it a fantastic

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WS'2 2018, October 3–5, 2018, Paris, France
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6438-6/18/10.
<https://doi.org/10.1145/3240431.3240434>

research field to study networks of actors. As a literary technology, hyperlinks concepts are anything but new: "Links are intrinsic to documents, and have been for millennia" [27]. According to Bardini [5], two main concepts are embedded in hypertexts: association and connection. Hypertexts allow to create conceptual associations between documents - a powerful function when used in a free and creative context - but they can also be very efficient in connecting documents to communicate ideas (i.e. conceptual links) within a community.

When Tim Berners-Lee proposed his World Wide Web (WWW) project, hypertexts were presented more as a way to connect documents to enhance communication through navigation than as form of conceptual associations: "The texts are linked together in a way that one can go from one concept to another to find the information one wants. (...) The process of proceeding from node to node is called navigation" [6]. But following Bardini [5], if associative hyperlinks are created freely by authors for their own use, connective ones which have a value within a community are more likely to be moderated. Although the goal of the WWW is to enhance information flows across communities, connections between documents are not controlled, as each individual website's author is responsible only for the connections from his website to the rest of the WWW.

1.1 Hyperlinks directionality: a bottom-up hierarchy

This directionality of the links reveals asymmetrical associations between the linked documents: the referer knows the referee but not necessarily the other way around. Considering hyperlinks as references provides powerful insights on the distribution of influence on the web. As the study of complex networks has demonstrated, online connections are not randomly distributed across the web. According to the principle known as the "Matthew effect" or "preferential attachment" [4], new web documents tend to cite the already most cited documents, reinforcing the concentration of links to a small fraction of pages. A hierarchy naturally emerges from this pattern across all scales of the web: it can be observed locally (eg. inside Wikipedia) as well as in its general structure. This hierarchy is bottom-up because it emerges spontaneously rather than by design, but also because hyperlinks tend to flow from a metaphorical bottom to a metaphorical top. Actors with high visibility drawing most of citations are a handful compared to the mass of low visibility actors who cite them. The structure emerging from the direction of hyperlinks was famously leveraged by Google's

<http://hyphe.medialab.sciences-po.fr/docs/20181004-ACM-WebStudies-HyperlinkIsNotDead.pdf>
<https://doi.org/10.1145/3240431.3240434>

1. From the web to platforms:
a (not so) brief history of hyperlinks
2. médialab: enabling digital field work
through design & engineering
3. Hyphe: curate hypertexts into
web corpora
4. Many angles of hypertext studies

1. From the web to platforms:
a (not so) brief history of hyperlinks
2. médialab: enabling digital field work
through design & engineering
3. Hyphe: curate hypertexts into
web corpora
4. Many angles of hypertext studies

The Web: a vast open documentary space

« *Links are intrinsic to documents,
and have been for millennia.* »

Nelson, T. H. (1992). *Literary Machines* 93.1. Sausalito, CA: Mindful Press

Hyperlinks: the backbone of the world wide web



« The texts are **linked together** in a way that one can go from one concept to another to find the information one wants. The network of links is called a **web**. [...] The texts are known as **nodes**. The process of proceeding from node to node is called **navigation**. »

Tim Berners-Lee, 1990, [WorldWideWeb: Proposal for a HyperText Project](#)

Hyperlinks: a familiar & transparent feature

« *But from the very beginning [...] the web has had **one defining feature** that we tend to overlook today, because it has become **so intuitive and natural that it goes unnoticed.*** »

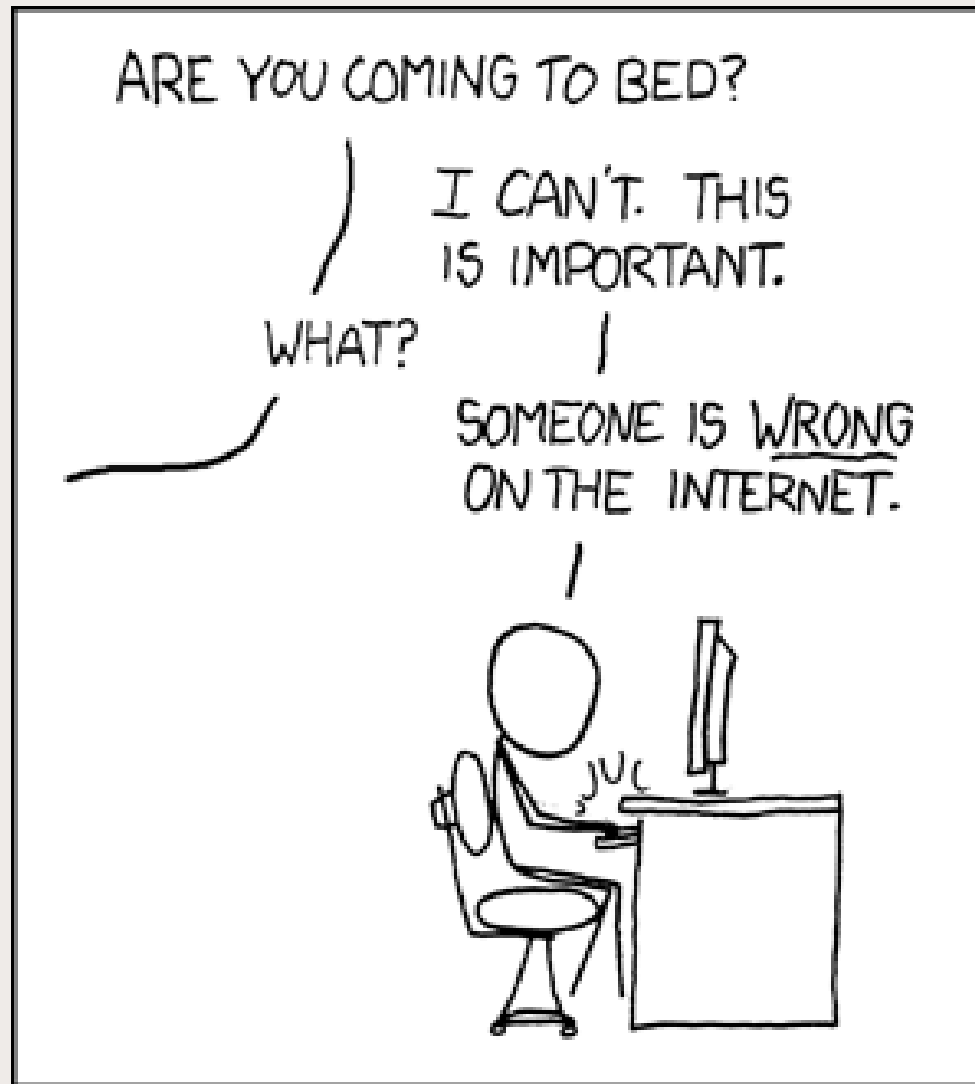
Brügger, 2017, [Connecting textual segments: A brief history of the web hyperlink](#)

Hyperlinks reflect asymmetrical associations

« *A hyperlink is a **manifestation of intention.***
By linking one page to another,
one piece of text to another,
people intend to do particular things. »

Ryfe, Mensing, & Kelley, 2016, What is the meaning of a news link?

The Web: a place of dialogue and debates

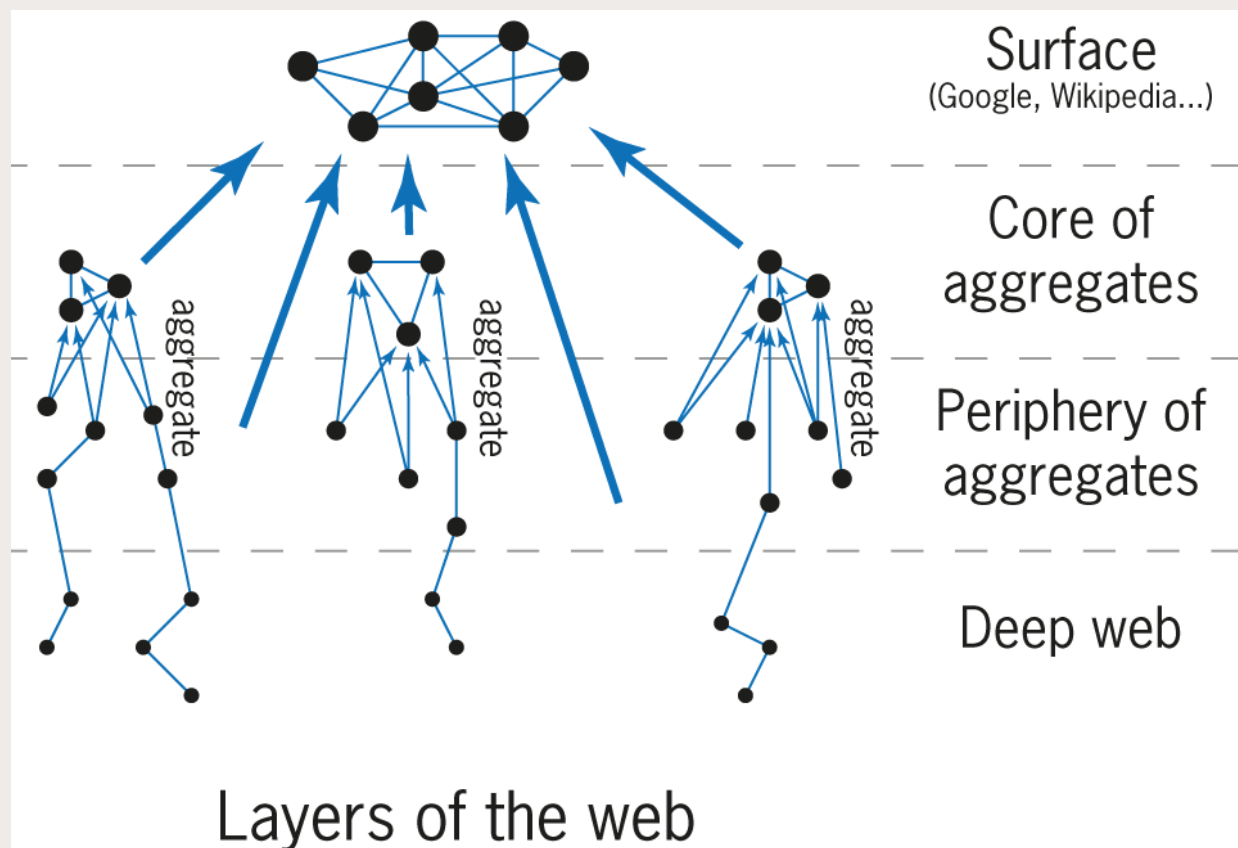


CC-BY-NC - Randall Munroe - XKCD

A bottom-up hierarchy emerged from hyperlinks

« Matthew effect » : preferential attachment

→ new web pages tend to cite the already most cited ones



The 2010's divide: the Web Vs. gated platforms

- 1999: Google's « PageRank » algorithm
 - L. Page, S. Brin & al. 1999. [ThePageRank Citation Ranking: Bringing Order to the Web.](#)
 - development of Search Engines Optimization
 - hyperlinks at the core of online visibility
- 2004: Facebook
- 2006: Twitter
- 2010's: « platformization » of the web, smartphones
 - capturing users with sharing buttons behind closed gates
 - increasing importance of likes & retweets
- 2018: legal definition of platforms in EU copyright reform

*« The definition of an online content sharing service provider under this Directive shall cover **information society service providers** one of the main purposes of which is to **store and give access to the public or to stream significant amounts of copyright protected content** uploaded / made available by its users, and that **optimise content**, and **promote for profit** making purposes, including amongst others displaying, tagging, curating, sequencing, the uploaded works or other subject-matter, irrespective of the means used therefor, and therefore act in an active way. »*

Are hyperlinks studies outdated?

- A new trend in Social Sciences: will LIKES replace LINKS?
- The rise of APIs in SHS:
 - simple access to massive and structured user centric data
(Twitter, Facebook...)
 - development of Single platform studies
 - risks of commercial & ethical close-back
(LinkedIn, Cambridge Analytica...)
- Working with the web otherwise can be hard:
heterogeneous formats, unstructured messy data...

Bruno Latour, médialab founder



« *Google is nice,
but we need
something better* »

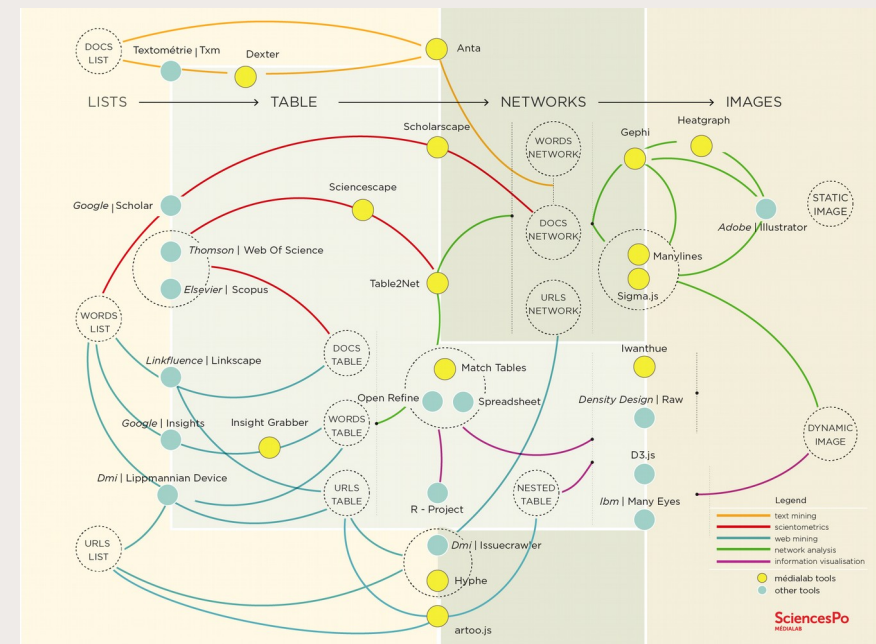
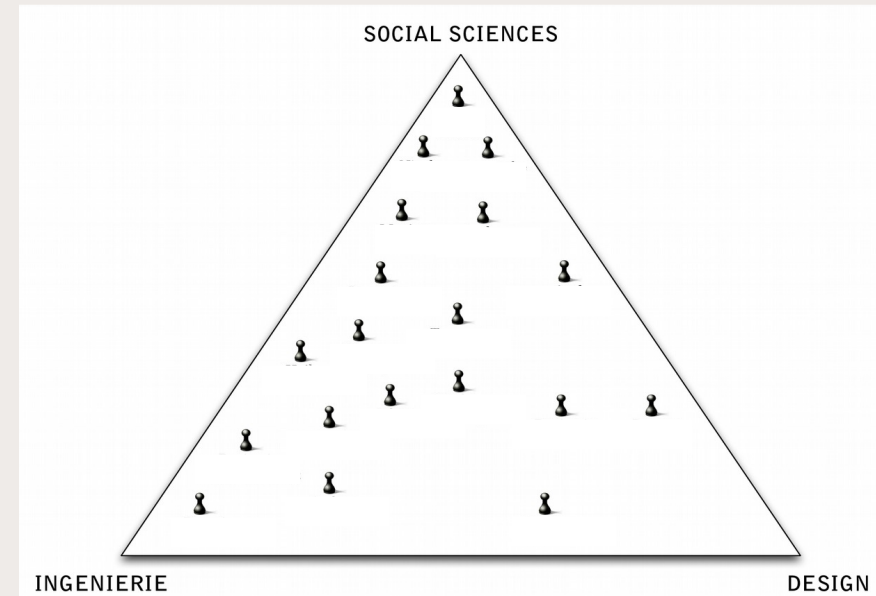
The Indian Express, 2011

1. From the web to platforms:
a (not so) brief history of hyperlinks
2. **médialab: enabling digital field work
through design & engineering**
3. Hyphe: curate hypertexts into
web corpora
4. Many angles of hypertext studies

médialab @ Sciences Po

<https://medialab.sciencespo.fr>

- Pluridisciplinary Research Lab created by Bruno Latour in May 2009, led by Dominique Cardon since 2017
- Social Sciences, Engineering & Design
- Articulate qualitative & quantitative methods through a digital approach
- Work with digital traces
- Deploy an ecosystem of tools
<http://tools.medialab.sciences-po.fr>
- METAT: a monthly Open Support Workshop
<https://www.sciencespo.fr/recherche/fr/content/metat-latelier-de-methodes>



The Web as a research field

DIME Web: one of three instruments of EQUIPEX DIME SHS

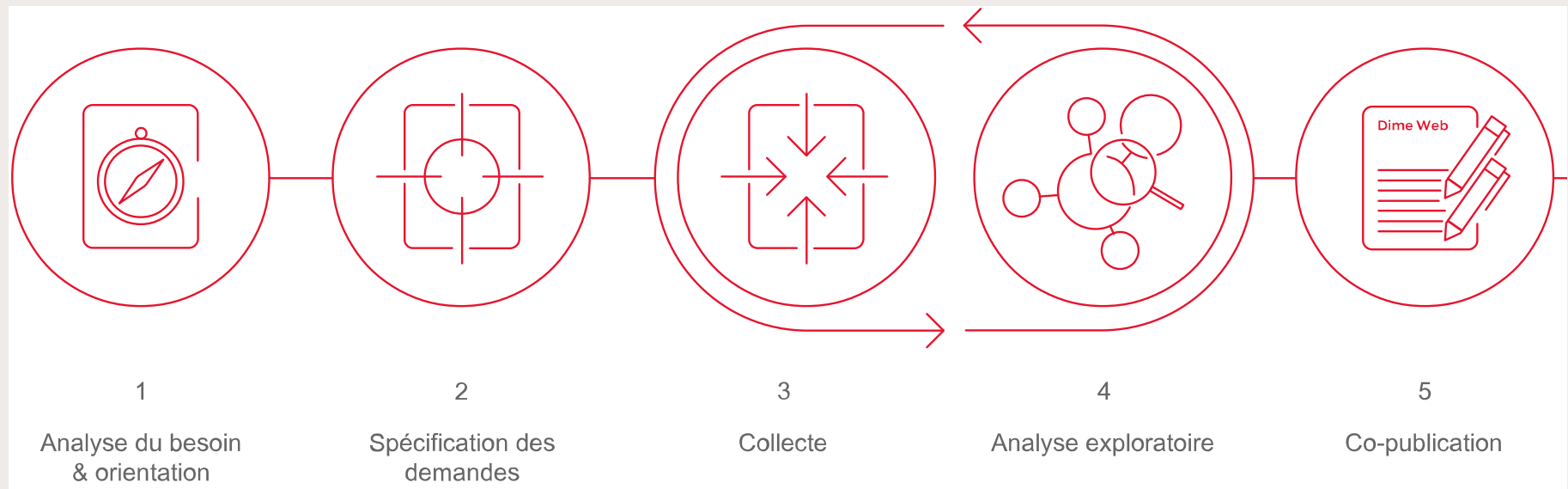
Methodological research and support

Network analysis, Controversies studies

Collect, enrich, clean, visualize & analyze digital traces:

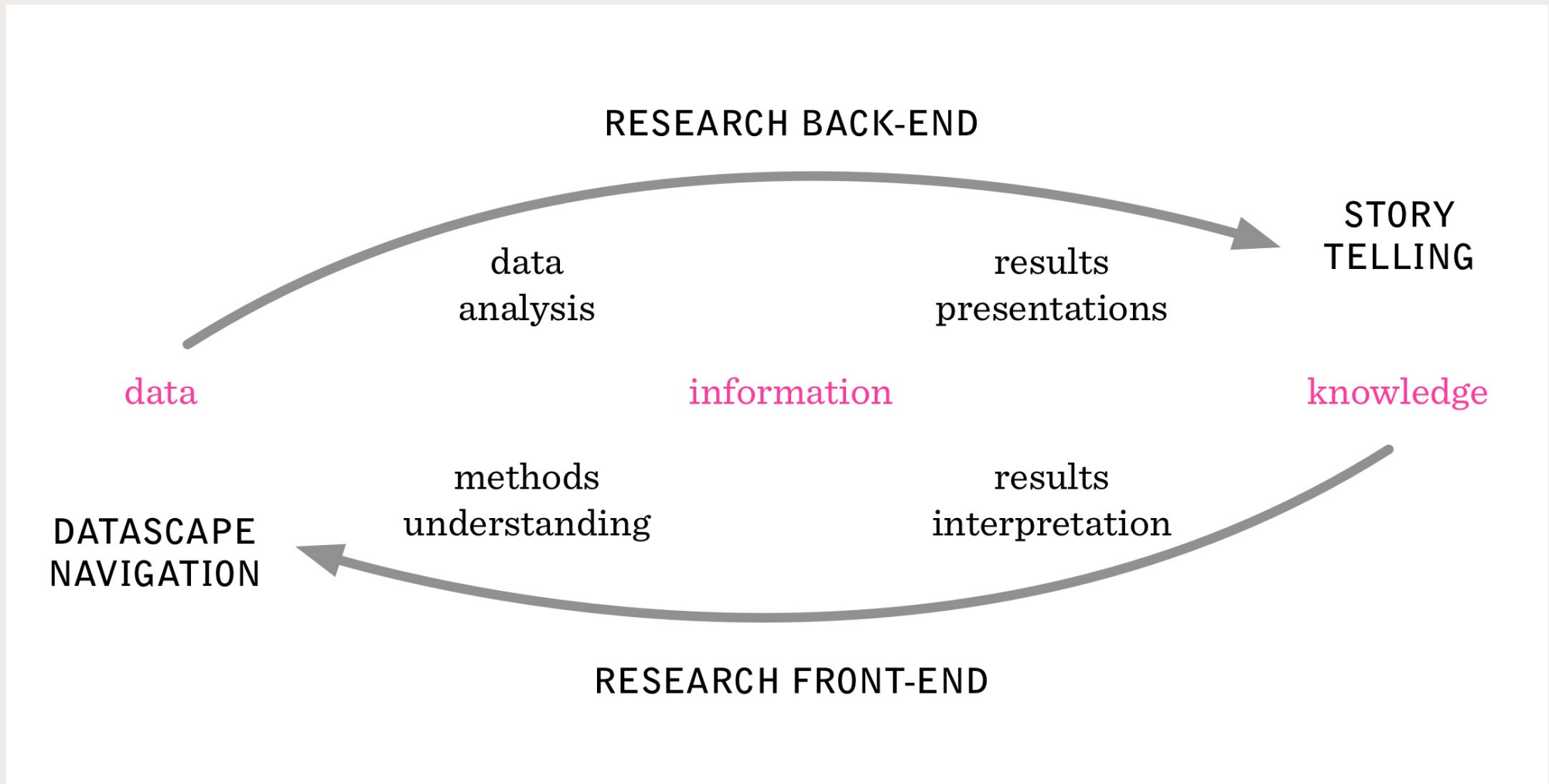
→ **Generic** collection & exploration tools

→ **Specific** extraction scripts



A Quali/Quanti approach: exploratory iterations

Digital ≠ Magic



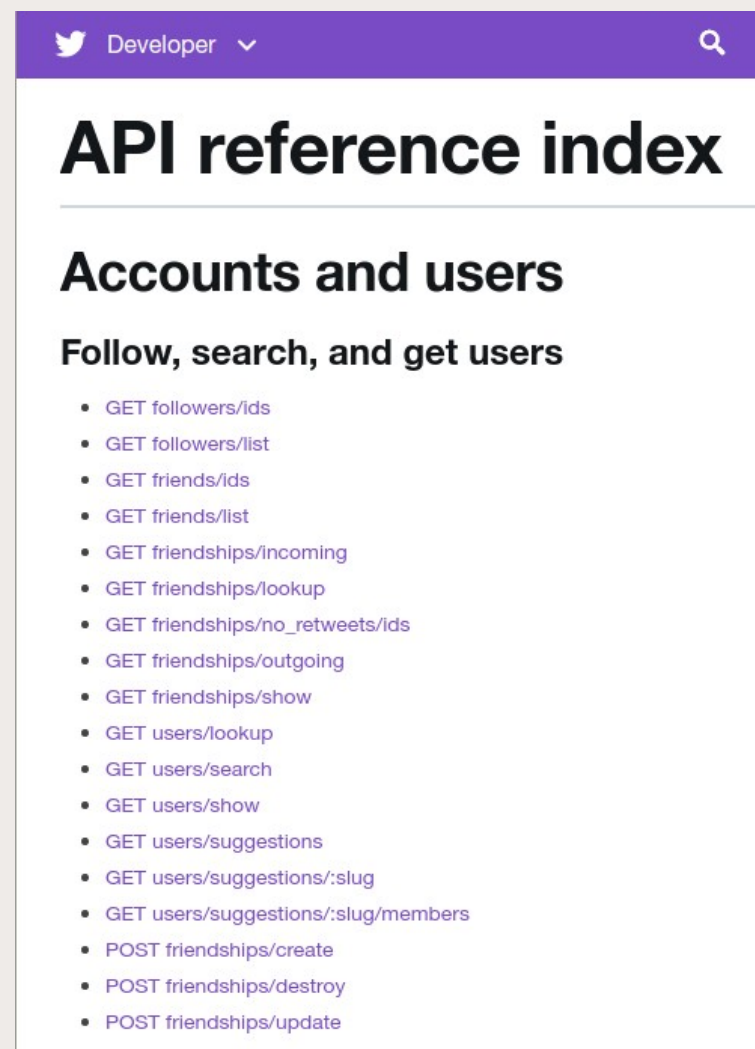
→ always avoid full automation

Research Driven Development

- Aim at large **Adoption**:
 - **build** tools fit for Social Scientists
 - **design** user centric interfaces
 - **publish** tools directly usable online
- Ensure maximum **Reusability**:
 - seize **opportunities** to implement new functionalities
 - Free Libre **Open Source** Software
(downloadable, installable, editable, reviewable, improvable)

Playground: working with platforms' APIs

- « Application Programming Interfaces »
- Structured data
- Massive and/or complete
- User behavioral data
- Problems:
 - volume
 - rate limits
 - black boxes
 - platforms' perspectives



CatWalk: qualitatively pick tweets from a corpus

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

The screenshot displays the CatWalk web interface. At the top left, it says "CATWALK". To its right are navigation buttons: "prev", "0", and "next". Further right is a "Download" button with a counter showing "0" in a red circle, "434" in a grey circle, and "2" in a green circle. Below this is a green bar with the word "IN" in white. The main content area shows a tweet from "RE•WORK @teamrework" with a "Follow" button. The tweet text is "Inside OpenAI, Elon Musk's Wild Plan to Set Artificial Intelligence Free ow.ly/4nfo2S #AI @open_ai" and is dated "7:15 PM - 29 Apr 2016". It includes a photo of Elon Musk and a link to a Wired article. Below the tweet are icons for reply, retweet (7), and like (15). On the right side, there is a vertical menu with buttons: "previous" (up arrow), "next" (down arrow), "IN" (green circle with right arrow), "OUT" (red circle with left arrow), "UNDECIDED" (grey circle with up arrow), and "save" (square icon).

« *Tinder-like* » review of all tweets listed in a CSV to decide to take them IN or OUT

SeeAlsology: semantic exploration from Wikipedia

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

Build & explore a semantic network of fields linked within « See Also » sections of Wikipedia pages

Humanités numériques [modifier | modifier le code]

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.

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.
- Phlcarto** 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

Paste your list of wikipedia articles here or [try an example](#)

https://fr.wikipedia.org/wiki/Humanit%C3%A9s_num%C3%A9riques

Stop words (press enter or separate the works with a comma)

Wikipedia: x Category: x File: x wikisource: x Commons: x
 liste d x index d x catégories d x portail x désambiguisation x
 résumé d x Catégorie: x Fichier: x add a word and press Enter

Distance Parent links

START CRAWLING **DOWNLOAD** **CLEAR CACHE**

Click a node to visit it on Wikipedia

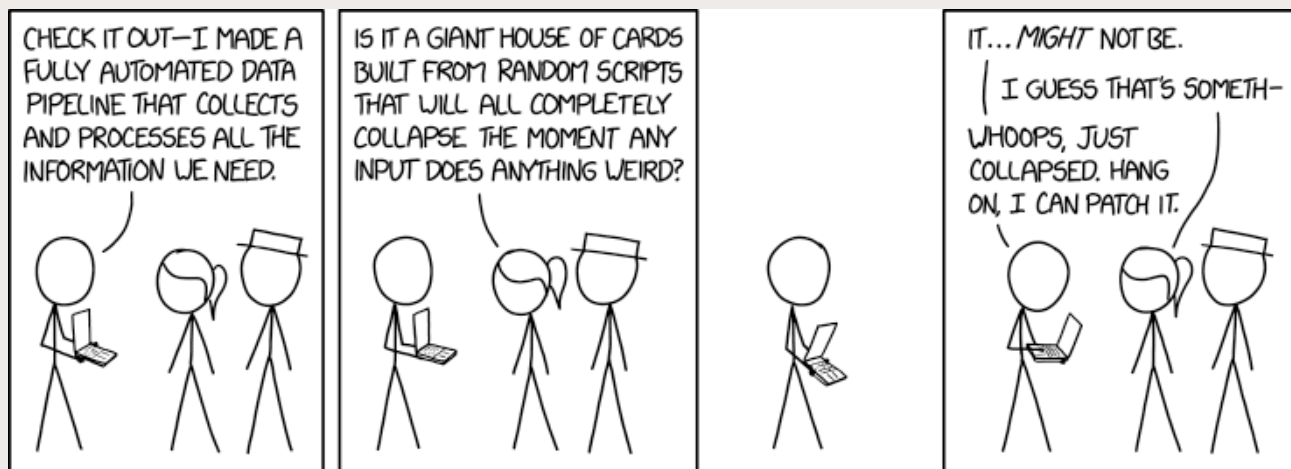
● seeds ● level -1 ● level 0 ● level 1 ● level 2

Ctrl+Click a node to add it to the seeds

APIs are rare: data Scraping to the rescue

- Extract information from a specific single source
- Build structured data from all kinds of web contents
url, date, author, image, description, place, values...
→ Enable statistics & quantitative analysis
- Whatever is public & formatted can be scraped
- Problem: The web constantly changes

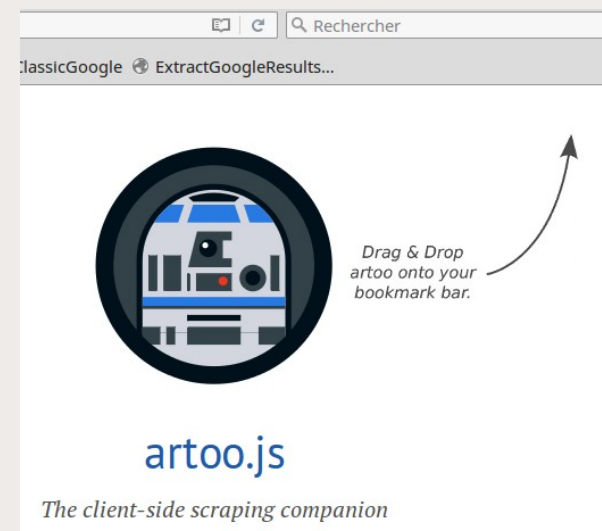
→ Maintenance issues



Artoo.js: the browser scraping companion

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

- Common scraping traps:
 - user authentication
 - cookies
 - dynamic pages with JS...
- Avoid simulating browsers: hack them!
- Embed jQuery helpers to scrap directly from within the browser's console (F12)



```
> 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
```

Google bookmarklets: search results as CSV

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

The image is a collage of screenshots illustrating the workflow of using Google bookmarklets to scrape search results. At the top left, a screenshot shows the 'Install Google Bookmarklets' page, which instructs users to drag and drop icons into their browser's bookmark bar. A red dashed box highlights the Google logo icon, with an arrow pointing to a search page. The top right screenshot shows a Google search for 'digital humanities' with a 'Redirect to Classic Google' popup. This popup allows users to select a language (currently 'en') and the number of results per page (currently '100'). It also displays the URL to which the user will be redirected. The bottom left screenshot shows the same search results page with an 'Extract Classic Google Results' popup. This popup shows the search query 'digital humanities', the page number (0), and the total number of results (103). It offers a button to 'Keep existing results & continue to the next page' and a button to 'Download CSV with 103 urls'.

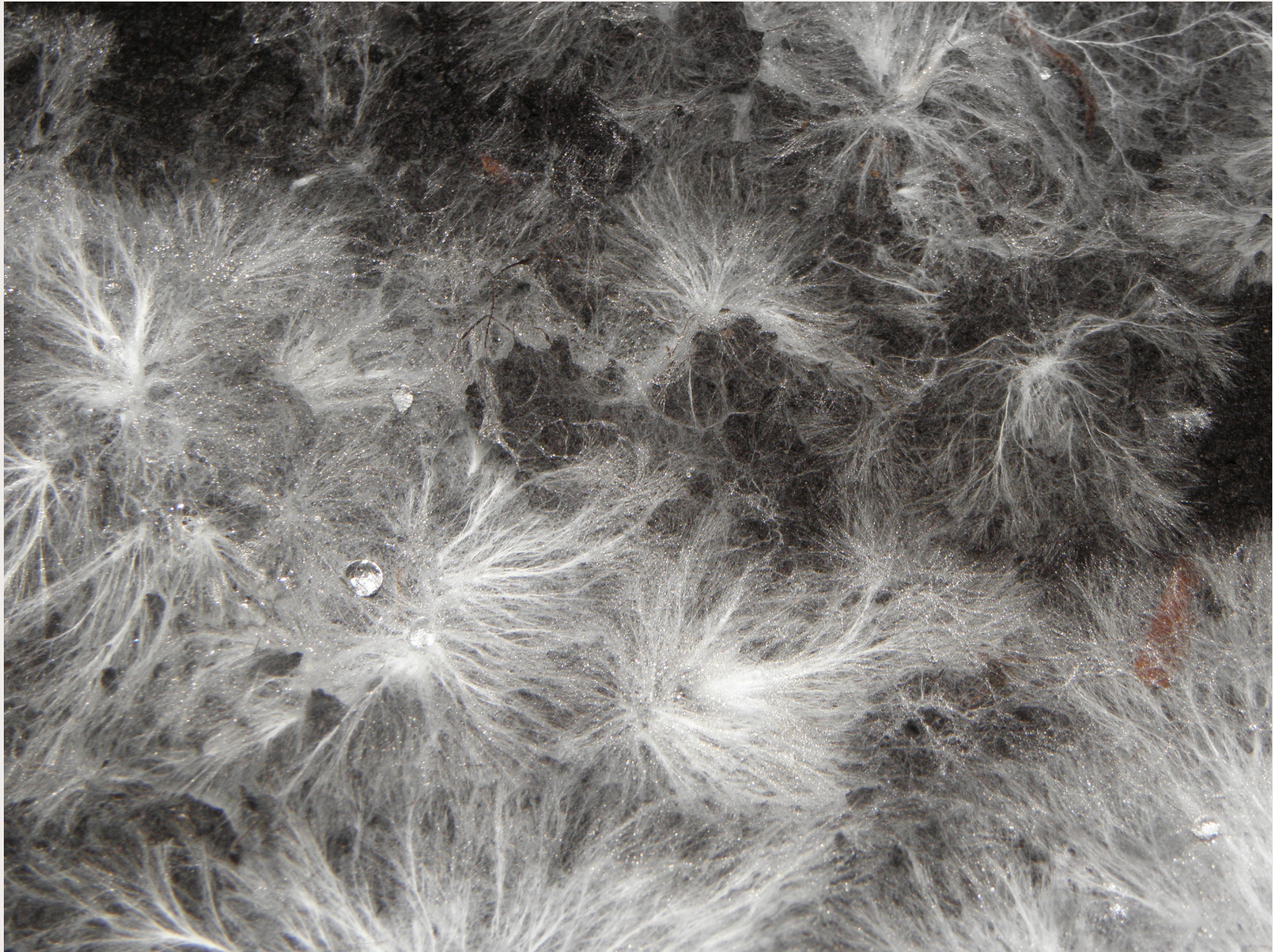
Simple buttons installable into the browser's bookmarks to let one easily scrap search results as tabular data

Crawling: harvest data from a variety of sources

- Extract coherent data from heterogeneous sources
- Build loosely structured data from common grounds found within web documents (pages)
 - text content → natural language processing
 - hyperlinks → network analysis
- Problem: the web is a « dirty mess »
redirections, erroneous links, dead links,
disappeared websites, bad encoding...

1. From the web to platforms:
a (not so) brief history of hyperlinks
2. médialab: enabling digital field work
through design & engineering
3. Hyphe: curate hypertexts into
web corpora
4. Many angles of hypertext studies

The interlinked mycellium: a network of hyphæ



[CC-BY-SA - Rob Hille on Wikimedia Commons](#)

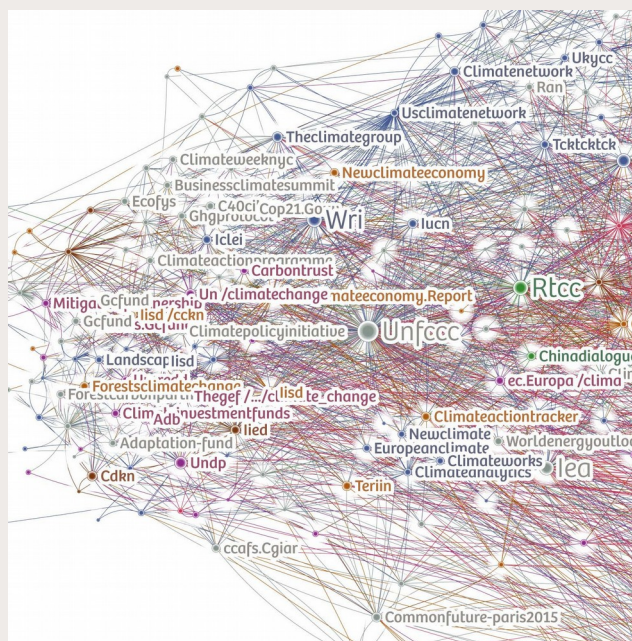
Hyphe: a research directed crawler

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

Build your own web documents corpus
to study social phenomena online

→ gather « web actors »

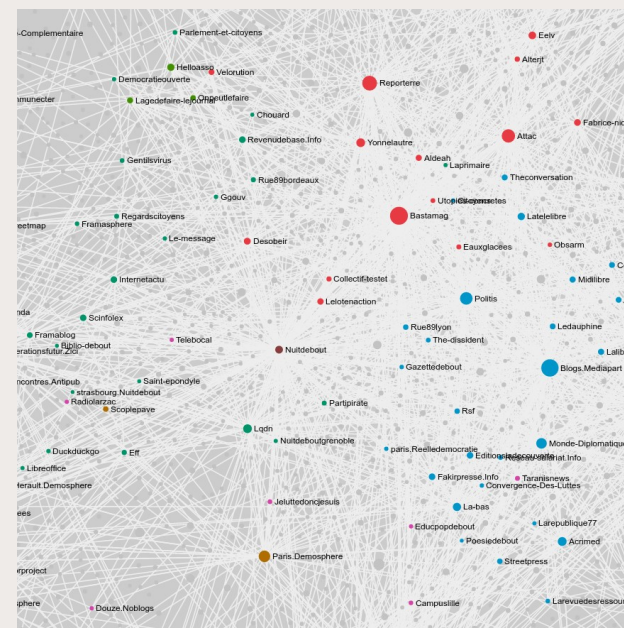
→ explore hyperlinks between them



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

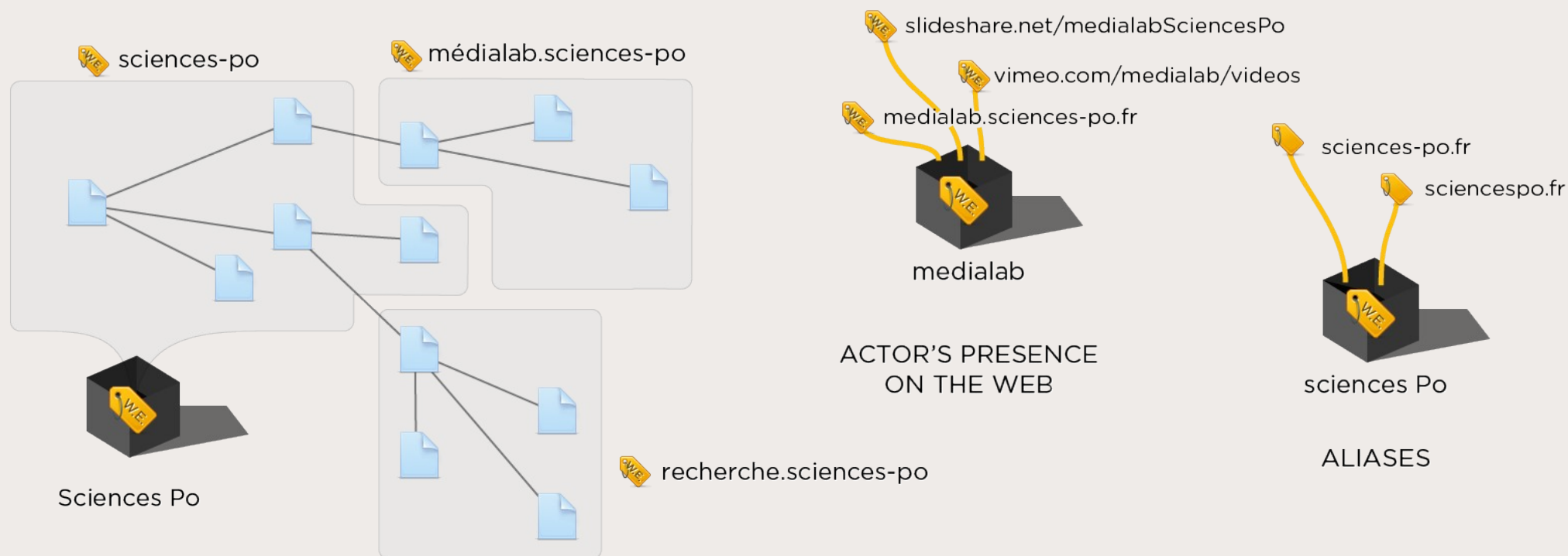
COP 21
Vie privée
Extrême droite
Tissu associatif
Produits laitiers
Cellules souches
Administrations culturelles

...



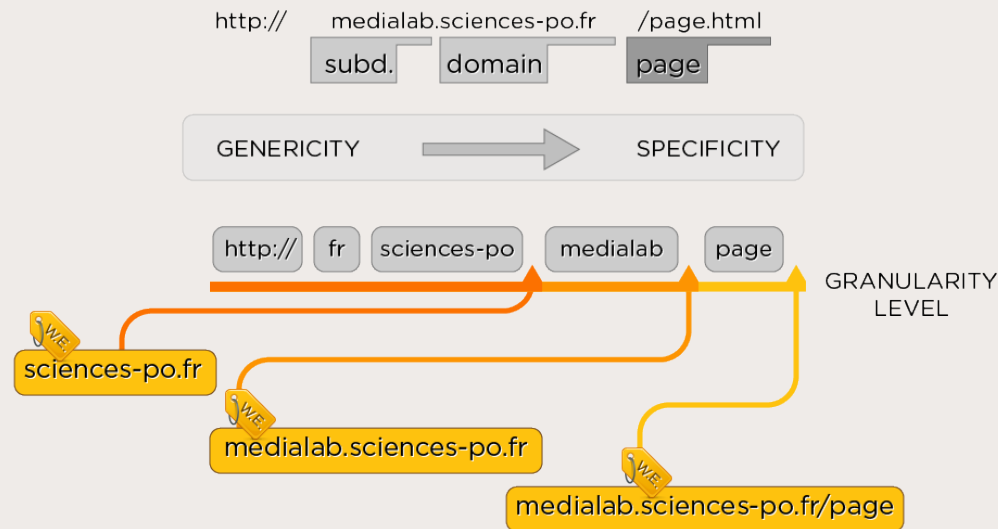
<http://utopies-concretes.org/>

There is no such thing as a « website » !



→ « **WebEntities** » : bundles of webpages aggregating coherent actors to answer a specific research question
= set of URL prefixes

Finely delimit the web territories of actors



Manually setup prefix patterns to adjust the cursor of « WebEntities »

DEFINE WEB ENTITIES

Check the boundaries of each web entity before creating it

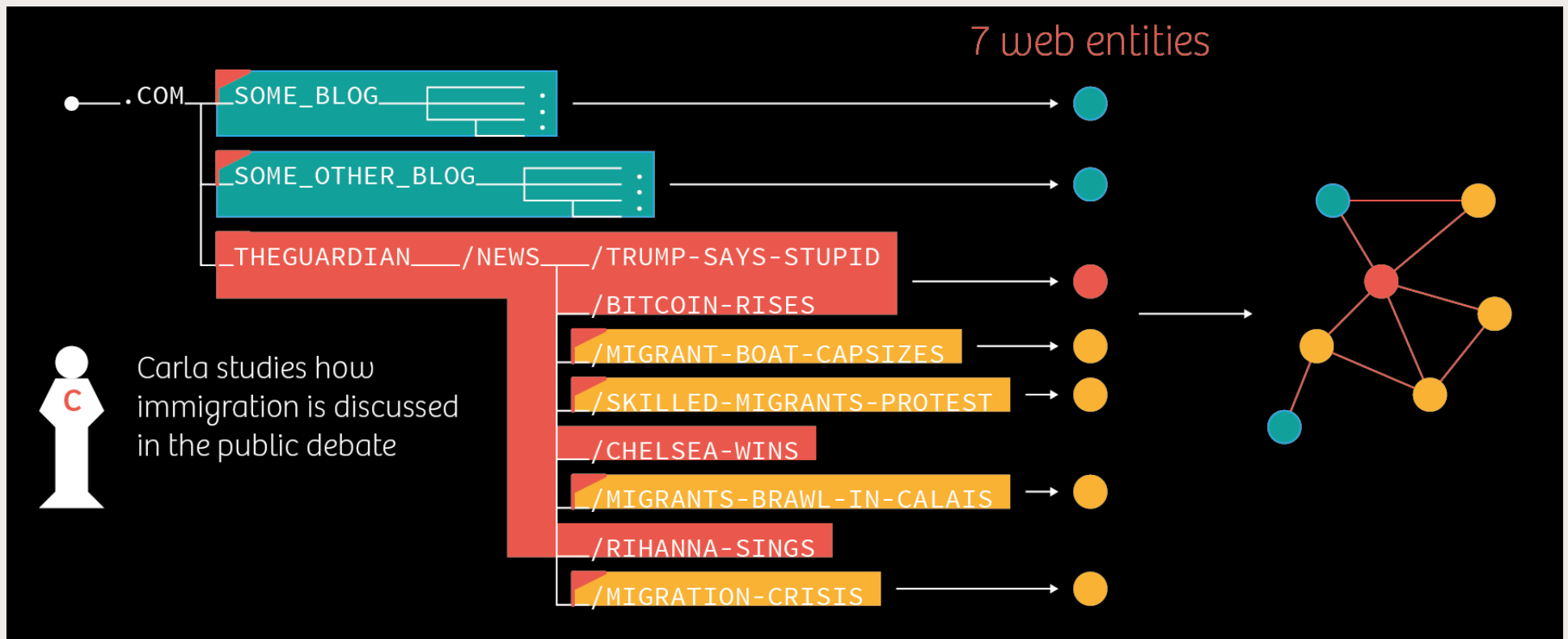
Move all sliders [TO THE LEFT](#) [TO THE RIGHT](#)

1	medialab.Sciences-Po.fr	http .fr sciences-po medialab.
2	tools.medialab.Sciences-Po.fr	http .fr sciences-po medialab. tools.
3	Sciences-Po.fr	https .fr sciences-po www.
4	Sciencespo.fr/bibliotheque	http .fr sciencespo www. /bibliotheque
5	Twitter.com /medialab_ScPo	https .com twitter /medialab_ScPo

Which data structure to manage hypertexts?

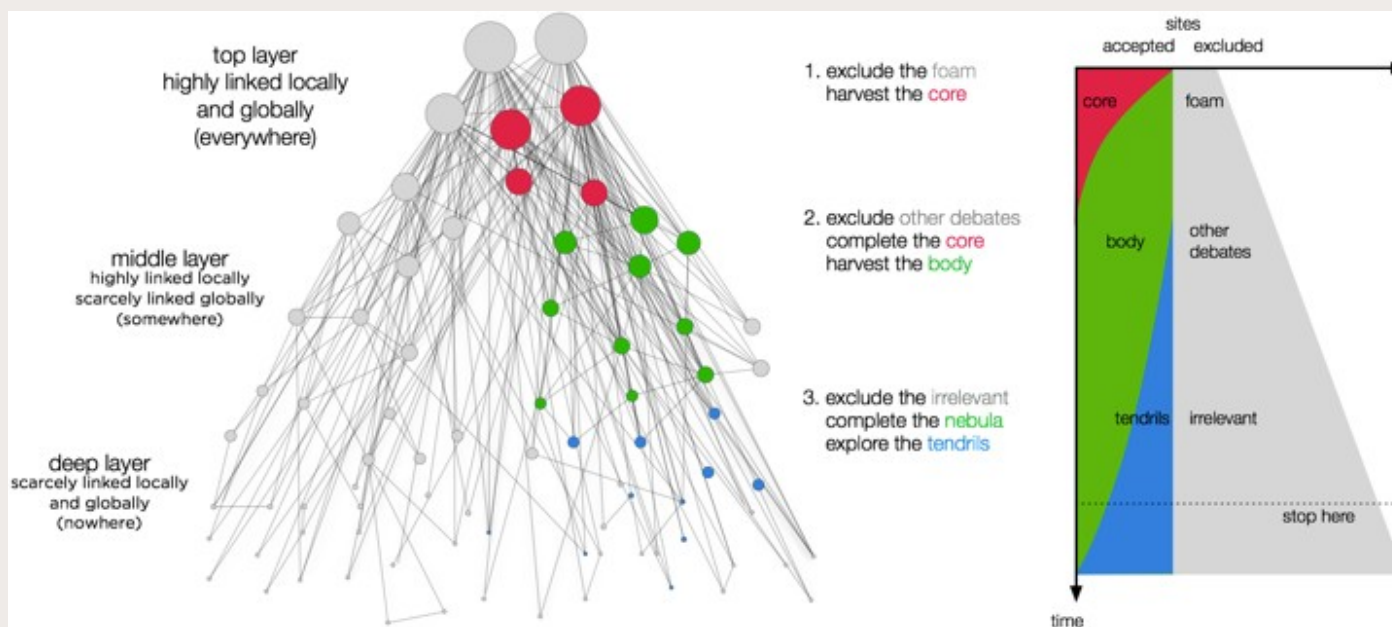
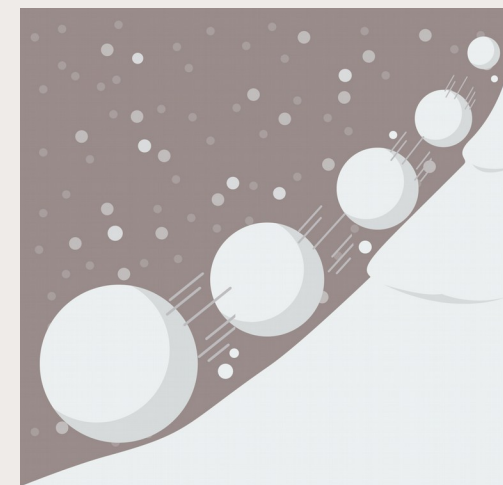
<https://medialab.github.io/hyphe-traph/fosdem2018/#/>

- **Tree** of urls
 - **Graph** of hyperlinks
 - Dynamic branches aggregates
- Hyphe's Traph



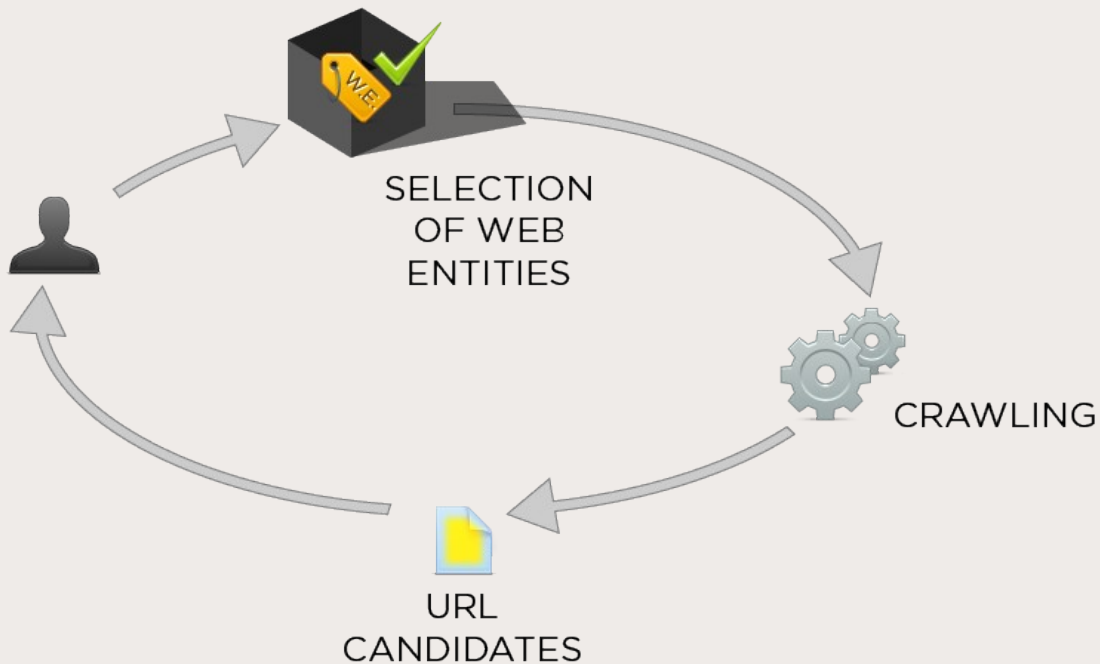
Hyphe's crawling strategy: leverage hyperlinks

- Classical crawlers (**DMI's IssueCrawler**): Snowball
 - Top Layer attraction (Google, YouTube, Wikipedia...)
 - Topic drifts
- Hyphe:
 - crawl exclusively pages within the chosen WebEntities
 - sort discovered entities by degree of citation
 - humanly select new entities to include and crawl



Web prospection loop: curating a corpus iteratively

- Step by step iterative expansion & curation of entities



- Human/Time cost
- How to know when to stop?
→ hyperlink citations threshold

PROSPECT 4,890 DISCOVERED

Search APPLY CHANGES CANCEL

Distribution of citations (log scale)

NAME	CITED ↑
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Google.fr	23
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Instagram.com	19
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Free.fr	16
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Wordpress.org	16
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Wp.com	13
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Blogger.com	12
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Twitter.com /home	12
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Gravatar.com	11
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Legifrance.gouv.fr	10
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Wordpress.com	10
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Collectifmarianne.fr	9
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Collectifracine.fr	9

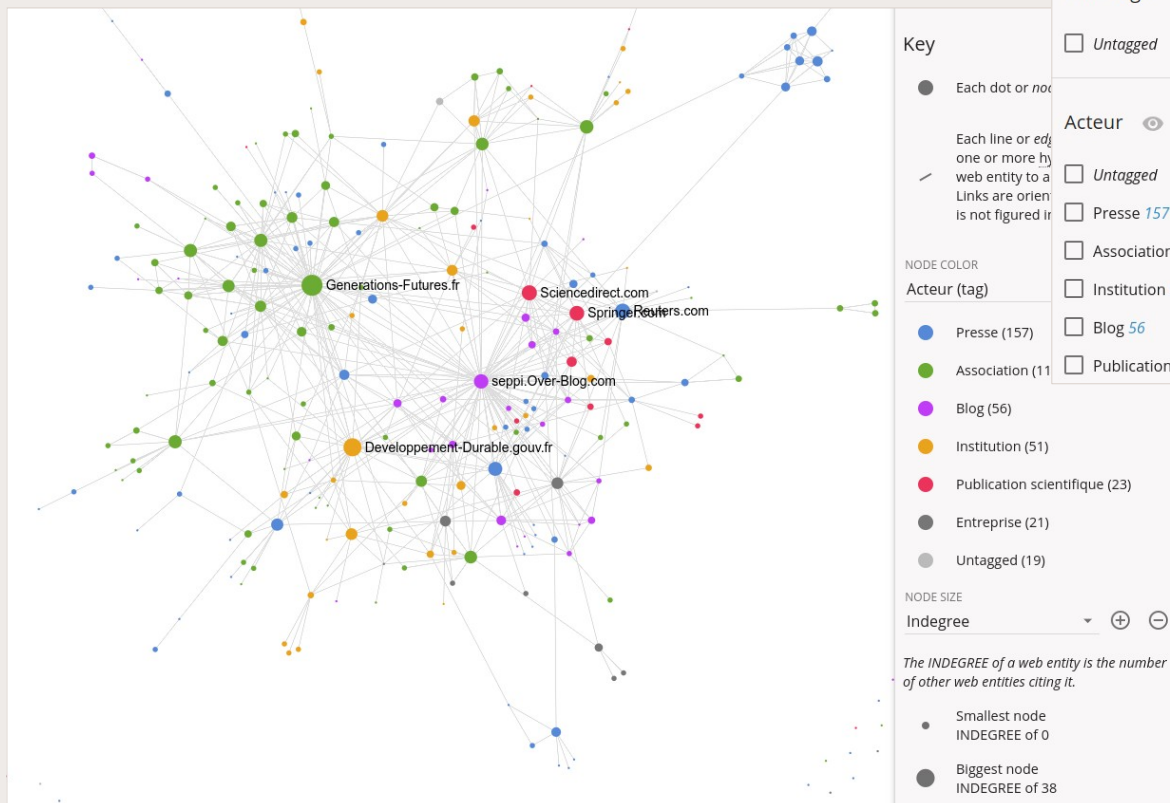
1 SET TO IN
Collectifmarianne... X

1 SET TO UNDECIDED
Legifrance.gouv.fr X

4 SET TO OUT
Gravatar.com X
Google.fr X

Qualify the corpus entities (tagging)

- Free notes
- Categories



TAGS

Filter [web entities](#) (status *IN* only). Tag one or a selection of web entities.

439
WEB ENTITIES

TAG FILTERS 439 WEB ENTITIES WEB ENTITIES NETWORK

Special filters

- Untagged
- Partially untagged
- Conflicts

Free Tags

- Untagged

Acteur

- Untagged
- Presse 157
- Association 111
- Institution 51
- Blog 56
- Publication scientifique 23

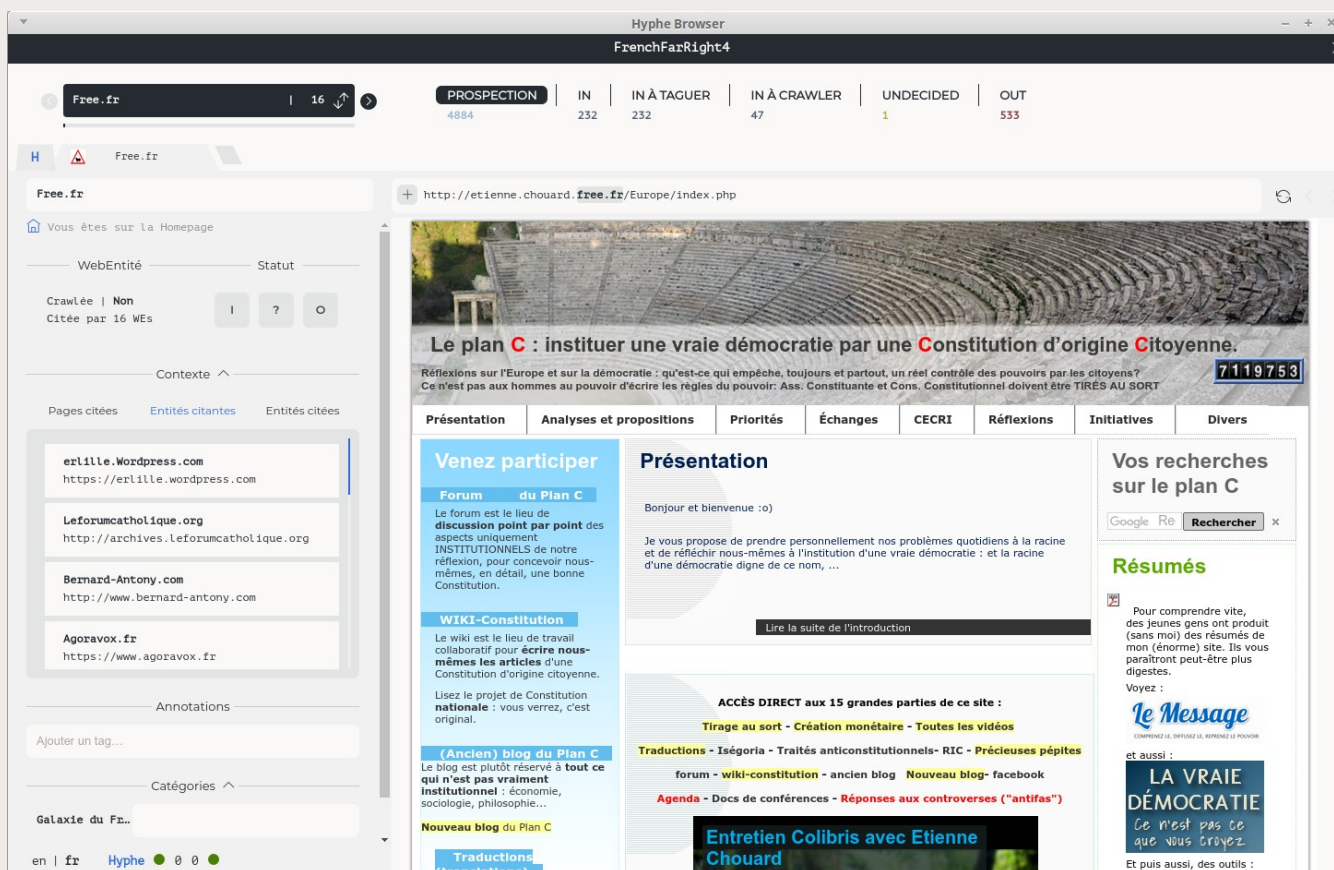
Display a category
Point de vue

Search

- Futura-Sciences.com /.../biologie-pesticide-9169 Neutre
- Lefigaro.fr /.../37002-20170627ARTFIG00002-pesticidepe-sti-sid-n-m-... Neutre
- Parents.fr /.../pesticides-et-grossesse-des-risques-confi... Contre les pesticides
- formulaires.Fondation-Nicolas-Hulot.org /.../stop_pestic... Contre les pesticides
- Contrepoints.org /.../270496-pesticides-lintox-discours-bio Pour les pesticides
- Observatoire-Pesticides.gouv.fr Neutre
- Letemps.ch /.../toxicite-pesticides-tueurs-dabeilles-confirmee-terrain Neutre
- Sciencepresse.qc.ca /.../neonicotinoides-pesticides-tue... Contre les pesticides
- Notre-Planete.info /.../4613-liste-fruits-legumes-pesticides Neutre
- Lepoint.fr /.../pesticides-tueurs-d-abeilles-bayer-interpelle-par-un-mil... Neutre
- Consoglobe.com /abeilles-pesticides-bayer-cg Contre les pesticides

HyBro: a web browser designed for corpus curation

<https://github.com/medialab/hyphe-browser/releases/>



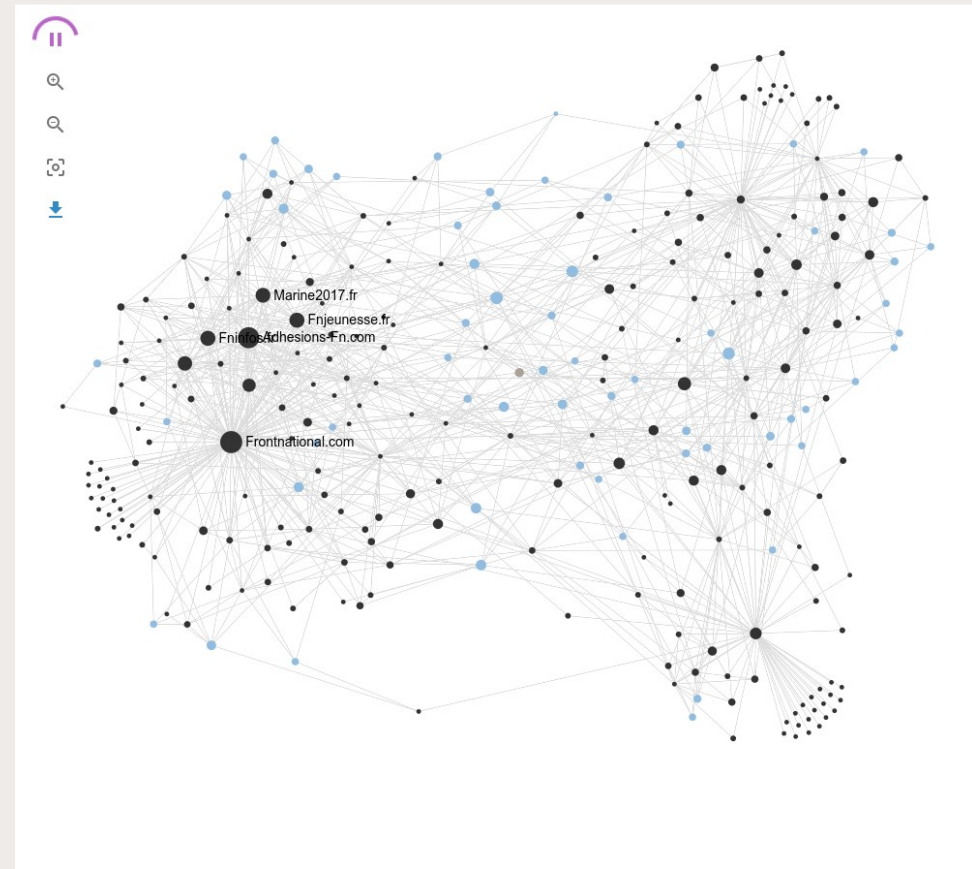
- « NaviCrawler » heritage : build a Hyphe web corpus while browsing
- « in-situ » prospection & tagging (digital field work)
- teach the web to students (IDEFI FORCCAST)

1. From the web to platforms:
a (not so) brief history of hyperlinks
2. médialab: enabling digital field work
through design & engineering
3. Hyphe: curate hypertexts into
web corpora
4. **Many angles of hypertext studies**

Many ways to use Hyphe

- Complete methodology includes:
 - sourcing, automatized collection, iterative corpus building
 - qualitative categorization, exploratory analysis,
 - network visualization, quantitative statistical analysis
- Diverse audiences:
 - Research: help social scientists work on digital fields
 - Pedagogy: teach students what the web is beyond Google & Facebook
- Possible small & large scale analyses:
 - a website's internal structure
 - a theme's ensemble of actors and their ties
 - a controversy's alliances & oppositions
 - etc.

From above: clusters, opposition & affinity



Network Viz Settings

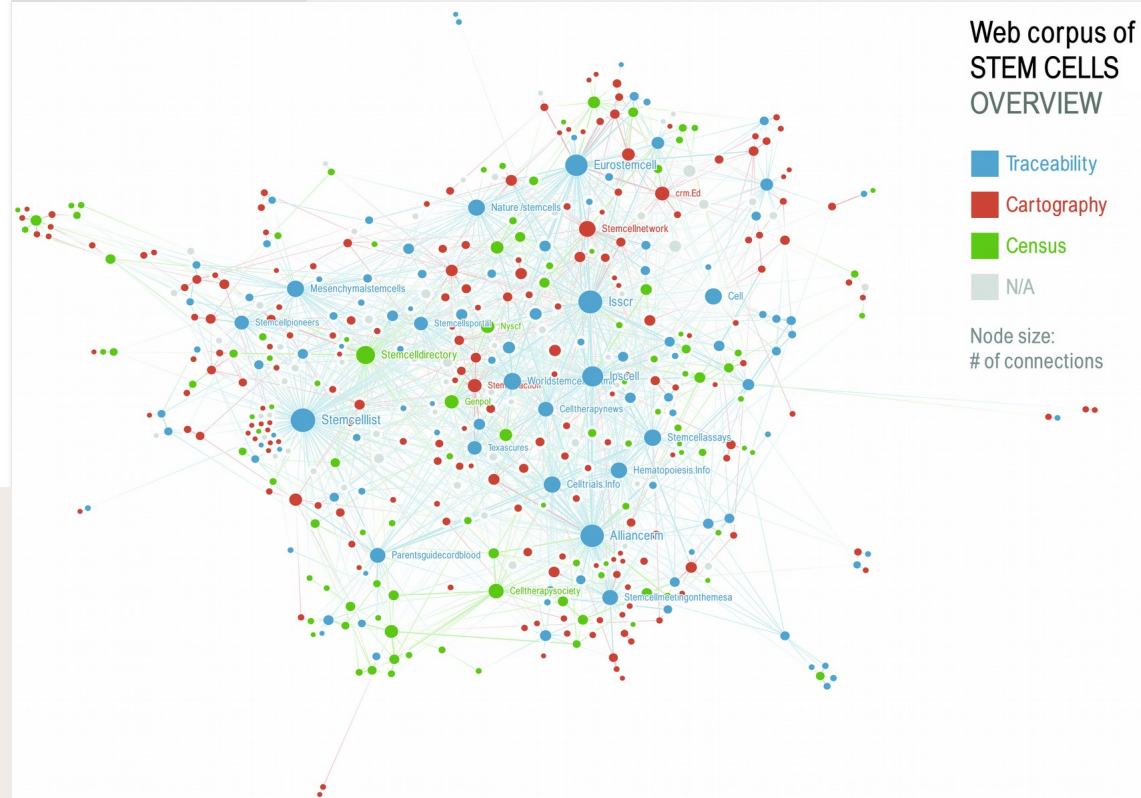
Filtering

- IN 232
- UNDECIDED 1
- OUT 533
- DISCOVERED 4,884

Filter DISCOVERED web entities

Display only DISCOVERED with ...

Filter ALL web entities



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

From within: explore webpages contents

PRIVACY WEB CORPUS

SciencesPo MÉDIALAB AXA Research Fund Data Innovation Lab

ABOUT

EXPLORE WEB ENTITIES

2,313 ENTITIES
7,549 entities represented as a cloud

Search

Q Apple FBI backdoor

nakedsecurity.Sophos
https://nakedsecurity.sophos.com/category/technologies/operating-s...
developers would rather quit than give FBI a backdoor A lead developer for the Tor Project said

Helpnetsecurity
https://www.helpnetsecurity.com/tag/backdoor/
encryption backdoors a bad idea March 4, 2016 backdoor cybercriminals encryption Apple and the FBI

nakedsecurity.Sophos
https://nakedsecurity.sophos.com/2016/03/23/tor-project-says-devel...
developers would rather quit than give FBI a backdoor 23 Mar 2016 1 Apple , iOS , Law & order , Privacy

nakedsecurity.Sophos
https://nakedsecurity.sophos.com/2016/03/23/tor-project-says-devel...
developers would rather quit than give FBI a backdoor 23 Mar 2016 1 Apple , iOS , Law & order , Privacy

Sidstamm
http://blog.sidstamm.com/2016_02_01_archive.html
their phones vulnerable is not the right approach. The current public discourse on the Apple vs. FBI "open

Laquadrature
https://mediakit.laquadrature.net/view.php?full=1&id=2374
20160219[AC] LCP Chiffrement Apple FBI Download : MP4 , WebM , How to embed ?

Laquadrature
https://mediakit.laquadrature.net/view.php?id=2374
20160219[AC] LCP Chiffrement Apple FBI Download : MP4 , WebM , How to embed ?

Laquadrature
https://mediakit.laquadrature.net/view.php?full=1&id=2374
20160219[AC] LCP Chiffrement Apple FBI Download : MP4 , WebM , How to embed ?

Topics

Surveillance FR

Business & Media

Surveillance US

Cybersecurity

Big data & Analytics

Data Regulation FR

Cookies & Tracking

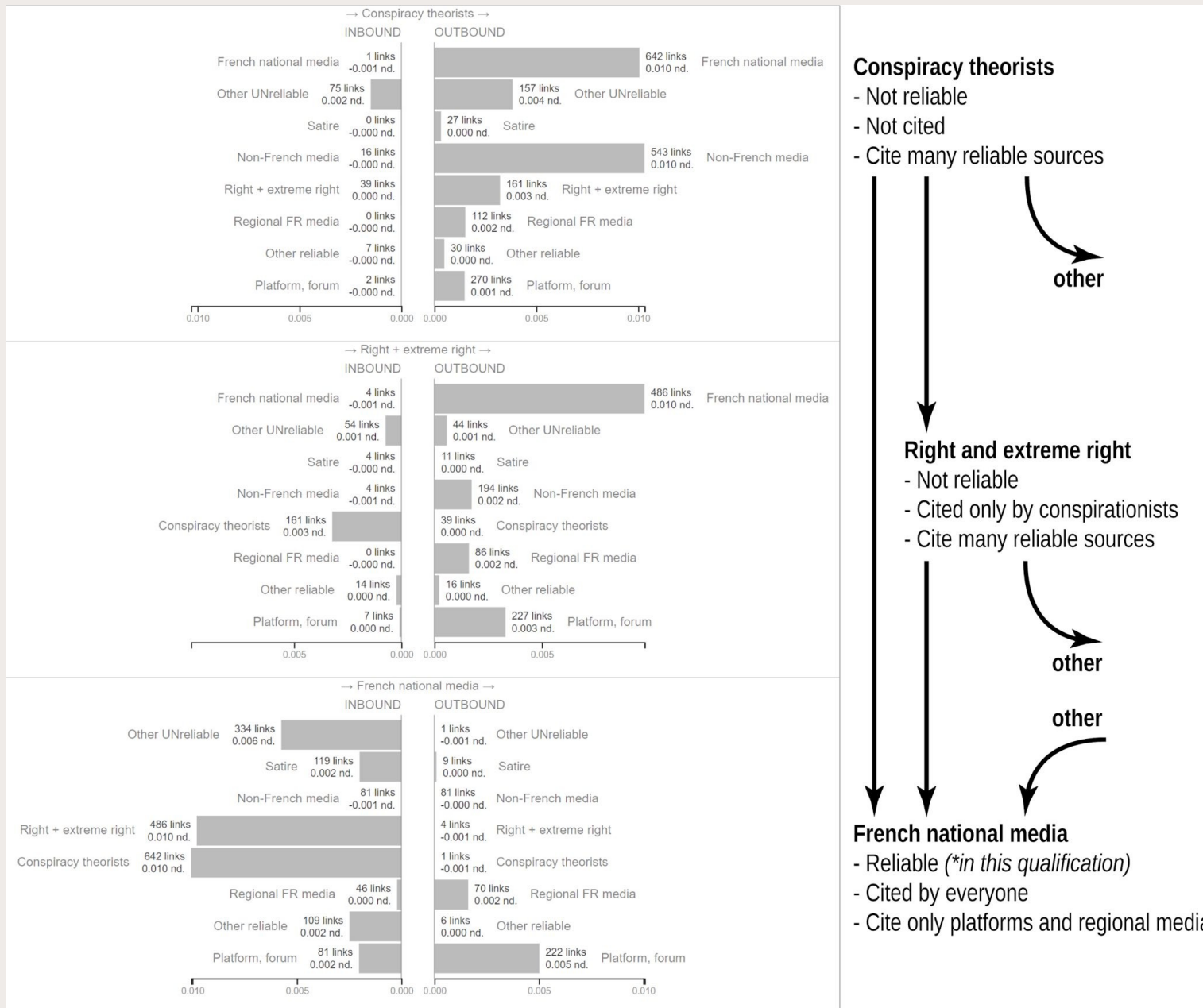
Telec Operators FR

Card and ID fraud

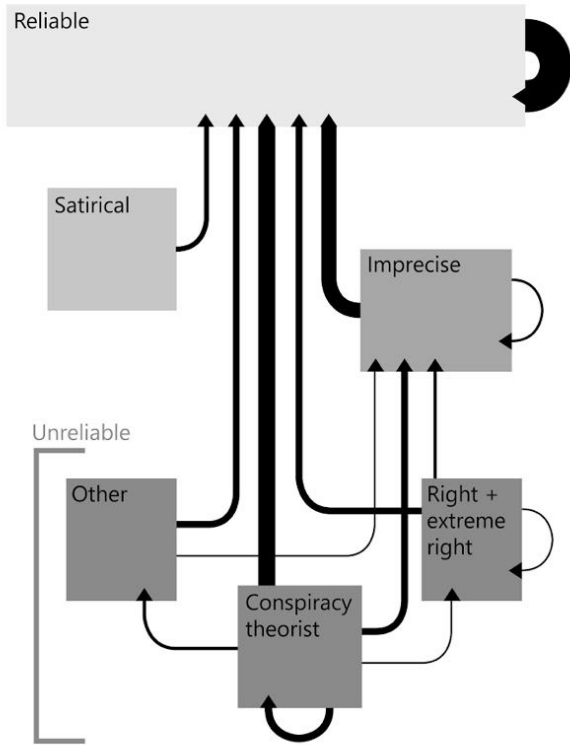
EXPLORE TOPICS

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

From the sides: a hierarchy of directed hyperlinks

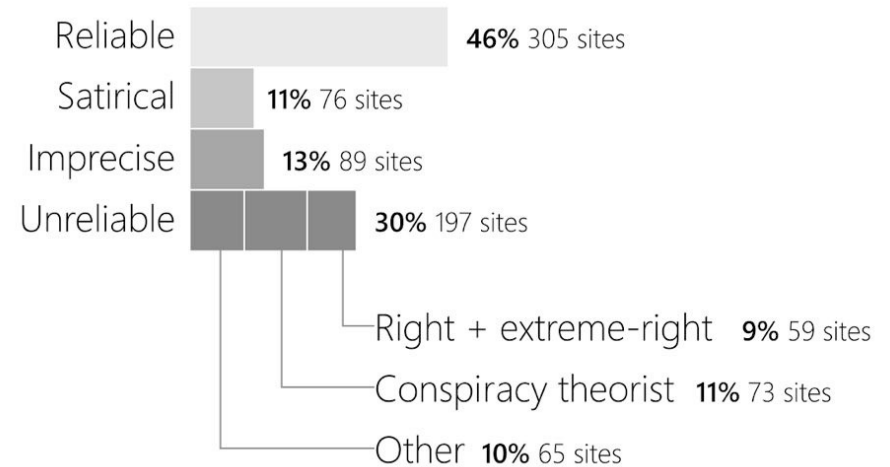


Hyperlinks directionality: a bottom-up hierarchy

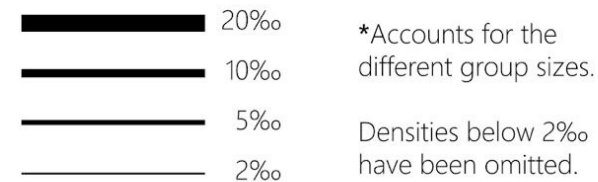


Most hyperlinks stem from the unreliable and aim at the reliable resources

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

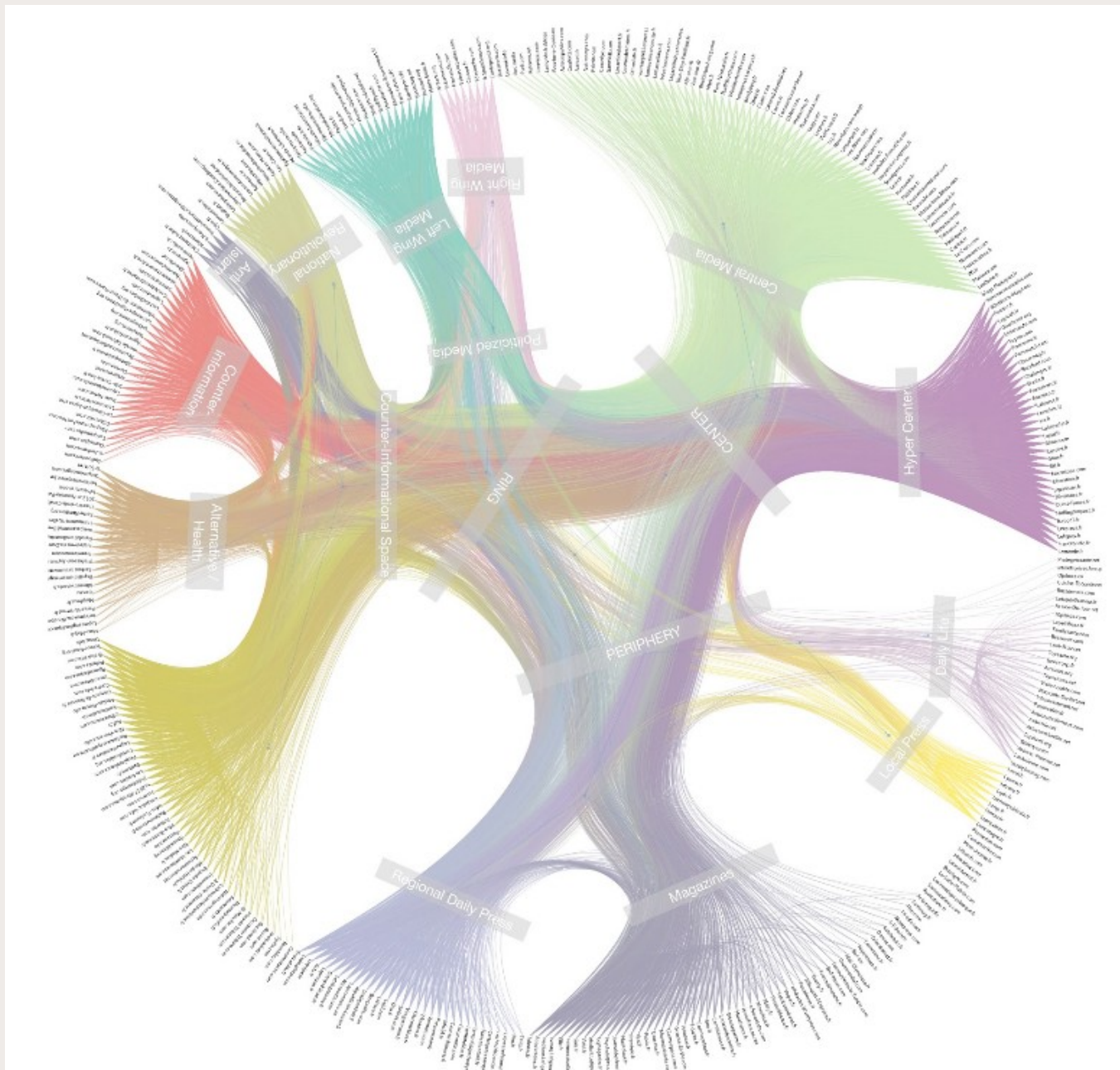


The thickness is proportional to the normalized link density*



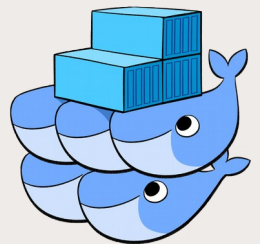
https://www.researchgate.net/publication/320225750_Visual_Network_Exploration_for_Data_Journalists

Explore polarization dynamics



Roadmap: what next?

- Import / export corpora or lists of webentities & crawls:
 - duplication, reproduction
 - longitudinal time exploration
- Integrated text processing (NLP) for content analysis
- Use modern web technologies to handle JavaScript powered contents (Facebook, React applications, etc.)
- Quality control tools for crawls
- Archival & exploration tools to publish finalized web corpora
- Propose automatic setups of Hyphe as Software As A Service



Hyperlink is not dead: long live hyperlink studies!

- 10-year financial support of EQUIPEX
→ building cutting-edge research instruments demands long-term support
- Hyphe is Free & Open Source, try it online with the demo!

<https://hyphe.medialab.sciences-po.fr/demo/>

<https://github.com/medialab/hyphe>

Questions?

benjamin.ooghe@sciencespo.fr

[@boogheta](#) [@medialab_ScPo](#)

Bibliography

Reference publications:

- Jacomy M., Venturini T., Heymann S., Bastian M. (2014), **ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software**, PLoS ONE, 9(6), 1-18
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679>
- Jacomy M., Girard P., Ooghe-Tabanou B., Venturini T. (2016), **Hyphe, a Curation-Oriented Approach to Web Crawling for the Social Sciences**, ICWSM 2016, Cologne
<https://spire.sciencespo.fr/hdl:/2441/6obemb2hsj9pboj9bbvc7sftne>
- Plique G., Jacomy M., Ooghe-Tabanou B., Girard P. (2018), **It's a Tree... It's a Graph... It's a Traph! Designing an on-file multi-level graph index for the Hyphe web crawler**, FOSDEM 2018, Bruxelles
<https://medialab.github.io/hyphe-traph/fosdem2018/#/>
- Ooghe-Tabanou B., Girard P., Jacomy M., Plique G. (2018), **Hyperlink is not dead!**, ACM Proceedings of the 2nd International Conference on Web Studies (WS.2 2018) Paris.
<http://hyphe.medialab.sciences-po.fr/docs/20181004-ACM-WebStudies-HyperlinkIsNotDead.pdf>