

Mapping the COVID-19 Literature

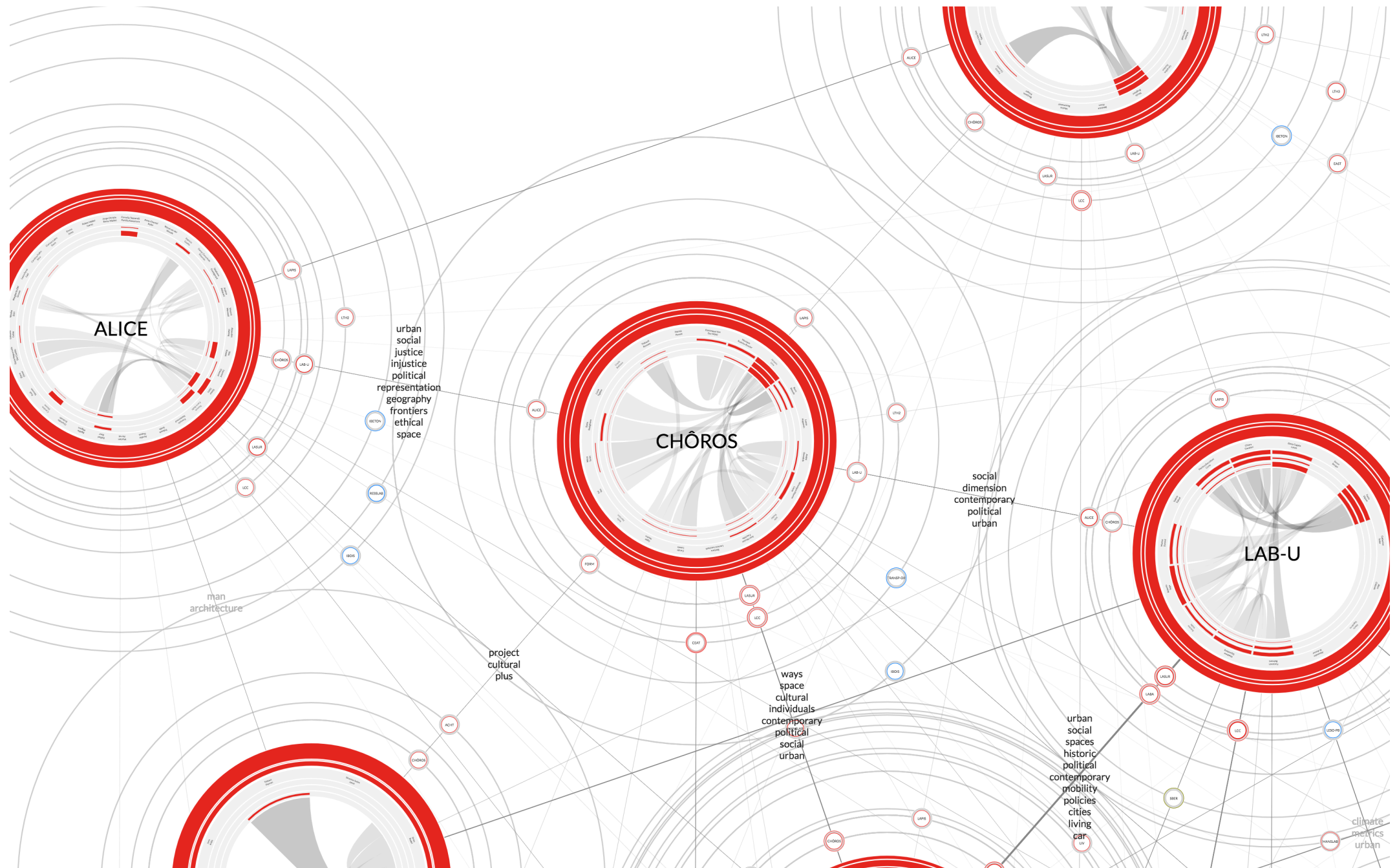
Dario Rodighiero (Harvard University)

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

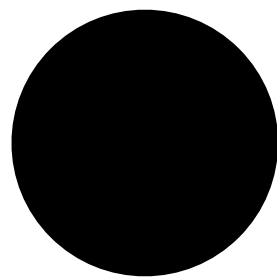
Apply Tokenization to each author

Apply TF-IDF to each author

Create links from shared tokens for each pair of authors

These links represent the **lexical similarity**

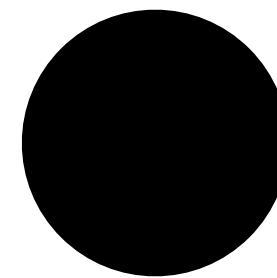
Draw network visualization



Author 1



List of tokens / Weight



Author 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

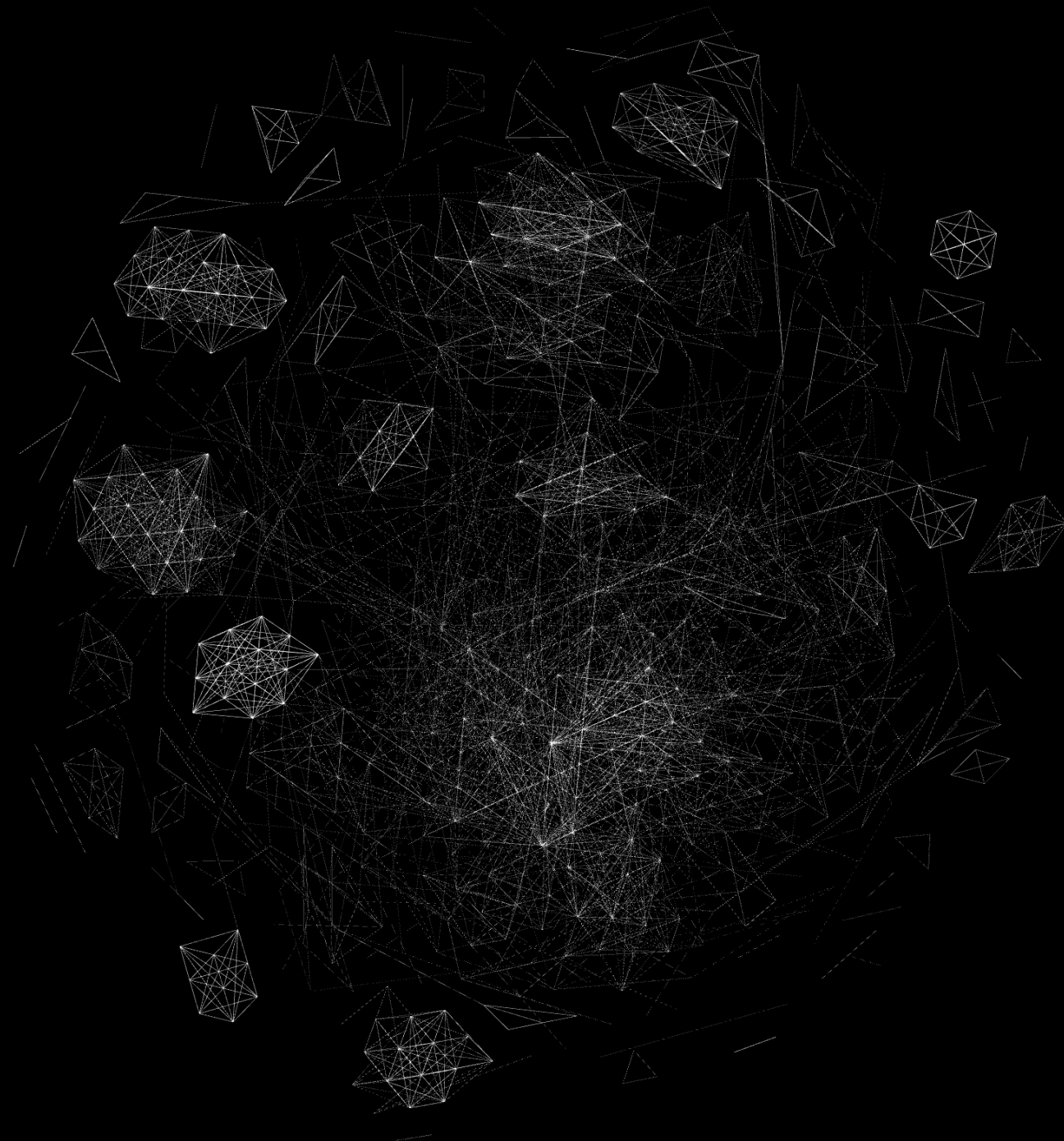
Links

Nodes

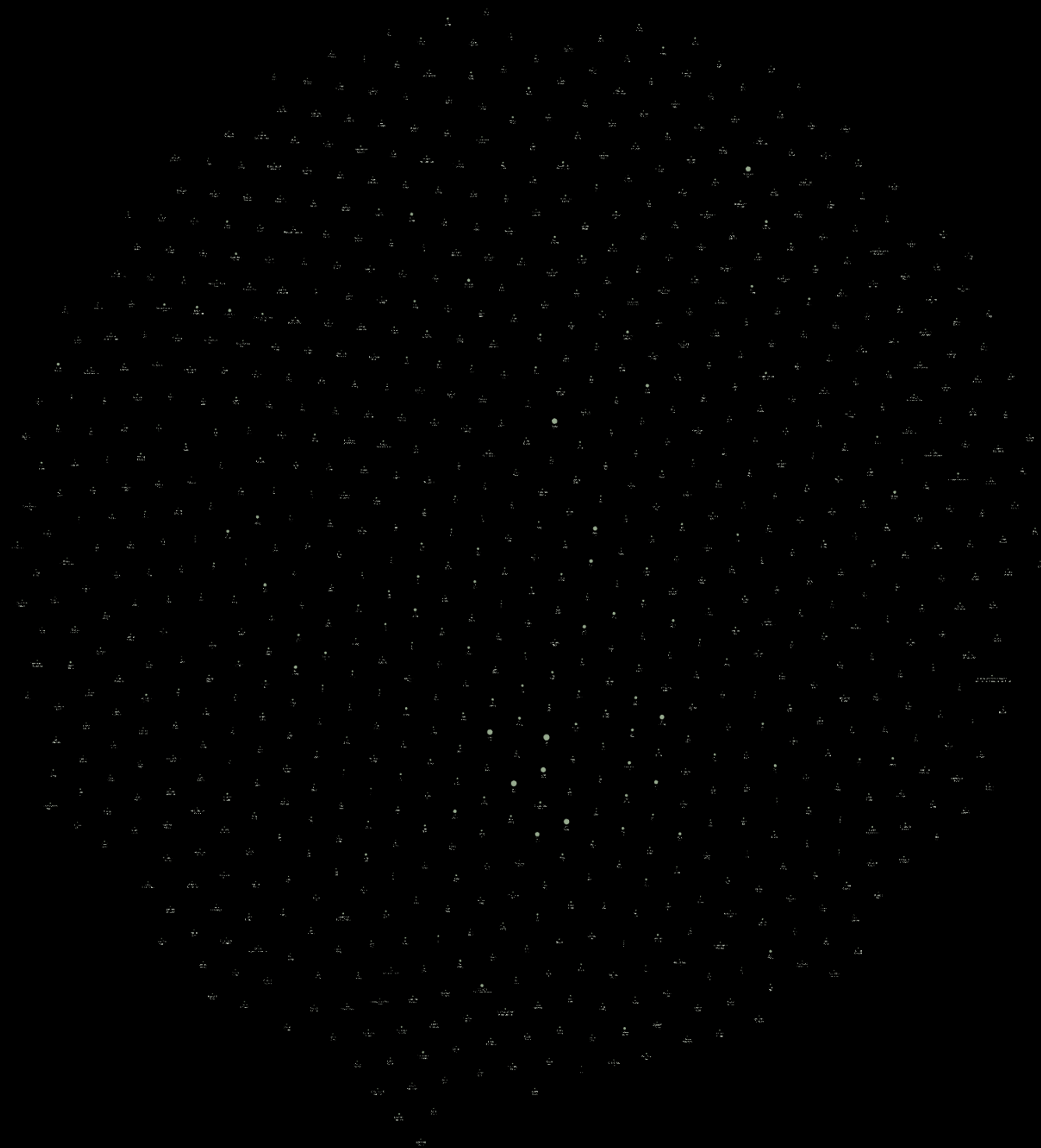
Contours

Link Keywords

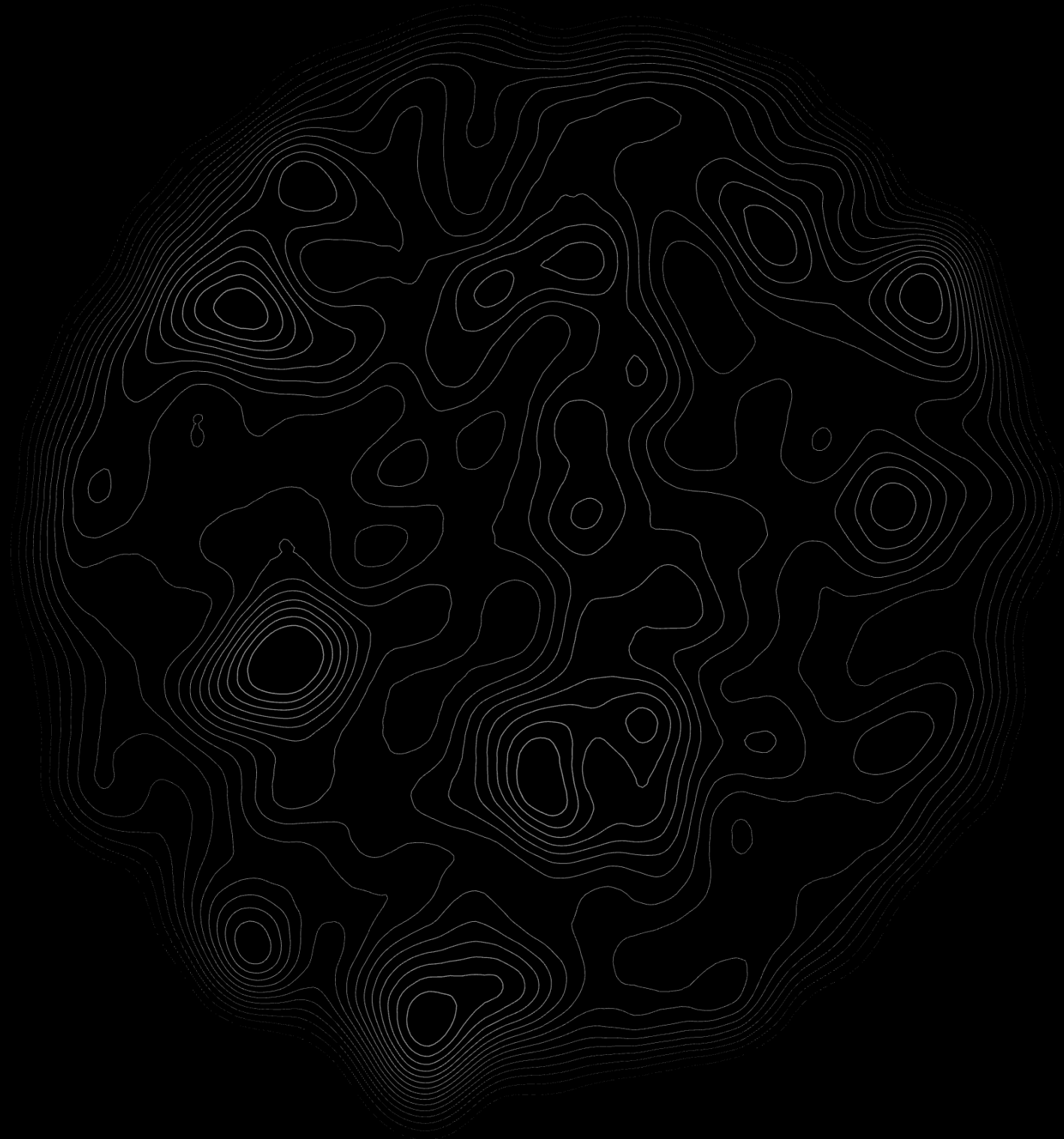
Semantic Background



Links



Nodes



Contours



Link Keywords



Semantic Background



Scholar's Details (zoom in)

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

@darioodighiero

<https://darioodighiero.com>