Cluster analysis of the association of COVID-19 mortality time series with pandemic intervention measures

Murat Razi, Manuel Graña

Abstract

- COVID-19 mortality time series is not uniform across countries
- Intervention measures (control stringency, vaccines) have been similar, but not uniform
- We look for association between intervention measures and COVID-19 mortality
  - Intersection between clusters of countries discovered independently over the COVID-19 mortality and each intervention measure
- We find little association between mortality and intervention measures

Contents

- Materials and methods
  - Data
  - Feature extraction
  - Spectral clustering
  - Overall computational pipeline
- Results
- Discussion
- Conclusions

Materials and methods

- Data
  - “Our world in data” (OWID) is a web site devoted to the publication of relevant data for the pressing geopolitical and economic questions. It is hosted by the University of Oxford
    - “new_deaths_smoothed_per_million”
    - “stringency index”
    - “new_vaccine_doses_smoothed_per_million”
  - From 1/1/2020 up to 1/1/2023

Materials and methods

- *NNMF:* Non-Negative Matrix Factorization
  - is a mathematical technique used for decomposing a non-negative matrix **V** into the product of two lower-dimensional matrices, **W** and **H** both non-negative
  - useful for data analysis, feature extraction, and topic modeling
  - Matrix **W** represents the basis vectors or features, and
  - matrix **H** represents the coefficients for combining these features to reconstruct the original data.

Materials and methods

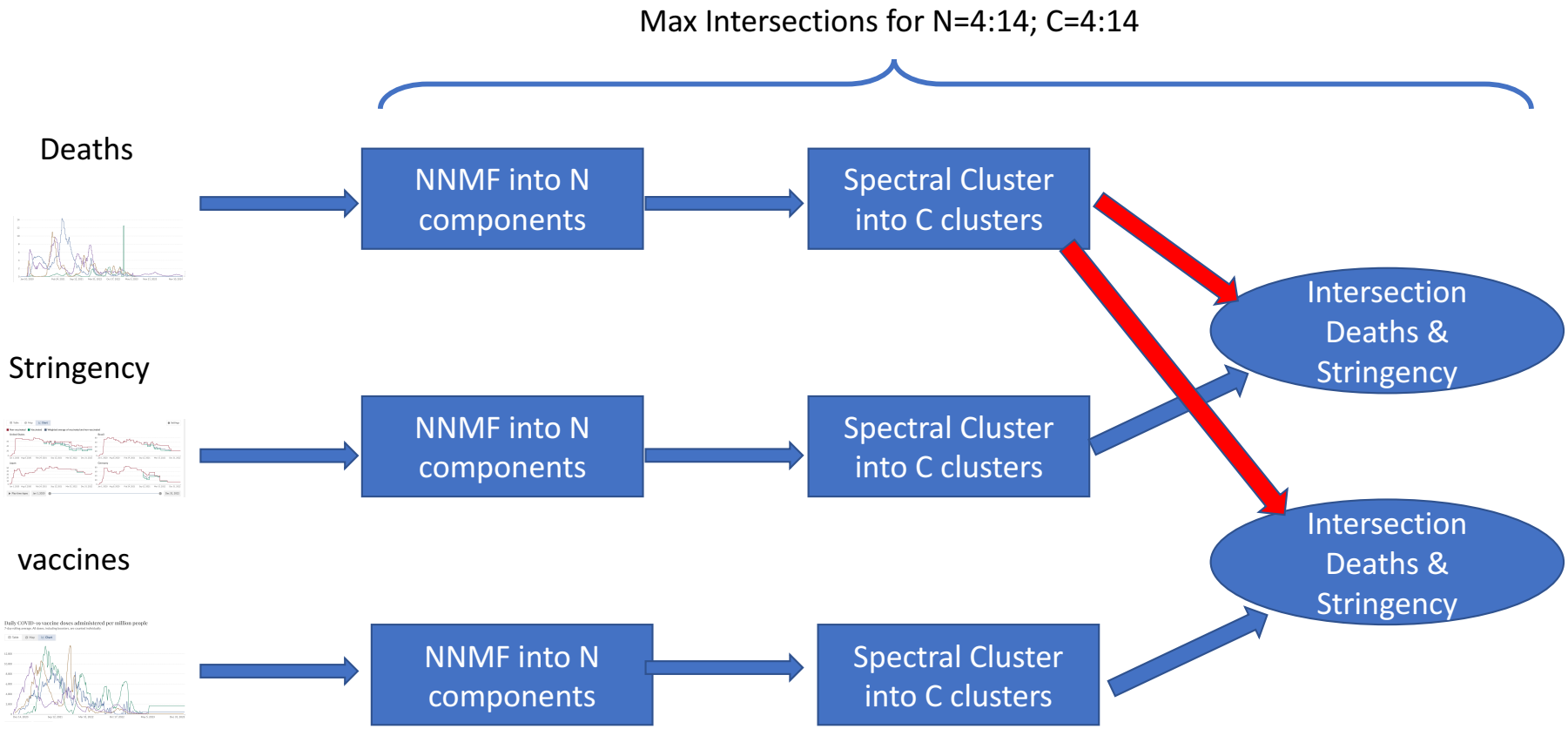
- Spectral clustering
  - Spectral clustering is a machine learning technique used for partitioning a dataset into meaningful and homogeneous clusters.
  - The second eigenvector of the Laplacian of the data samples similarity graph provides the optimal graph cut.
  - In this paper, spectral clustering is applied over coefficients of NNMF to cluster countries into different latent topics
  - This process is carried out independently for each kind of time series.
    - i.e. we find clusters according to mortality, stringency index and vaccine doses

Materials and methods

- Overall processing pipeline
  - is a grid search over the number of NNMF components and clusters of time series looking for the number that achieves the maximal intersection of clusters of countries in the range 4:14.
  - The rationale of this process is that if an intervention measure has any influence on the mortality from the virus, then clusters of countries with similar intervention measures patterns should be similar to clusters with similar mortality patterns.
  - This second degree similarity should correspond to the **association** between intervention measures and mortality.

Algorithm 1 Grid search for the selection of the optimal numbers N^* and C^* of NNMF components, and clusters, respectively, for the Stringency Index and the Vaccine doses maximizing the in intersection with per million deaths clusters.

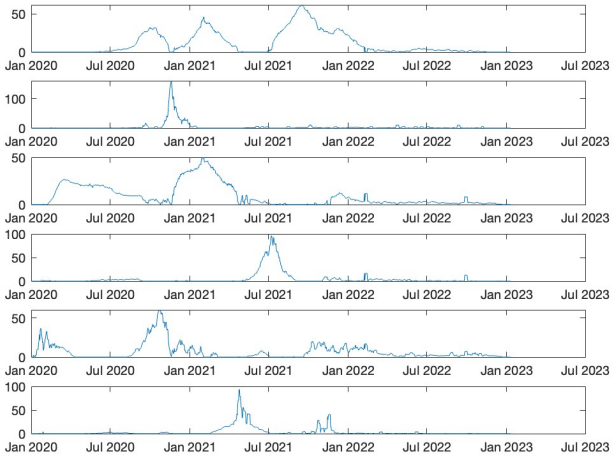
1) For $N = 4:14$ and $C = 4:14$
   a) Extract N NNMF components from the Stringency Index time series.
      i) Compute the NNMF feature extraction of the Stringency Index time series, denote this collection of feature vectors $SI(N)$
   b) Extract N NNMF components from the Vaccination doses time series.
      i) Compute the NNMF feature extraction of the Vaccination doses time series, denote this collection of feature vectors $V(N)$
   c) Extract N NNMF components from the per million death time series.
      i) Compute the NNMF feature extraction of the per million death time series, denote this collection of feature vectors $PMD(N)$
   d) Compute the spectral clustering of $SI(N)$ into C clusters. Each cluster contains countries with similar Stringency Index policies.
      i) Denote the collection of country clusters as $SI(N, C)$
   e) Compute the spectral clustering of $V(N)$ into C clusters. Each cluster contains countries with similar patterns of Vaccination policies.
      i) Denote the collection of country clusters as V(N,C)
   f) Compute the spectral clustering of $PMD(N)$ into C clusters. Each cluster contains countries with similar patterns of mortality along the pandemic.
      i) Denote the collection of country clusters as $PMD(N, C)$
   g) Compute the $C \times C$ matrix of intersections of clusters in $SI(N, C)$ with clusters in $PMD(N, C)$, denoted $\mathcal{I}[SI(N, C)]$.
      i) Each entry $\mathcal{I}[SI(N, C)]_{ij}$ contains the countries that are in cluster i of $SI(N, C)$ and in cluster j of $PMD(N, C)$.
   h) Perform the one-to-one assignment of $SI(N, C)$ clusters into $PMD(N, C)$ clusters) maximizing the sum of the cardinality of the intersections in the entries of $\mathcal{I}[SI(N, C)]$. Denote this assignment $\mathcal{A}[SI(N, C)]$.
   i) Compute the $C \times C$ matrix of intersections of clusters in $V(N, C)$ with clusters in $PMD(N, C)$, denoted $\mathcal{I}[V(N, C)]$.
      i) Each entry $\mathcal{I}[V(N, C)]_{ij}$ contains the countries that are in cluster i of $V(N, C)$ and in cluster j of $PMD(N, C)$.
   j) Perform the one-to-one assignment of $V(N, C)$ clusters into $PMD(N, C)$ clusters maximizing the sum of the cardinality of the intersections n the entries of $\mathcal{I}[V(N, C)]$. Denote this assignment $\mathcal{A}[V(N, C)]$.
2) Find the numbers of NNMF components N_{SI} and clusters C_{SI} with maximum value of the sum of cardinalities of the cluster assignments $\mathcal{A}[SI(N, C)]$.
3) Find the numbers of NNMF components N_V and clusters C_V with maximum value of the sum of cardinalities of the cluster assignments $\mathcal{A}[V(N, C)]$.
4) Find the number of components and clusters with maximum cardinality of the cluster assignments A_V_PMD(N,C), denote them N_V and C_V


Max Intersections for N=4:14; C=4:14
Deaths
Stringency
vaccines
NNMF into N components
NNMF into N components
NNMF into N components
Spectral Cluster into C clusters
Spectral Cluster into C clusters
Spectral Cluster into C clusters
Intersection Deaths & Stringency
Intersection Deaths & Stringency

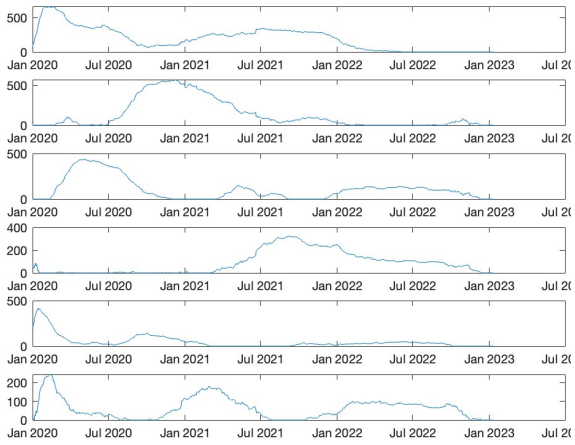
Results

$(C_{SI} = 4)$ $(N_{SI} = 6)$

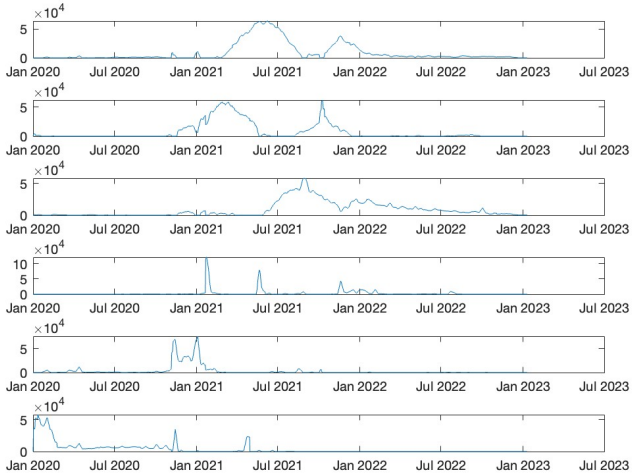
NNMF Mortality time series patterns



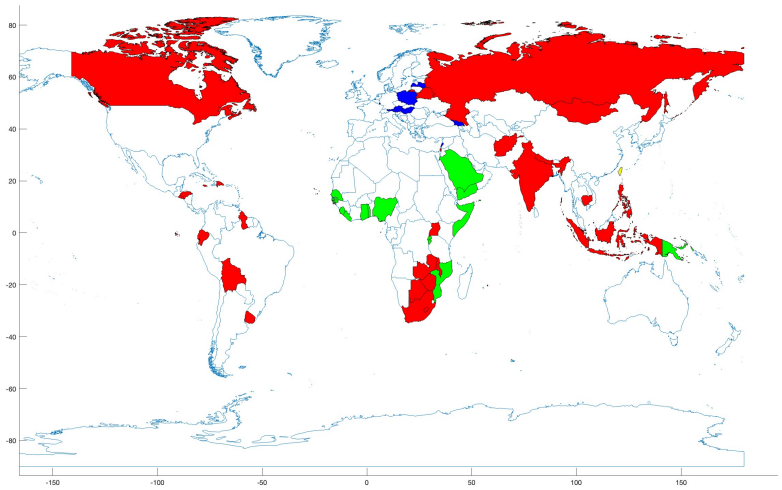
Stringency index patterns



Vaccine doses patterns

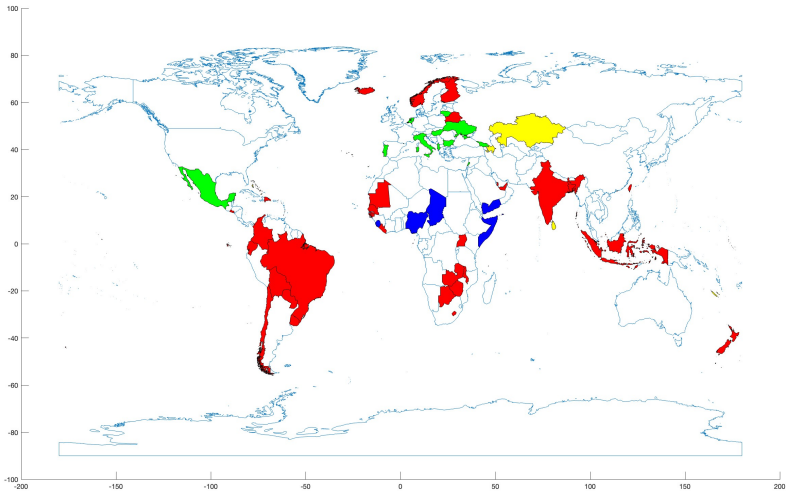


Results



Maximal intersection of clusters of **Stringency Index** and clusters of per million deaths. Countries without color do not appear in any intersection, therefore the non-pharmaceutical interventions in these countries appear to be **not associated** with the mortality.

Results



Maximal intersection of clusters of **vaccination doses** and clusters of per million deaths. Countries without color do not appear in any intersection, therefore the pharmaceutical interventions in these countries appear to be **not associated** with the mortality.

Discussion

- The maximal overall association of Stringency Index with mortality is relatively weak (70 of 200 countries), as it is the association of vaccination doses with mortality (75 out of 200 countries).
- These results point to the consideration of other factors, such as political, demographic, or cultural factors, to look for explanations of the actual mortality burden of the pandemic.

Discussion

- Regarding the association between stringency index and mortality,
  - In the western Europe countries apparently similar social control measures produced dissimilar mortality patterns, or the other way around.
  - Our interpretation is that, at least, there is no causal effect of the social measures and the actual mortality patterns.
  - In other words, the social control intervention appear to be irrelevant for the mortality outcomes in these countries.

Discussion

- Regarding the association between vaccine doses and mortality
  - Most countries showed no association
  - We can not assess the direction of the association between vaccination and mortality in the few intersection countries,
  - but we can infer that these countries achieved similar mortality results following similar temporal patterns of mass vaccination.
  - for most western European countries and the North America countries (Canada and the USA) the patterns of mass vaccination appear to be irrelevant for the mortality patterns.

Conclusions

- In summary:
  - After clustering the countries according to each of the kinds of time series (Stringency Index, Vaccination doses per million, deaths per million), we look for maximal assignments between clusters of each kind, maximizing cluster intersections.
  - The association strength can be measured by the number of countries found in these maximal assignments.
- Findings
  - We have found little overall association between intervention measures and mortality results, only 70 out of 200 countries.
  - This result should allow to question the blind repetition of assertions about the efficacy of intervention measures that can be found in the literature.
  - many countries show no association between intervention measures and mortality, hinting to the irrelevance of these measures to the actual evolution of the pandemic mortality toll in these countries.