Thematic and spatial analysis of technologies using CorText Manager and RISIS patent database

RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

RISIS training sessions

Lionel VILLARD

LISIS, IFRIS, INRAE, CorText, ESIEE Paris

CorText Manager

CorText platform

Build a scientific and technical platform, with its specific position in order to support a **research area on infrastructures, traces and digital uses of science and innovation in society**.

Marc Barbier
Member of CorText platform, Researcher at LISIS, Head of IFRIS

Antoine Schoen
Member of CorText platform, Researcher at LISIS, Senior lecturer at ESIEE Paris

Lionel Villard
Head of CorText platform, Researcher at LISIS, lecturer at ESIEE Paris

Patricia Laurens
Member of CorText platform, Researcher at CNRS and LISIS

Philippe Breucker
IT engineer from INRAE, LISIS. Technical Director of the Cortext Digital Platform. Web Designer and developer.

Bilel Benbouzid
Researcher, Senior lecturer at LISIS

Alexandre Hannud Abdo
Post-doctorant, LISIS

Pierre-Yves Bulot
IT Engineer Assistant, Cortext

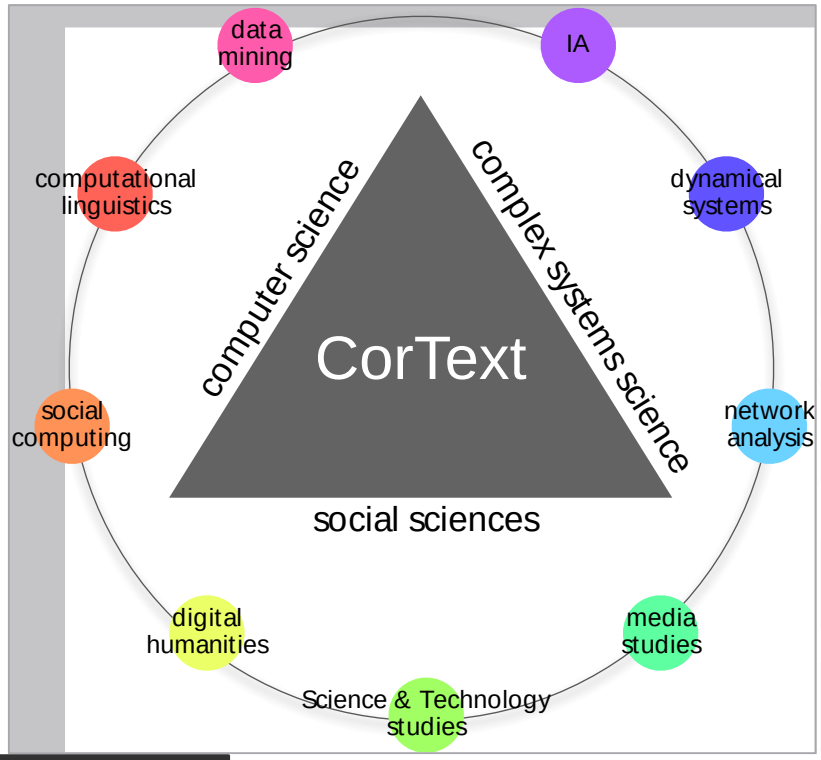
Luis-Daniel Medina
IT Engineer, Cortext

Diego-Fernando Gómez Peña
IT Engineer, Cortext

Tatiana Andrea Sánchez Castaño
IT Engineer, Cortext

Joenio Marques da Costa
Research Software Engineer, Cortext

Géraldine Enderli
Engineer specialised in the production, processing and analysis of data and survey at INRAE - LISIS, CorText



Antoine Mazières
Research scientist in the Computation Social Science team at Centre Marc Bloch

Constance De Quatrebarbes
Fondateur Présidente - DRISS (Digital Research In Science & Society)

Chloé Duloquin
Web Designer, Graphiste, Intégratrice web

Jean-Philippe Cointet
Associate Professor, Sciences Po Paris, Medialab

Guillaume Orsal
Computer engineer, data mining, web development and SEO

Cristian Martinez
PhD Engineer in Computer Science, NLP/Data Senior Consultant at Cogniteva

Nicolas Turenne
Assistant professor in data science, Beijing Normal University & Hong Kong Baptist University United International College

Tam Kien Duong
Data & design, Etalab

Nicolas Baya-Laffite
STSLab, Université de Lausanne

Loic Boudoulec
IT Engineer

Bertha Brenes
IT Engineer, Trainee, Cortext

Anis Arabi
Big data engineer

Nicolas Ricci
Web developper and data

Audrey Baneyx
Project manager Data science, Sciences Po Paris - médialab

Andrei Mogoutov
Bullescence

Élise Tancoigne
Researcher, University of Geneva, Switzerland

CorText Manager in 2021

Between October 2020 and September 2021

- 1,200 active users, generating more than 63,000 calculations;
- From around 450 institutions (universities, companies, ministries, firms, intelligence services, etc.) and cities.

CorText Manager uses for publication

- Since 2016, more than 488 authors have published using CorTexT Platform; which represents less than 10% of the community of users: https://www.cortext.net/publications/
- Increase of users from Asia (Wuhan et Manille) and Brazil (growth rate +200% during Covid)

The most frequent topics of Users

Emergence of research and innovation fields (bioenergy, nanotechnology, biodiversity,…);

Controversies and **hot topics** (pesticides, synthetic biology, global food security);

Socio-semantic and relational **mapping of research output** (publications, patents, projects)

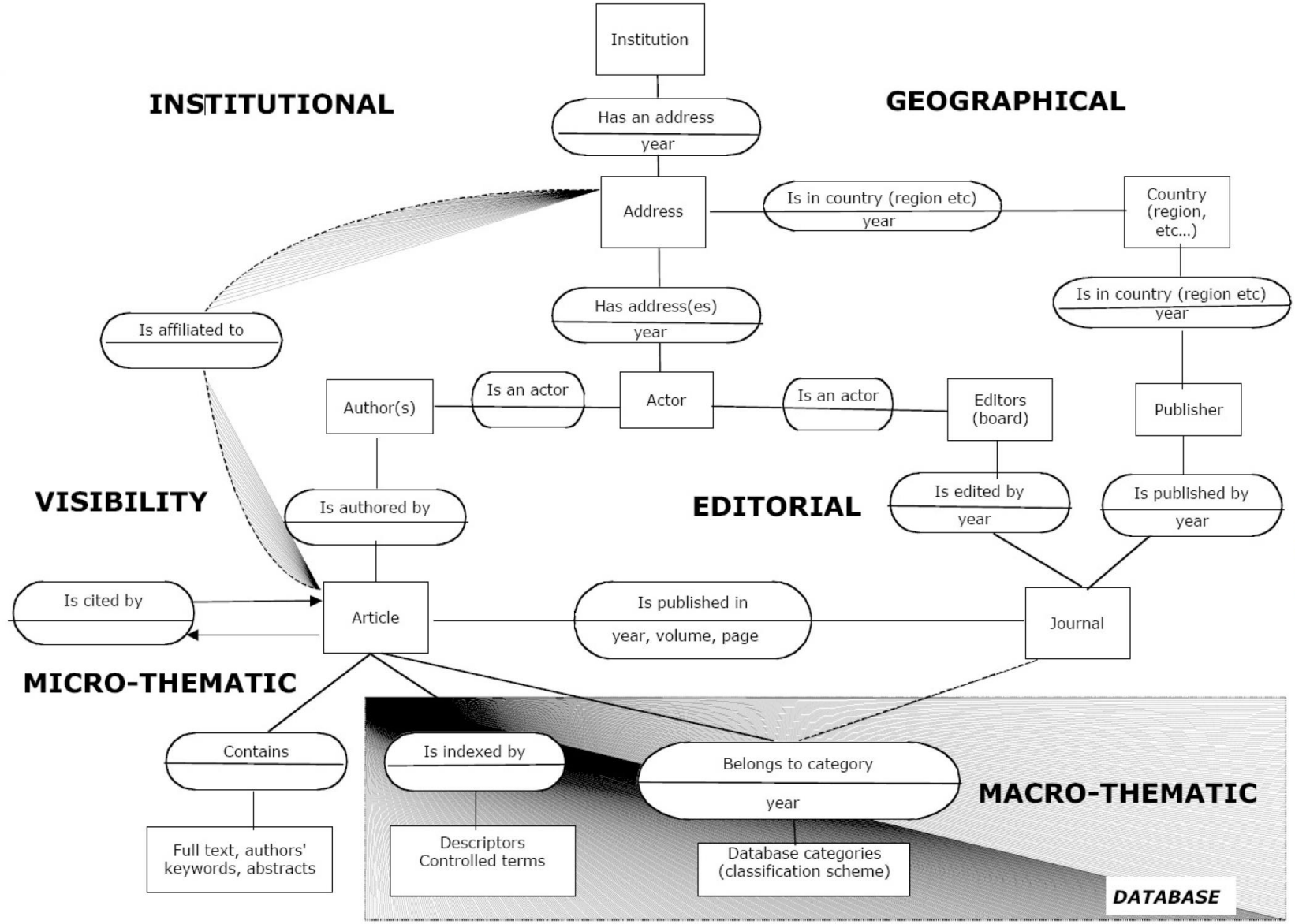
Analysis of the **web and social media** to characterise the digital turn and its phenomenology (US election, Yeallow Jacked, Youtube ecosystem).

- Two other types of uses:
  - Support for qualitative approaches: maps are used as an effective medium to stimulate and feed the interaction with the actors;
  - Literature review: some authors use CorText Manager to define and situate their work (pdh, in a paper)

→ https://docs.cortext.net/trainings/risis-cortext-patents-2022/session1-background/session1b/03-examples-papers/

The scientometric roots of CorTexT Manager

Scientific articles and bibliographic notices as hub of information



Zitt M. 2004

What CorText Manager produces

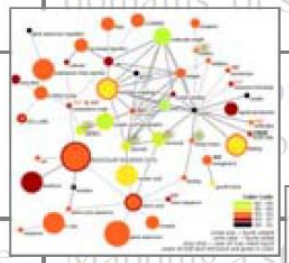
From co-citation analysis (Callon et al., 1993), to coword analysis and beyond (CorText Manager, 2021).

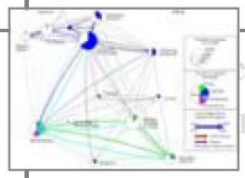
Two main types of indicators of scientific activity:

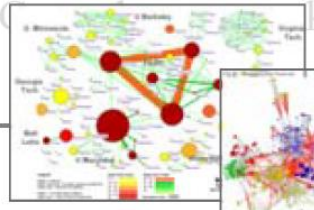
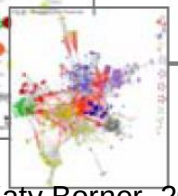
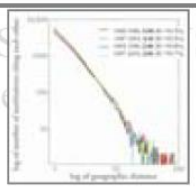
- Basic indicators and simple descriptive statistiques: stocks, ranks, frequencies;
- **Relational indicators** (networks) with, also, two main types of relational indicators:
  - **direct**: relationships do not enter directly into the content of the documents (eg: collaborations of author names);
  - **indirect**: relationships built from an analysis of the content of documents: 1/words from titles, abstract 2/ Geographical coordinates from author addresses …

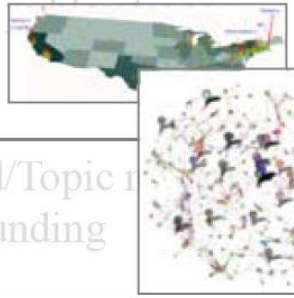
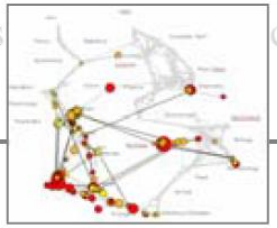
Mining relationnal information in scientific publications and bib. notices

	Micro/Individual (1-100 records)	*Meso/Local (101–10,000 records)*	*Macro/Global (10,000 < records)*
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NS SA, all of sci
Temporal Analysis (When)	Funding portfolio of one individual	ic bursts f PNAS	113 Years of P Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a s intellectual l	PNAS
Topical Analysis (What)	s.	flows in research	VxOrd/Topic m NIH funding
Network Analysis (With Whom?)	NSF work of one	k	NIH's cy





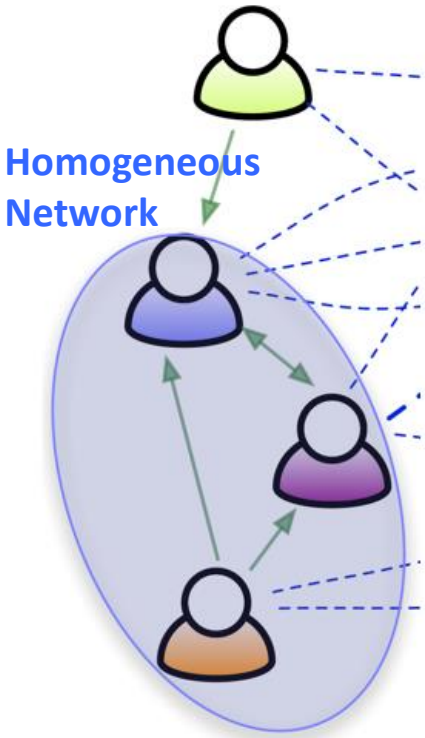




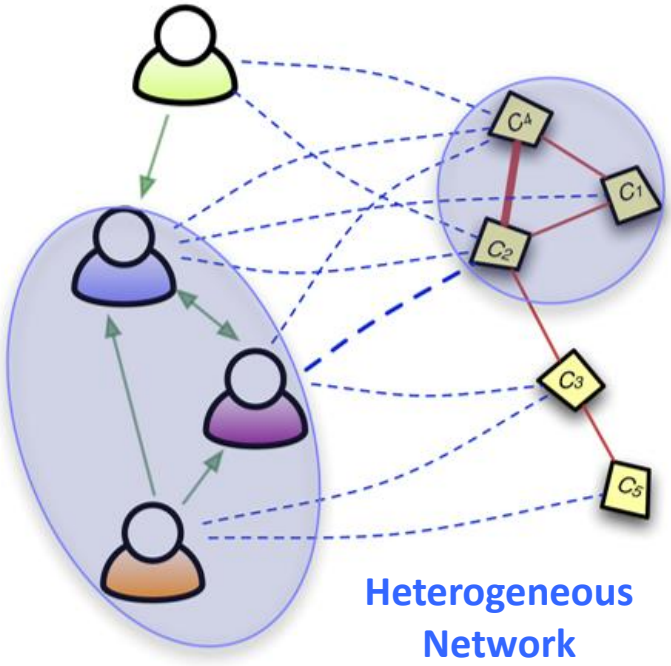
Atlas of science - Visualizing What We Know, Katy Borner, 2010, The MIT Press, 272 p.

Mining, Mapping, and Acceleration Science and Technology, Katy Borner, 2012, Sciences Po, paris

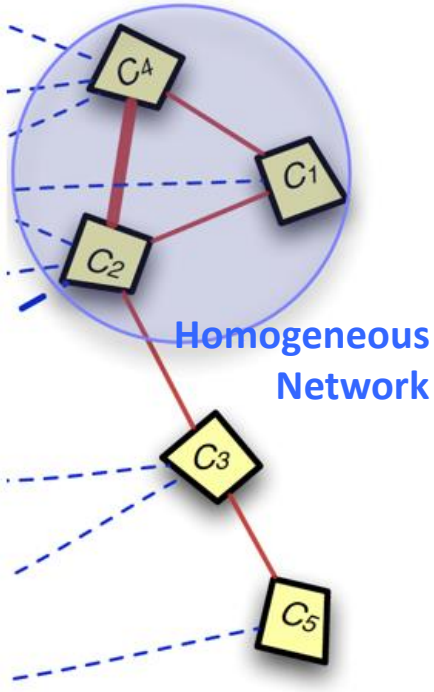
Combine direct and indirect relations within metrics and visualisations methods



Humans have relationships: sociological networks



Humans produce texts: heterogeneous socio-semantic networks

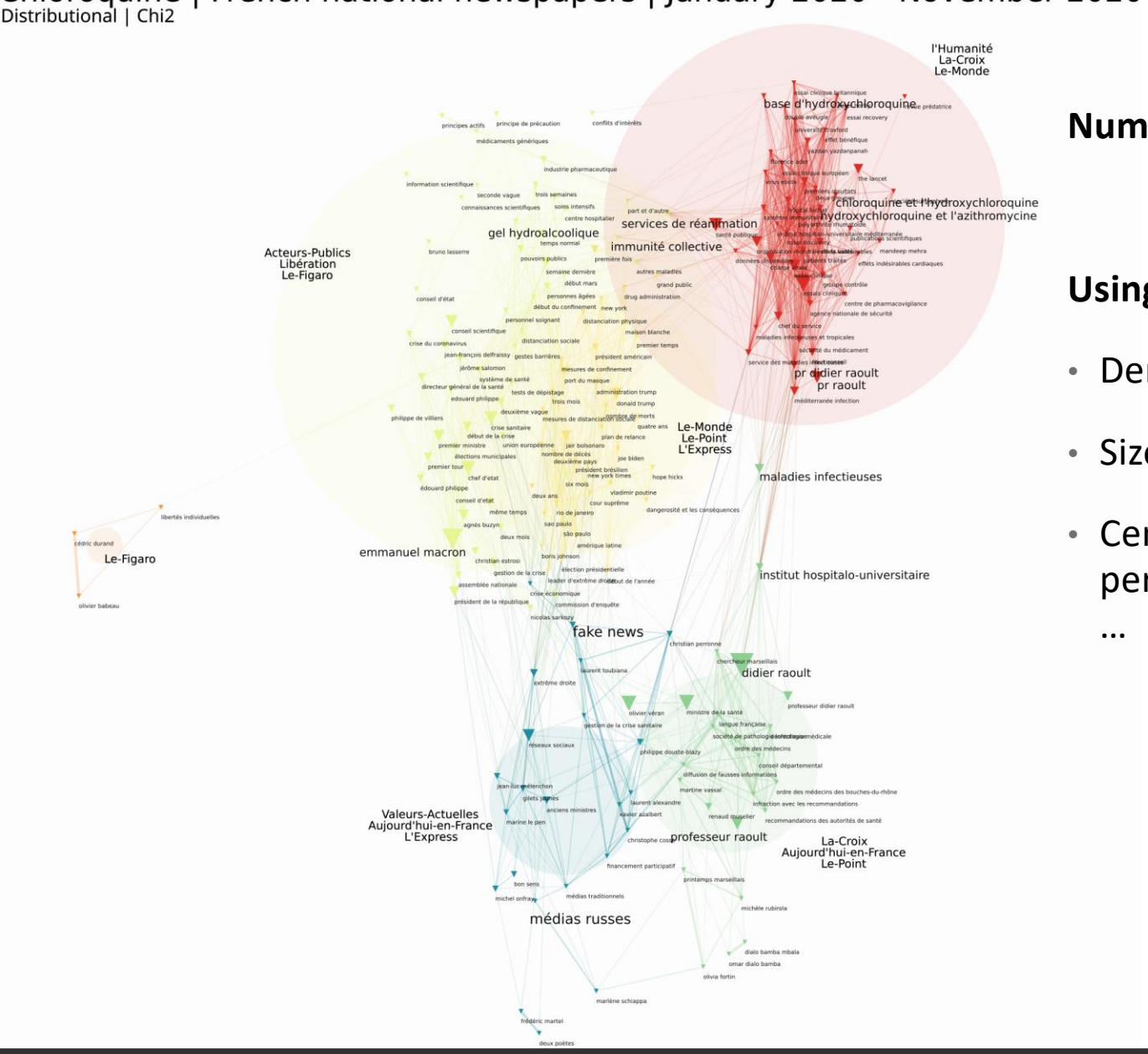


Textual contents bring relations between concepts: semantic networks

The portfolio of available metrics of similiarity

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

Macroscopic interpretation

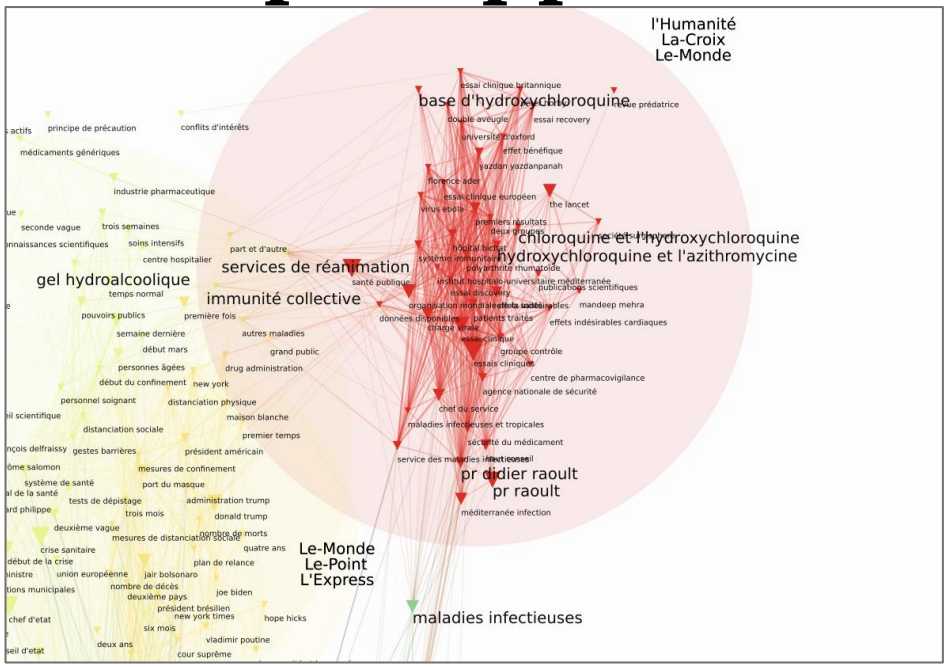


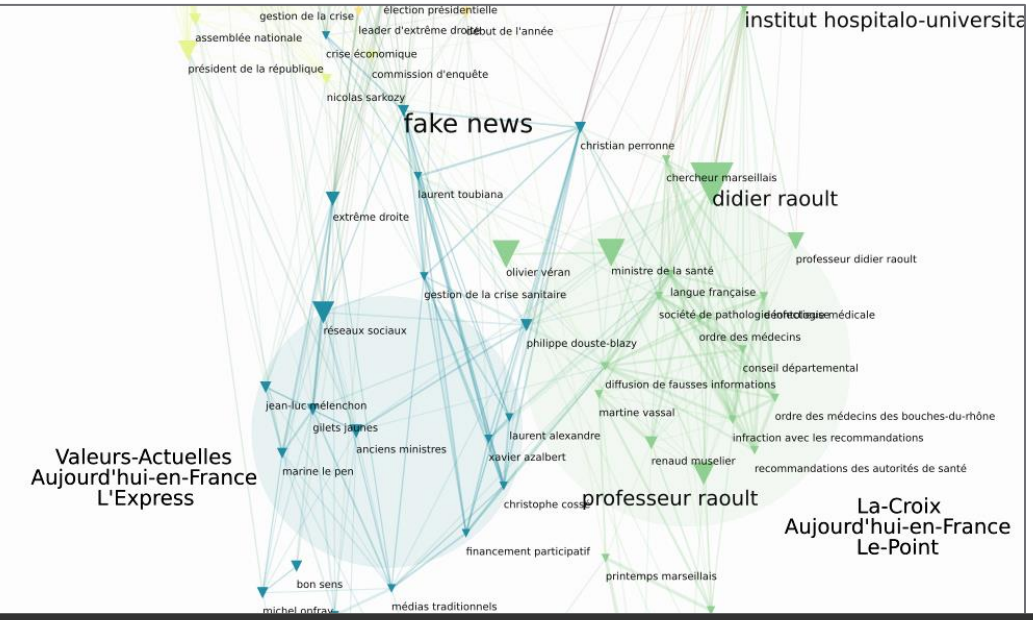
Number of Clusters

Using metrics

- Density of the graph
- Size of Clusters
- Central or peripherical positions
- …

Microscopic Approach





Local interpretation of cluster composition

Paying attention to relative positions of nodes

Using metrics

- Centrality of nodes
- Compositions of clusters (nodes, links)
- Connectivity and relational density