

Computational modeling and data analysis of COVID19 outbreak

J. Tenreiro Machado*

* Institute of Engineering, Polytechnic of Porto, Dept. of Electrical Engineering,
Rua Dr. António Bernardino de Almeida, 431, 4249 – 015 Porto, Portugal

Abstract. This paper addresses the study COVID19 outbreak by means of three computational approaches. The first consists of using real-world data about the number of infected people around the world. The second analyses and compares the genetic code of several virus including the SARS-CoV-2. The third adopt evolutionary concepts to construct an evolutionary algorithm that somehow mimics real-world situations. In the three cases, hierarchical clustering and the multidimensional scaling allow a superior visualization of the results.

Introduction

The Coronavirus disease 2019 (COVID-19) outbreak is an example of an extreme event. This paper studies the dynamics of the spreading across different countries of COVID-19.

The first case of COVID-19 was officially reported in China on December 31, 2019, in Wuhan of Hubei province. At an early stage, the Chinese authorities seemed not to give importance to the problem. However, with the rapid emergence of new cases, the attitude changed dramatically. In the meantime, new cases have been emerging in many countries. In particular, the rapid evolution in Iran, Italy, Spain, Brazil and USA became the most dramatic cases. The analysis of the evolution of the confirmed cases versus time has considerable interest from the point of view of delivering good information to health organizations and to the general public. Several statistics have been presented, adopting different forms for organizing and visualizing the data. However, a comprehensive representation of the COVID-19 spreading dynamics is still missing.

Results and discussion

The COVID-19 data are made available by the European Centre for Disease Prevention and Control (<https://www.ecdc.europa.eu/en>). One strategy for processing data we use the Multidimensional scaling (MDS). The MDS is a computational technique for clustering and visualizing multidimensional data. Fig. 1 left shows the MDS plot the 79 countries using the item-to-item Canberra distance. A second strategy is to use the virus genomic sequences. The information of 133 publicly released genomic sequences was collected in the Global Initiative on Sharing Avian Influenza Data (GISAID) and the GenBank of the National Center for Biotechnology Information (NCBI) databases (<https://www.gisaid.org/>, <https://www.ncbi.nlm.nih.gov/genbank>). Fig. 1 right depicts MDS 3-dim locus for the set of 133 virus using the Normalized Compression Distance (NCD) and $zlib$ based on the Kolmogorov complexity theory with the cluster of SARS-CoV2 connected by a line.

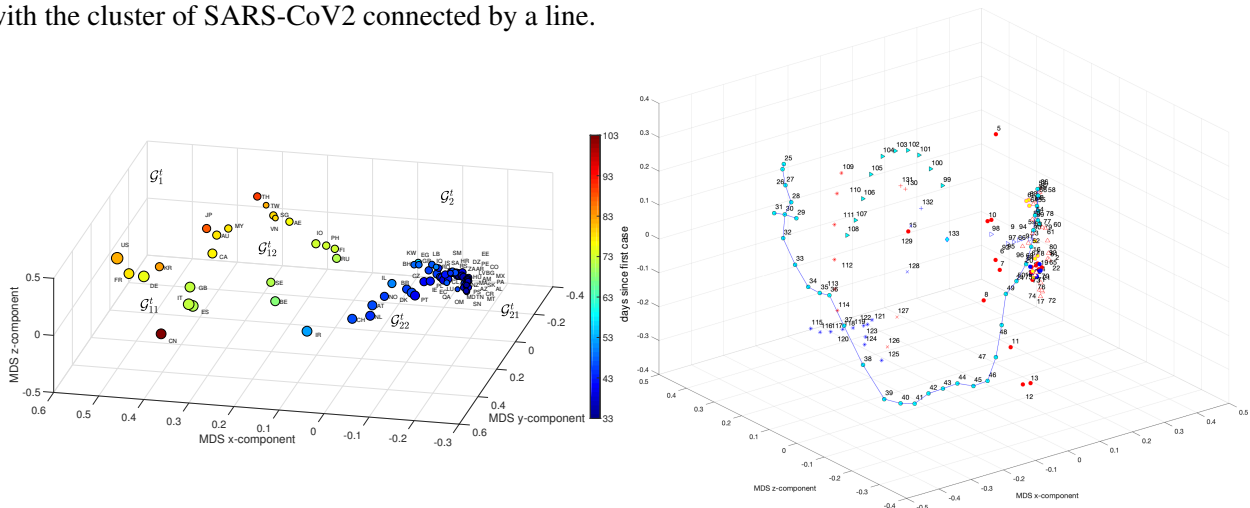


Figure 1: The 3D MDS locus of: (left) the 79 countries using the item-to-item Canberra distance in the time-domain. The size of the dots is proportional to the logarithm of the number of infections and the color is proportional to the time elapsed since the first reported case in each country from December 31, 2019, up to April 12, 2020, (right) the set of 133 virus using the NCD and $zlib$ based on the Kolmogorov complexity theory with the cluster of SARS-CoV2 connected by a line.

References

- [1] Machado J. T., Lopes A. (2020) Rare and extreme events: The case of COVID-19 pandemic. *Nonlinear Dynamics* **100**:2953-2972.
- [2] Machado J. T. (2020) An Evolutionary Perspective of Virus Propagation. *Mathematics* **8**:1-22.
- [3] Machado J. T., Rocha-Neves J.M., Andrade J.P. (2020) Computational analysis of the SARS-CoV-2 and other viruses based on the Kolmogorovs complexity and Shannons information theories. *Nonlinear Dynamics*.
- [4] Quaranta G., Formica G., Machado J. T., Lacarbonara W., Masri S.M (2020) Understanding COVID-19 nonlinear multi-scale dynamic spreading in Italy. *Nonlinear Dynamics*.