

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

RISIS training sessions, 2021

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**

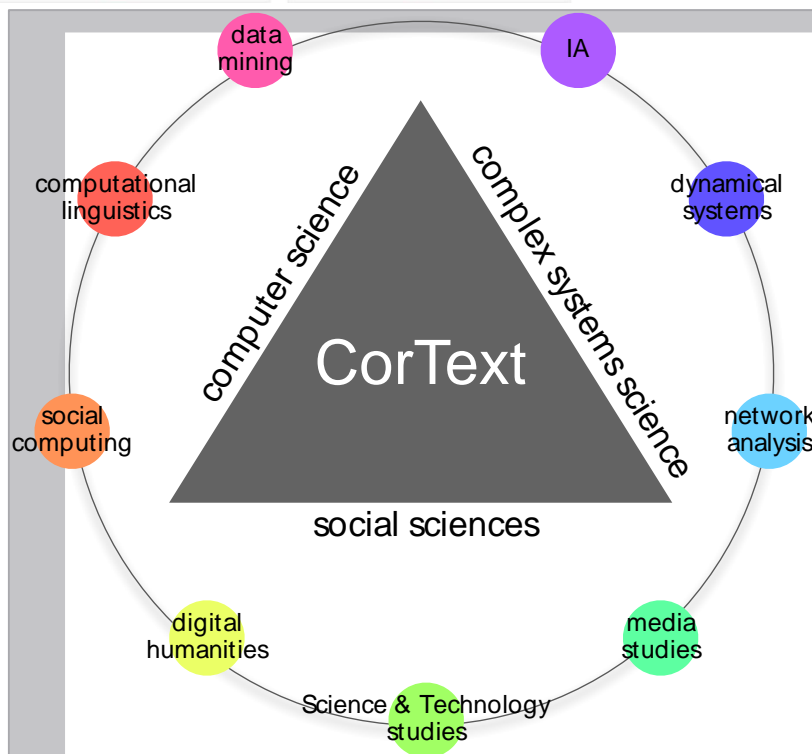
IT Engineer, CorText


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 Président - 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


Loïc Boudoulec

IT Engineer


Bertha Brenes

IT Engineer, Trainee, CorText


Anis Arabi

Big data engineer


Nicolas Ricci

Web developer and data


Audrey Baneyx

Project manager Data science,
Sciences Po Paris - medialab


Andrei Mogoutov

Bullsecence


Élise Tancoigne

Researcher, University of Geneva,
Switzerland

CorText Manager in 2021

In the last 12 months (October 2020 to 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)

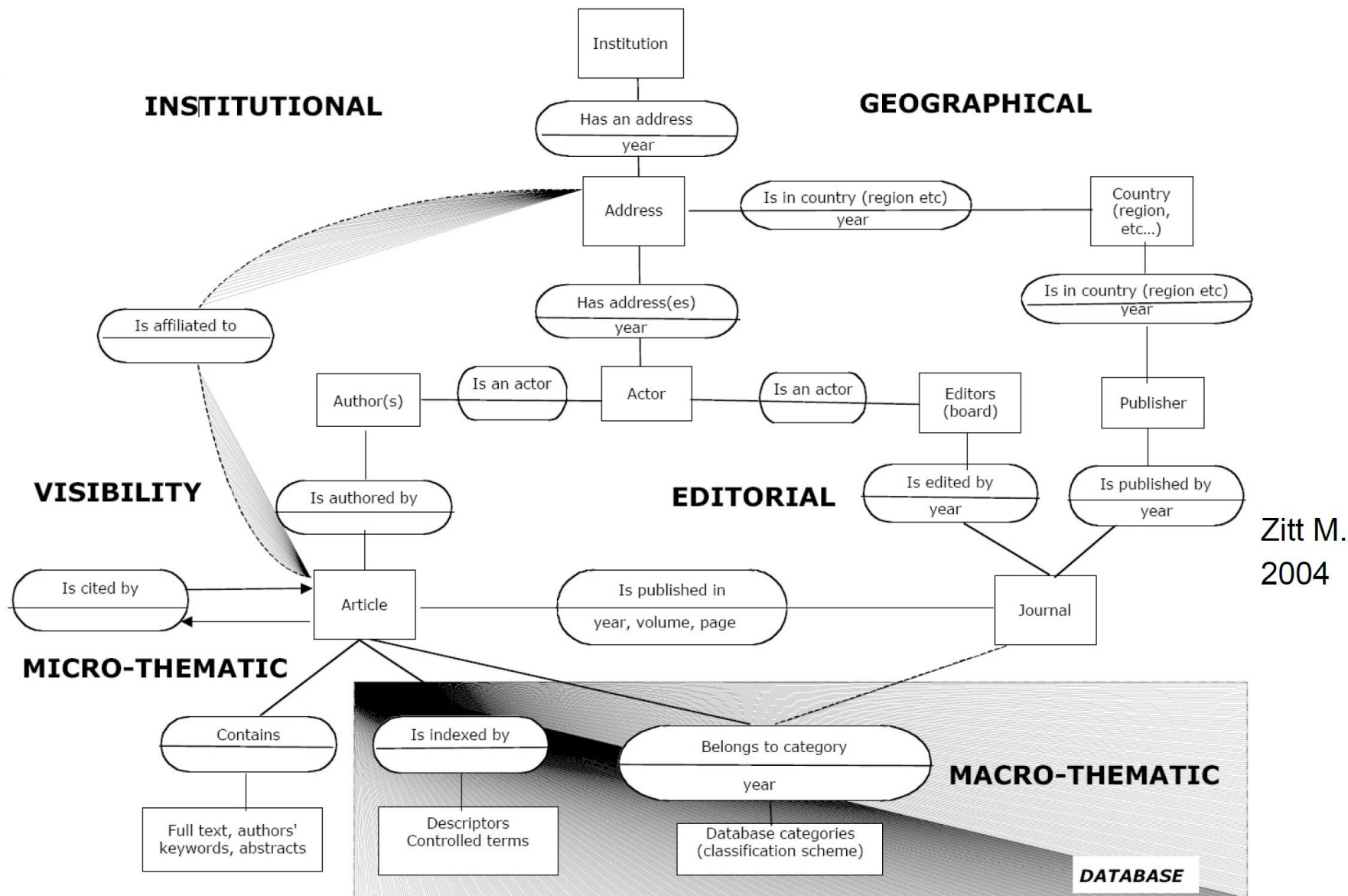
Analysis of the web and social media to characterise the digital turn and its phenomenology (US election, Yeallow Jacked, Youtube ecosystem).

- 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/risi-cortext-patents-2021/session1-background/session1b/03-examples-papers/>

The scientometric roots of CorTexT Manager

Scientific articles and bibliographic notices as hub of information



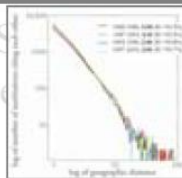

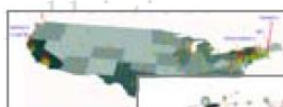
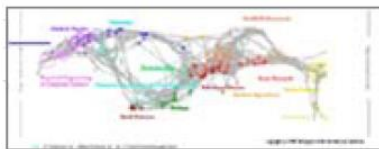
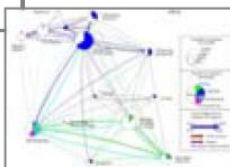
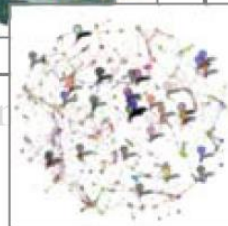

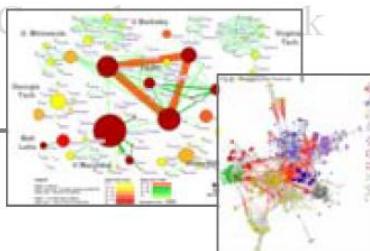
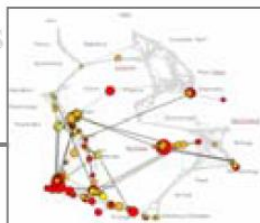
What CorText Manager produces

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

Two main types of indicators of scientific activity:

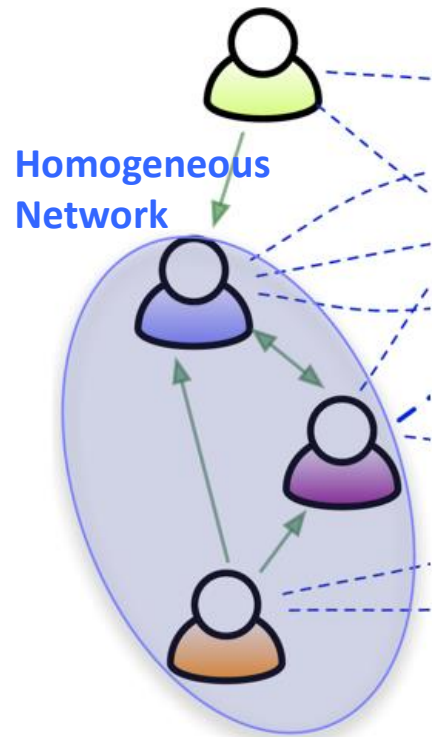
- Basic indicators and simple descriptive statistiques: stocks, ranks, frequencies;
- **Relational indicators** (networks) with, also, two main types of relational indicators:
 - **direct**: relationships do not enter directly into the content of the documents (eg: collaborations of author names);
 -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

Mining relational information in scientific publications and bib. notices

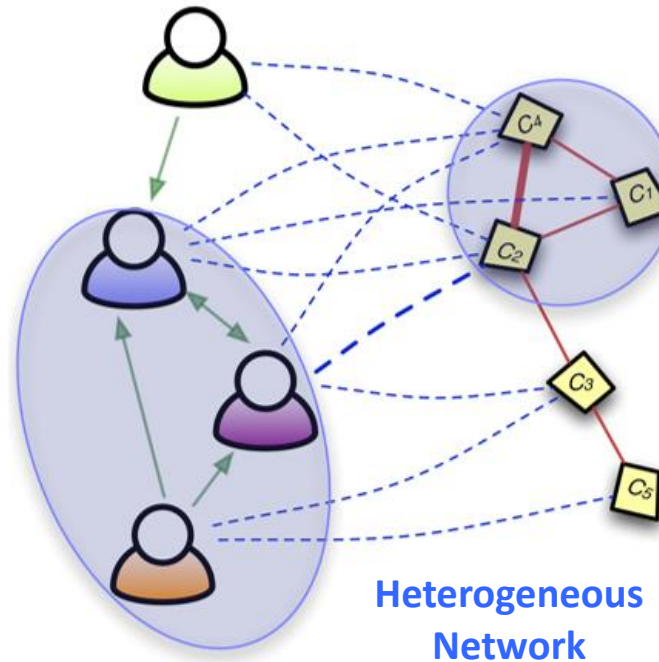
	<i>Micro/Individual (1-100 records)</i>	<i>Meso/Local (101-10,000 records)</i>	<i>Macro/Global (10,000 < records)</i>
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of science 
Temporal Analysis (When)	Funding portfolio of one individual	Scientific bursts of PNAS	113 Years of P Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a scientist's intellectual life	PNAS 
Topical Analysis (What)			VxOrd/Topic research NIH funding 
Network Analysis (With Whom?)	NSF work of one 		NIH's 

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

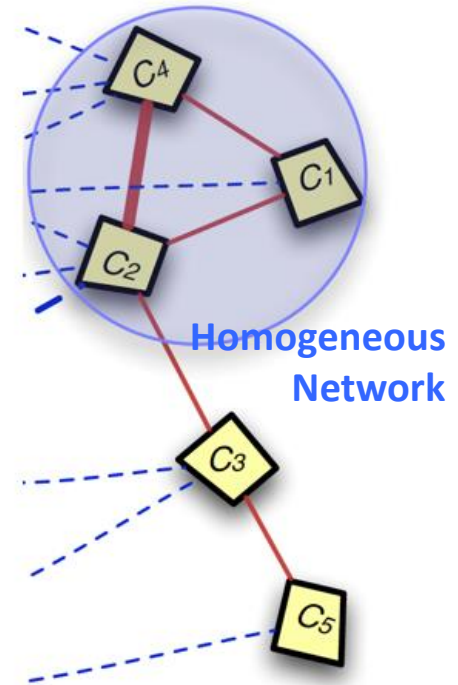
Combine direct and indirect relations within metrics and visualisations methods



Humans have relationships: sociological networks



Humans produce texts: heterogeneous socio-semantic networks

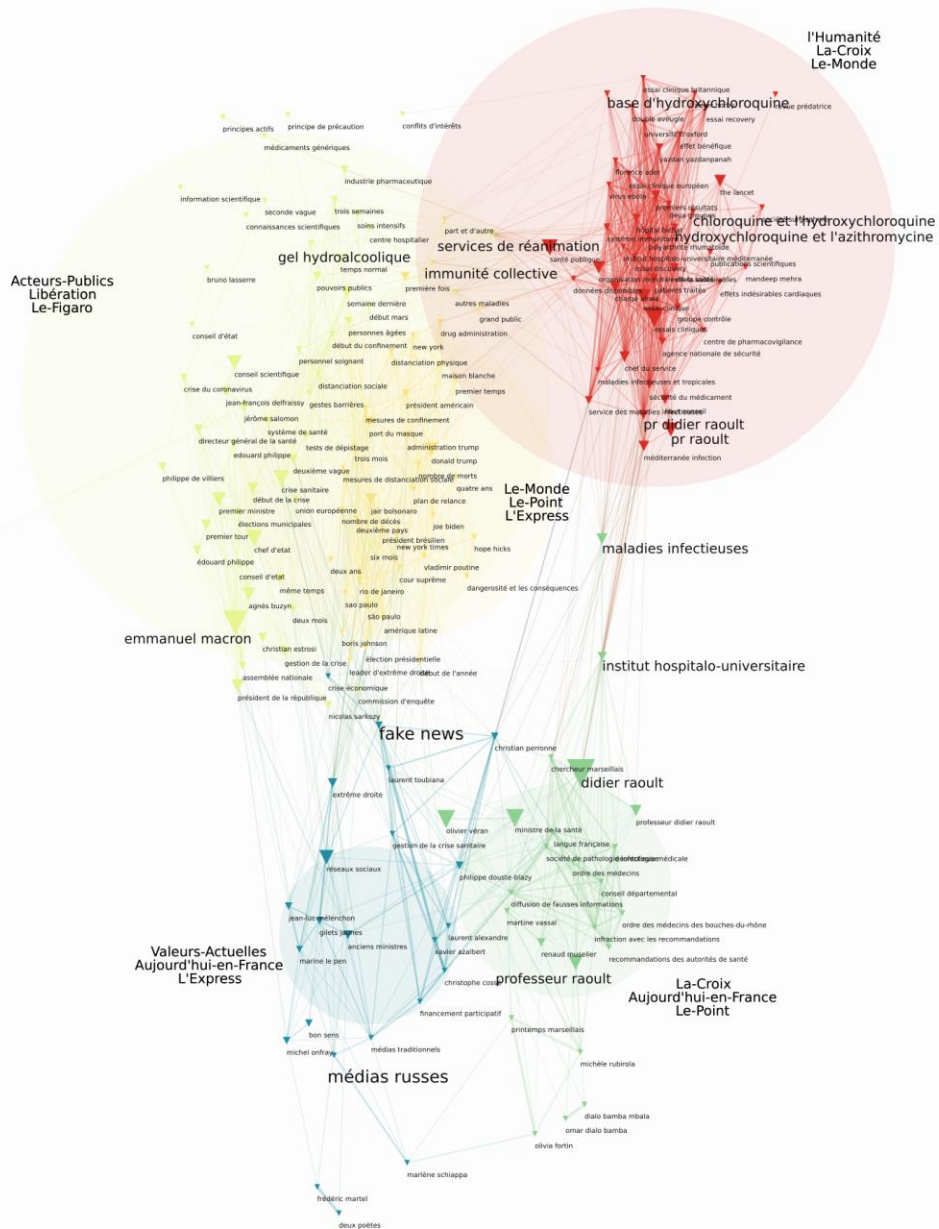


Textual contents bring relations between concepts: semantic networks

The portfolio of available metrics of similarity

proximity measures	type of network	normalisation	special properties
raw	interaction network (e.g. social network)	no	-
χ^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

RISIS training sessions | Session 1b | Introduction

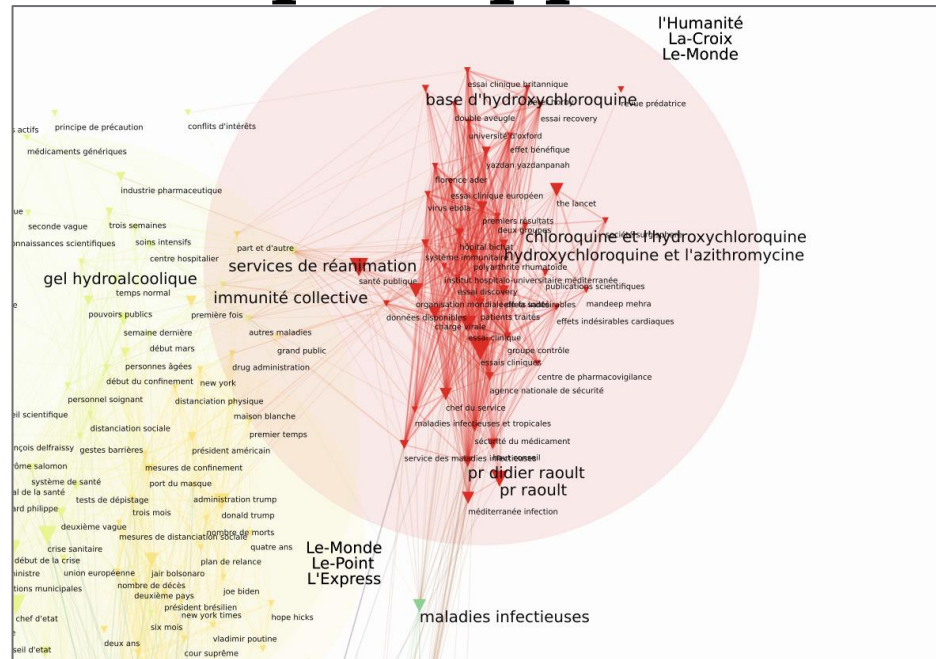
Chloroquine | French national newspapers | January 2020 - November 2020
Distributional | Chi2

Number of Clusters

Using metrics

- Density of the graph
- Size of Clusters
- Central or peripheral 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

