جامعة نيويورك أبوظبي

NYU | ABU DHABI

Report of Project

Guanghui Pan | Abbas K. Rizi

University of Oxford | Aalto University

January 2024

Contents

Abstract

Strategizing Pandemic Preparedness: Optimizing Scientist and Policymaker Collaboration for Future Challenges

In this paper, we delve into the critical question of which scientific experts should be integrated into teams to enhance preparedness and informed decision-making in the face of COVID-19 and future pandemics. Our research presents a comprehensive strategy for assembling a collaborative framework between scientists and policymakers, specifically tailored to address the complexities of pandemic response. Using COVID-19 as a pivotal case study, we refine our methodology, drawing vital insights from various international approaches to managing such multifaceted crises. Our findings reveal a significant convergence of research areas within epidemic studies over the past two years, highlighting an increase in interdisciplinary collaboration. This paper demonstrates the importance of understanding the dynamics of scientific teamwork and the evolution of research themes, as these elements are key to gaining a strategic advantage in controlling and mitigating the impacts of forthcoming epidemics. Through our analysis, we aim to provide a blueprint for effective synergy between scientific inquiry and policy formulation, paving the way for more resilient and responsive systems in anticipation of future global health challenges.

1 Introduction

The COVID-19 pandemic has served as a stark reminder of our vulnerability to infectious diseases. It wasn't the first and certainly won't be the last pandemic that humanity will face. This recent global health crisis has underscored the critical need for preparedness against future pandemics. The nature of pandemics, inherently devastating, can rapidly overwhelm healthcare systems and cause extensive social and economic disruptions. The unpredictable nature of viral mutations, combined with the interconnectedness of our world, only exacerbates these risks. Therefore, it's crucial to understand the mechanisms of pandemic spread, including the role of social networks and human behavior [1, 2]. This knowledge is vital for developing effective intervention strategies [3] and crucial in mitigating the severe consequences that unchecked pandemics can have on global health, economies, and societies [4]. Our future resilience hinges on learning from past experiences, investing in public health infrastructure, advancing scientific research, and fostering global cooperation to manage and ultimately prevent the devastating impact of pandemics.

The reality of COVID-19 caught everyone off guard. "We don't want to be in that situation in the future" is a common sentiment. Our goal is to have something akin to a Weather Service for epidemics and epidemic threats. Seasonal flu and RSV outbreaks have already pushed community hospitals to their limits. This envisioned service would be a hub for the country's monitoring, forecasting, and scenario analysis regarding all possible threats, including known diseases like Ebola and Zika and novel infections. We need a program that aims to create a first-of-its-kind, long-term academic collaboration on pandemic preparedness using advanced mathematical modeling and systematically collected health data from various sources. Participants include epidemiologists, statisticians, mathematicians, and computer scientists, with several representatives from each public health institute and other disciplines. The program aims to use clinical health data combined with real-time data streams representing social activity and human mobility, alongside advanced mathematical modeling and computational methods, to address several urgent questions for COVID-19 and future pandemics.

The program will develop methods, tools, and operational procedures for implementing interoperable health data pipelines, novel methodology published in international scientific journals, and support the public health institutes in their aim to keep disease spreading low without causing too high a burden on societies. However, forming such a diverse group of scientists is challenging in practice. This endeavor's complexity and interdisciplinary nature make it a formidable task, yet it is a necessary one. How we respond to this call for action against such complex challenges will define our preparedness for future pandemics.

In this project, we will implement a data-oriented approach to recruit top experts for pandemic preparedness and decision-making. Utilizing comprehensive datasets and analytics, we'll identify professionals with specific expertise relevant to pandemic management challenges according to the closeness of topics in the epidemic literature. This method ensures that our team is highly knowledgeable and adept at innovative, interdisciplinary collaboration. By focusing on data-driven selections, we aim to assemble a task force uniquely equipped to efficiently address the complexities and dynamics of future pandemic threats.

2

Data & Method

The **COVID-19 Open Research Dataset (CORD-19)** is a comprehensive resource of scholarly articles on COVID-19 and related coronaviruses [5]. Developed by the *Allen Institute for AI*, in collaboration with leading research groups, CORD-19 aims to facilitate the application of recent advances in natural language processing and computational techniques to generate new insights into the COVID-19 pandemic.

As a dynamic dataset, CORD-19 encompasses over 44,000 scholarly articles, with more than 29,000 featuring full text. The dataset is designed for the global research community and is updated weekly to include new research from peer-reviewed publications and archival services like bioRxiv and medRxiv.

Key components of CORD-19 include:

- `metadata.csv`: A file released weekly, documenting COVID-19 updates and new research, containing fields such as `cord_uid`, `title`, `doi`, `abstract`, `publish_time`, `authors`, and `journal`.
- Full text documents: JSON files with full-text parses of articles, available in `pdf_json` and `pmc_json` formats.
- Document embeddings: Precomputed embeddings for each paper, facilitating various NLP tasks.

Researchers can utilize the dataset for a multitude of purposes, from extracting specific information like titles and abstracts to conducting in-depth analyses. The versatility of CORD-19 makes it a vital tool in ongoing COVID-19 research and analysis. For detailed information and access to the dataset, please refer to the GitHub repository of CORD-19 and the Allen Institute for AI's CORD-19 page.

Our methodology utilizes a comprehensive dataset comprising over 200,000 academic papers spanning from 1970 to 2022. We employ rigorous sampling techniques to analyze the textual content of these documents. To categorize the diverse range of topics, we apply the elbow method, a robust statistical technique, to determine the optimal number of clusters. Our analysis indicates that dividing the topics into nine distinct groups provides the most coherent and meaningful categorization.

Our primary objective is to examine the evolution of these thematic clusters over time, particularly during the COVID-19 era. To achieve this, we have segmented our post-2020 data into five distinct six-month intervals. This temporal division allows us to meticulously track and analyze the shifts in topic clustering, offering valuable insights into how academic discourse and research priorities have adapted in response to the pandemic. This approach not only enhances our understanding of the evolving scientific landscape but also guides our expert recruitment strategy, ensuring a dynamic and responsive team composition.

3
Results

The ensuing Fig.3.1 vividly illustrates the dynamic shifts within the academic landscape over just two years. They effectively highlight the convergence of various previously distinct topics, demonstrating an increased interdisciplinary integration. Additionally, these figures reveal a notable trend where specific topics, initially sparse and less explored, have gained substantial traction and become more densely populated with research and discussion. This visualization not only underscores the rapid evolution of academic focus areas but also reflects the responsive nature of research to emerging global challenges, particularly in the context of pandemic preparedness and response.

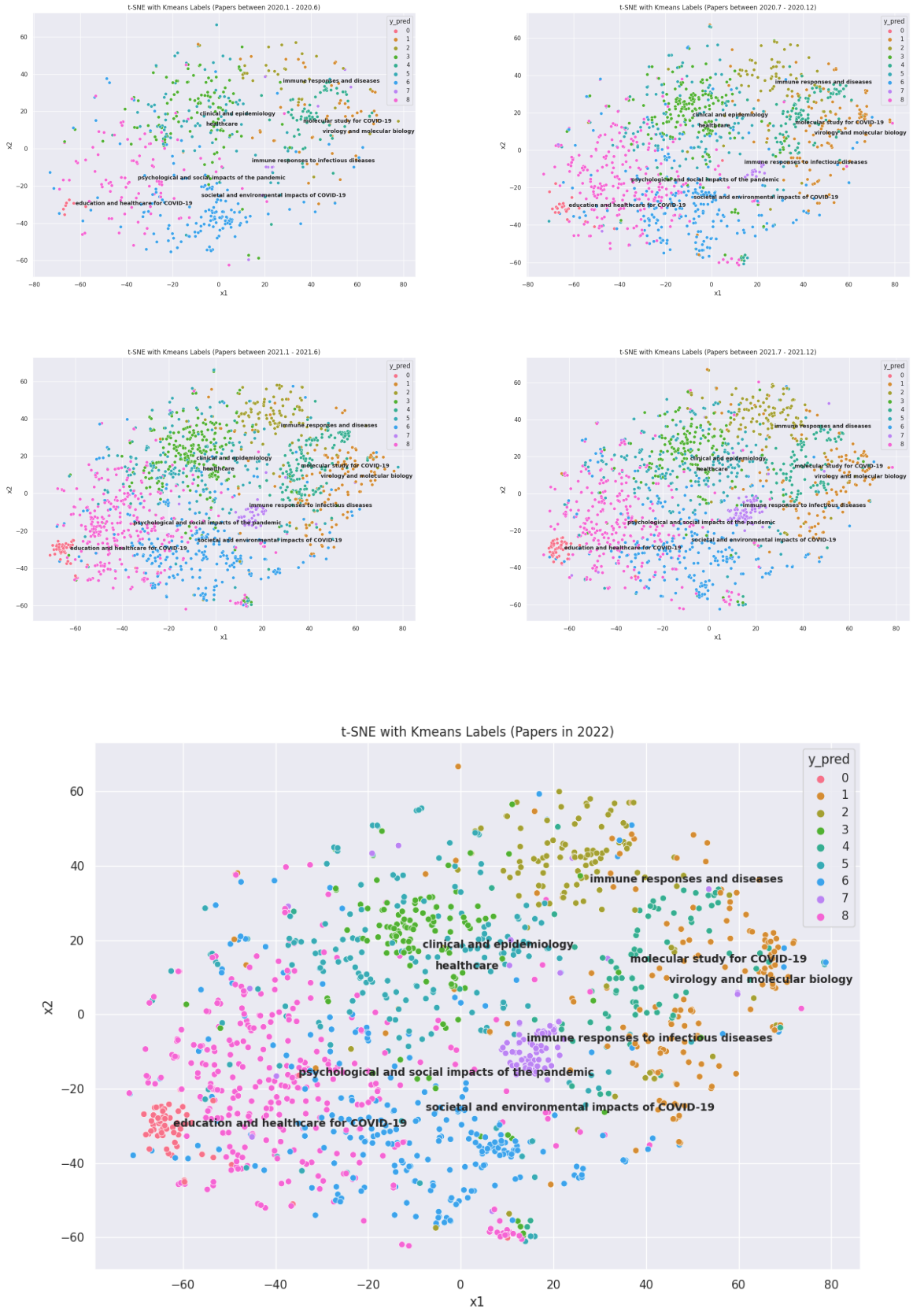

Figure 3.1: Progression of Topic Clusters in Epidemic-Related Research: A Biannual Analysis from 2020 to 2022

Bibliography

[1] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

[2] István Z Kiss, Joel C Miller, Péter L Simon, et al. Mathematics of epidemics on networks. *Cham: Springer*, 598:31, 2017.

[3] Nicola Perra. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Physics Reports*, 913:1–52, 2021.

[4] Melissa Leach, Hayley MacGregor, Ian Scoones, and Annie Wilkinson. Post-pandemic transformations: How and why covid-19 requires us to rethink development. *World development*, 138:105233, 2021.

[5] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Douglas Burdick, Darrin Eide, Kathryn Funk, Yannis Katsis, Rodney Kinney, et al. Cord-19: The covid-19 open research dataset. *ArXiv*, 2020.