# Mapping the COVID-19 Literature

Dario Rodighiero (Harvard University) Eveline Wandl-Vogt (Austrian Academy of Sciences) and Elian Carsenat (NamSor)

> July 20, 2020 COVID-19 Satellite of Sunbelt



Affinity Map (PhD) https://affinitymap.epfl.ch

## Lexical Cartography of Scientific Communities

Select a Scientific Community Collect its articles and group them by author Apply Tokenization to each author Apply TF-IDF to each author Create links from shared tokens for each pair of authors These links represent the **lexical similarity** Draw network visualization



### Author 1List of tokens / WeightAuthor 2

Lexical similarity

### Dataset

COVID-19 Open Research Dataset (CORD-19) by the Allen Institute for AI July 1, 2020 release 170,000 articles

#### **FILTER 1**

### Metadata include "COVID-19"

20,000 articles

81,952 authors

### ↓

### FILTER 2

### Authors with at least 8 articles

Almost 900 authors (nodes) More than 3,000 shared tokens (links)

### Network Visual Layers

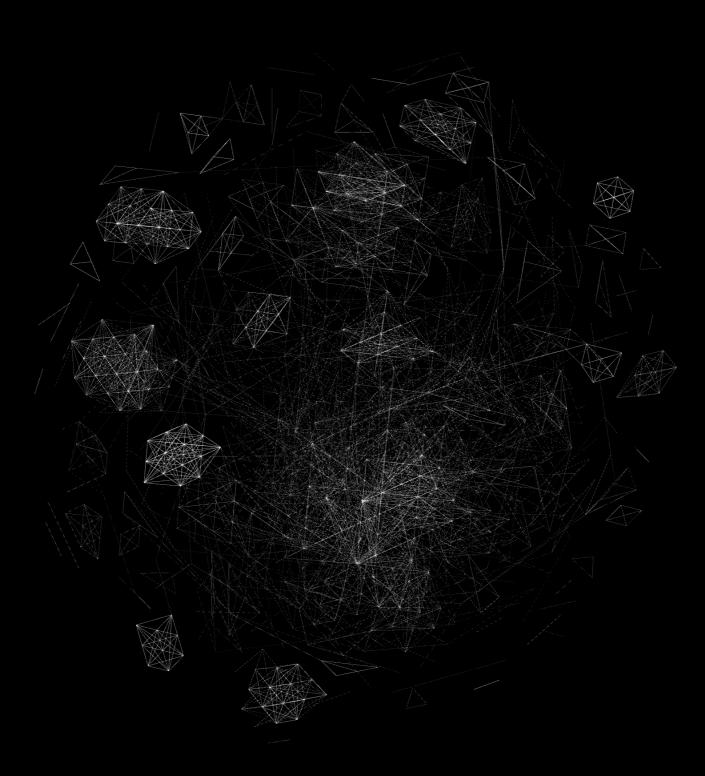
Links

Nodes

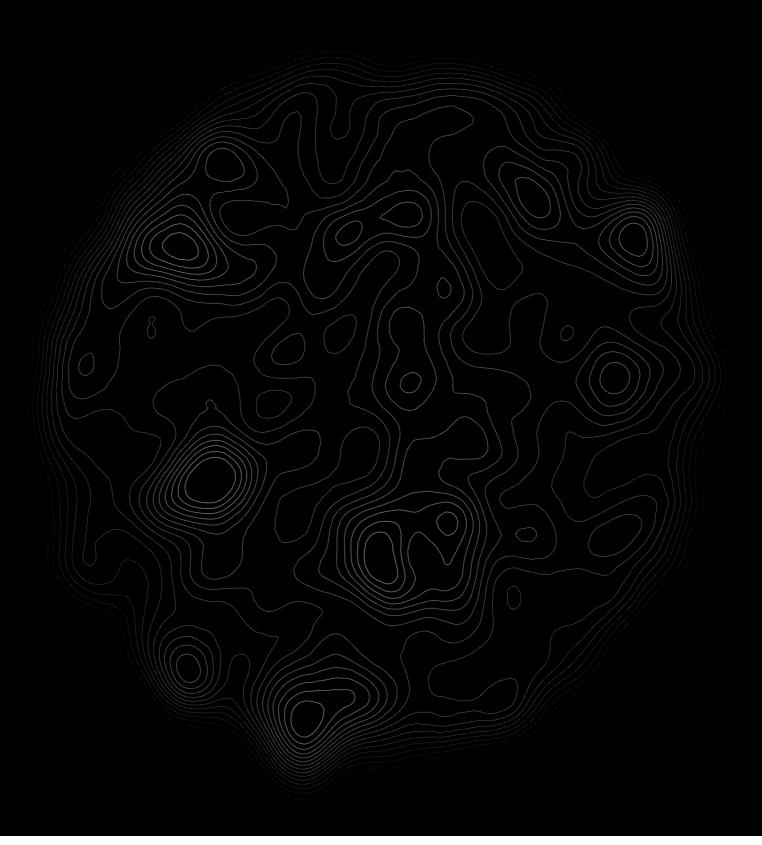
Contours

Link Keywords

Semantic Background



#### Nodes



Contours

laboratory

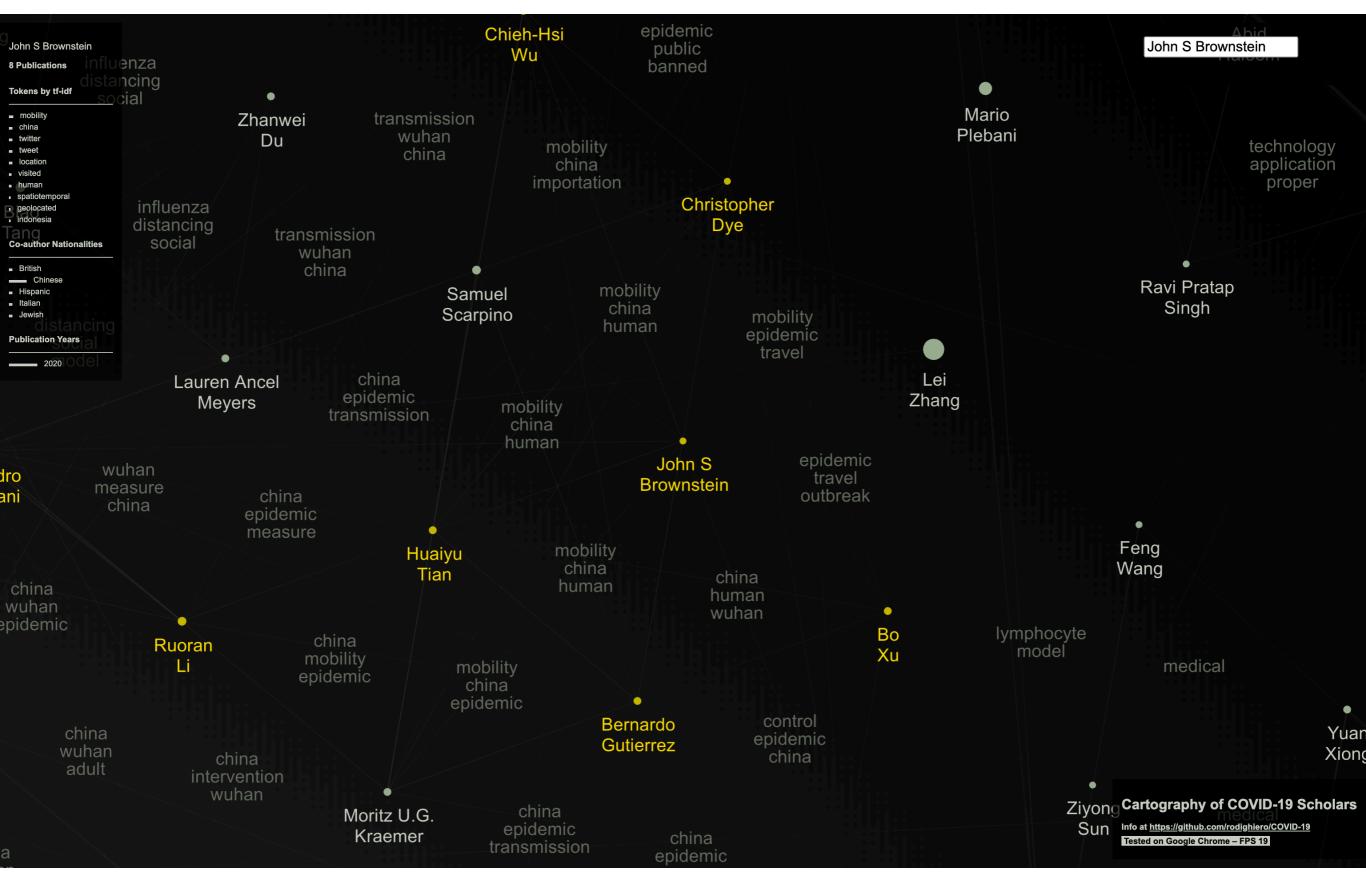
france professor france france france diabetes cardiac cardiac azithromycin azithromycin cardiac tria electrophysiology diabetes electrophysiology azithromycin physical simulation simulation simulation environmental azithromycin distancing recommendation recommendation orthopaedic isolation orthopaedic contact measure estimate taiwan source source taiwan gastrointestinal technology hamster hamster estimate estimate assay technology stemi digestive singapore mobility hamster assay dispersion korea italy hamster mobility resident assay interferon republic extremely republic assay dispersion ventilated probable resident olfactory fracture republic republic probable angiotensin hksar republic cancer olfactory quality asthenia mexican antibody disease milan mexican milar placebo emission emission anakinra decoction seroconversion mexican placebo placebo anakinra child antibiotic child child antibiotic child child child child child child child child antibiotic angiotensin cancer meteorological antibiotic emission . placebo incidental incidental anakinra placebo placebo . placebo learning surgical child child antibiotic child antibiotic placebo image statin surgical clinical incidental image feature statin child antibiotic antibiotic surgical antibiotic antibiotic statin pregnant celebrex statin statin quantitative shedding marker balance treatmer transfusion statin balance balance interactome mortality transfusion viral documento suppressor interactome balance transfusion survivor adverse radiology balance store africa convalescent factor africa severe severe store zhejiang death exhaustion alveolar clinical zhejiang liver asymptomatic zoonotic liver shanghai clinical hetil hetil cluster anosmia furin cancer symptom anosmia shanghai nucleic shanghai hetil hetil renal renal liver group hetil renal hetil renal disease clinical shanghai heparin hubei symptom tracheostomy severe heparin symptom hubei worker worker clinical worsening rheumatic refractory worker hubei discharge hubei worker variate survivor survivor survivor obstetric telehealth ncd4lr survivor ventilation telehealth vaccine asymptomatic telehealth arima urological pharmacist taiwan extra urological

ophthalmologist

#### Link Keywords



Semantic Background



Scholar's Details (zoom in)



Cartography of COVID-19

## Network visualization available at

https://rodighiero.github.io/COVID-19

## Principles

Lexicon can be more democratic metric Visualize both local and global scales

## Features

Regular pattern to host semantic information Activate information with zoom level Mapping bigger communities using video game technology

## Limitations

## Before was the number of elements to visualize Now it is the computational time for lexical distance

## Conclusion

The most prolific researchers of COVID-19 published more than 50 papers in six months.

COVID-19 Scientific Literature is so large that there are technical limits to visualize it.

If we can't even visualize it, how do we think we can read it?

# Thanks

<u>@dariorodighiero</u>

https://dariorodighiero.com