



AUTHOR COLLABORATION NETWORK ANALYSIS ON "CORONAVIRUS"

Edited by:
Francesco Musetti
Enrico Caria
Ilaria Muru
Alberto Paschina

INTRODUCTION

The project aims, through network-based methodologies, to analyze collaboration networks in the context of the "Coronavirus". Specifically, it will focus on collaboration networks among authors in scientific research, describing how this phenomenon has evolved over time, including in response to the Sars-Covid19 pandemic. A second analysis seeks to match the results of topic modeling with the graph, in order to observe collaborations among authors and highlight them by the topic most addressed by each author.

DATASET E METHODOLOGY

For the analysis we have used open data from the PUBMED database, that is a free bibliographic database that contains millions of information about articles published in medical and biomedical literature and is managed by the National Center for Biotechnology Information (NCBI).



The source code was written in Python and mainly used the NetworkX, Matplotlib, Pandas, and Gensim libraries to generate the graphs, plots, and topic modeling.



RESULTS

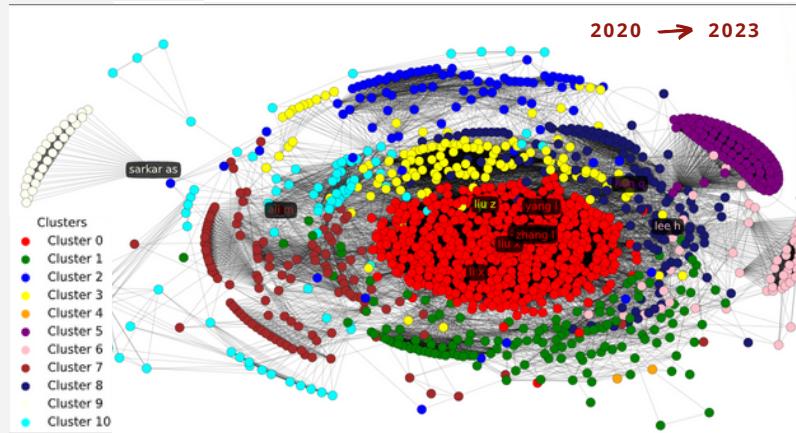
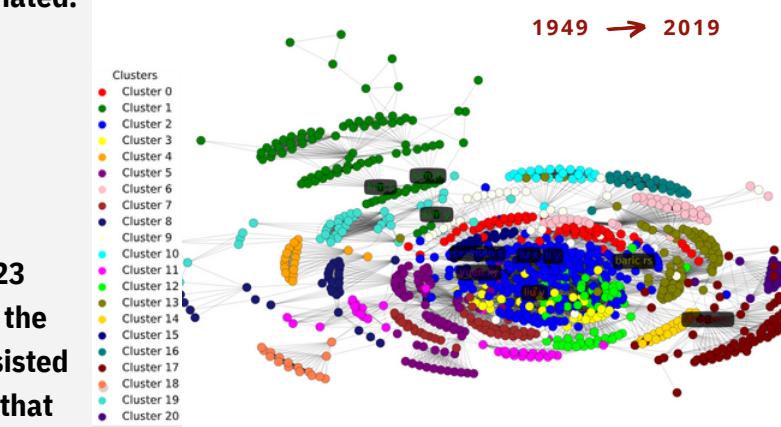
The analyses carried out by us aim to find answers to natural questions that arose in the initial phase. The two main analyses, the results of which are shown below, concern, on one hand, the analysis carried out to verify whether the Scientific Community on the topic of "Coronavirus" was extremely connected or whether it was possible to identify sub-communities that showed more frequent collaboration.

On the other hand, they concern the analysis carried out to verify which topics were most discussed by the Scientific Community on the topic of "Coronavirus" and their evolution over the years, also through collaborations between authors depending on the topics they dealt with.

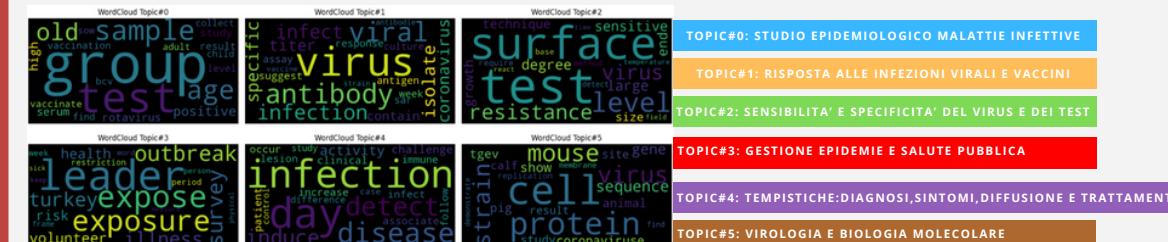
Regarding the first proposed analysis, using Louvain's algorithm, we applied a cluster analysis on the principal components of the graphs constructed for each period, both of which were filtered for computational purposes so that authors with a low number of published articles were eliminated.

We identified 20 clusters for the period 1949-2019.

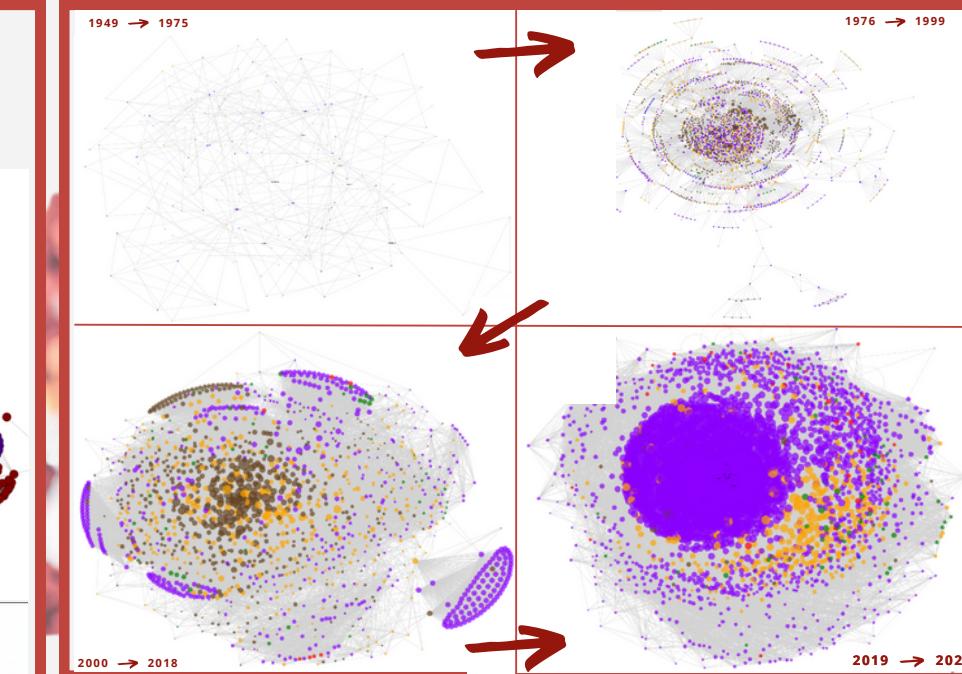
While for the period 2020-2023 there are 10 clusters, despite the fact that the latter graph consisted of many more nodes, proving that among the most prolific authors from a publication point of view, collaboration increased following the pandemic.



Regarding the second proposed analysis, we opted for the application of a Topic Modeling algorithm, which is used in Natural Language Processing (NLP) and that allows associating a topic or theme to a document in a collection of documents. From the analysis of the Abstracts associated with each publication, it was possible to distinguish the documents into six categories shown in the WordCloud, which were interpreted to generate the following 6 dominant Topics proposed:



By associating each author with one of the identified Topics based on the themes addressed by the authors themselves, it was possible to visualize the evolution of collaborations over time, highlighted by the Topics discussed, as shown in the following graphs:



CONCLUSIONS

Through the analyses performed, therefore, it was possible to provide answers to the questions initially proposed:

- It was highlighted that the Collaboration Graph can be divided into multiple components indicating clusters of more intense collaborations between authors.
- It was highlighted how the advent of the Sars-Covid19 Pandemic affected the aforementioned clusters of collaboration.
- It was highlighted how, in the pre-pandemic period, the literature on coronavirus topic reflects the ranking of the most authoritative journals in the field and how this is not the case during and post pandemic period.
- It was highlighted how indeed First Authors play a role as a "bridge" between collaborative communities.
- It was pointed out how it is also possible to identify the most widely covered topics in Scientific Research in the field of "Coronavirus" and their evolution over time through graph analysis combined with Topic Modeling.

SOURCES

- Official [NetworkX Documentation](#)
- Official [Gensim Documentation](#)
- Official [Matplotlib Documentation](#)
- Official [Pandas Documentation](#)
- [PubMed website](#)
- [MDPI website](#)
- Compendio insegnamento di Economia dei Network - A.A 2021-2022 - Prof. Stefano Matta
- Università degli Studi di Cagliari
- [SCImago Journal Rank o SJR indicator website](#)



SCAN ME