Multidimensional scaling analysis of virus diseases

António M. Lopes [a],*, José P. Andrade [b], J.A. Tenreiro Machado [c]

[a] *UISPA—LAETA/INEGI, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal*
[b] *Department of Anatomy, Faculty of Medicine, University of Porto, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal*
[c] *Institute of Engineering, Department of Electrical Engineering, Polytechnic of Porto, Rua Dr. António Bernardino de Almeida, 431, 4249-015 Porto, Portugal*

ARTICLE INFO

ABSTRACT

Background and Objective: Viruses are infectious agents that replicate inside organisms and reveal a plethora of distinct characteristics. Viral infections spread in many ways, but often have devastating consequences and represent a huge danger for public health. It is important to design statistical and computational techniques capable of handling the available data and highlighting the most important features.

Methods: This paper reviews the quantitative and qualitative behaviour of 22 infectious diseases caused by viruses. The information is compared and visualized by means of the multidimensional scaling technique.

Results: The results are robust to uncertainties in the data and revealed to be consistent with clinical practice.

Conclusions: The paper shows that the proposed methodology may represent a solid mathematical tool to tackle a larger number of virus and additional information about these infectious agents.

1. Introduction

Viruses exert enormous damage on humans worldwide and are the single most important cause of infectious morbidity and mortality. History was, and still is, shaped since ancient times by viral diseases. These diseases began to be characterized in the 19th century leading to the identification and differentiation of many viral illnesses [1]. The first viruses were identified at the end of the 19th century and since then the process of discovery has continued steadily with a growing momentum in these years. In fact, in recent years it is possible to visualize viral structure at an atomic level of resolution, nucleotide sequences of viral genomes are known, and functional domains of numerous viruses and enzymes have been established [1,2]. This information is now being applied to the development of diagnostic tools and effective antiviral therapies.

The classification of viruses has also evolved. Firstly, sub-classifications were based on pathologic features such as the preference of a specific organ (for example, the liver in viral hepatitis). Secondly, some epidemiologic characteristics were defined as the transmission by arthropods (arbovirus, for example) [1]. The current classifications are based on the type and structure of the viral nucleic acid and its replication strategy, the symmetry type of the capsid of the virus, and the presence or absence of a lipid envelop [1,2].

---

* *Corresponding author.* UISPA—LAETA/INEGI, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal. Tel.: +351 220413486; fax. +351 225081445.
E-mail address: aml@fe.up.pt (A.M. Lopes).

More than 2000 species of viruses have been identified and approximately 650 are capable of infecting humans and animals [2]. Diseases can range from the common cold to fatal events such as Ebola, Smallpox or Rabies [2]. Globally, viral diseases are very diverse and present several degrees of complexity.

In this study we will adopt multidimensional scaling (MDS) to visualize the relationships between 22 selected human viral infectious diseases. Some viruses were selected based on recent viral outbreaks and presence in the media (for example, Influenza A virus subtype H5N1, Ebola and Chikungunya), others were chosen due to historical reasons (for example, Rabies, Poliomyelitis and Smallpox), and still others due to their prevalence and incidence in human populations (for example, Influenza, Rhinovirus and Norovirus). In two viral diseases (Human Immunodeficiency Virus and Rabies) we consider both the treated and untreated paradigms of the disease due to the huge discrepancy in mortality.

MDS is proven to obtain a new perspective on visualizing global data associated with human pathologies. MDS is a set of techniques used to analyse similarities in data that produce spatial or geometric representations of complex objects [3–5]. MDS had its origin in behavioural sciences for its help in understanding judgements of individuals (as preference, or relatedness) concerning elements in a set of objects [6–8]. Nowadays, MDS is used with a large variety of real data, such as biological taxonomy [9–12], finance [13,14], marketing [15], sociology [16], physics [17], geophysics [18–20], communication networks [21,22], biology and biomedics [23,24], among others [25,26].

Bearing these ideas in mind, the paper is organized as follows. In Section 2 we present the MDS technique. In Section 3 we study and compare data regarding 22 virus diseases. Finally, in Section 4 we draw the main conclusions.

2. Multidimensional scaling

Given s objects in a m-dimensional space and a measure of proximity, δ_{ij}, between objects i and j, a symmetric $s \times s$ matrix, $\mathbf{C} = [\delta_{ij}]$, of item to item (dis)similarities is calculated in a first step. The MDS algorithm produces a $s \times q$ $(q < m)$ configuration, $\mathbf{X}$, representing point coordinates (items), where q is specified by the user. Thus, row i from matrix $\mathbf{X}$ gives the coordinates of object i in the q-dimensional embedding space. Configuration $\mathbf{X}$ preserves, as best as possible, the proximities between pairwise elements in the higher m-dimensional space and unveils the underlying data structure. MDS is, consequently, different from other similar techniques, such as factor and cluster analysis, because there are no assumptions concerning which factors might drive each dimension. Additionally, MDS is able to treat distinct types of data, has better convergence rates, and is less complex than other methods [3,27].

In order to arrive at the best configuration $\mathbf{X}$, MDS evaluates different alternative configurations while minimizing a goodness-of-fit function. This problem, equivalent to minimizing the raw stress function, σ^2, can be formulated as [28]:

$$\sigma^2 = \sum_{i=2}^{s} \sum_{j=1}^{i-1} z_{ij} \left(\delta_{ij} - d_{ij} \right)^2, \tag{1}$$

where z_{ij} is a user chosen non-negative weight and d_{ij} is a measure of the (dis)similarities among the items in the embedding space. Therefore, d_{ij} is usually a distance measure. Smaller (larger) distances between two objects translate into more (less) similarities between them. For example, the Minkowski distance provides a general way to specify distance for quantitative data in a multidimensional space:

$$d_{ij} = \left(\sum_{k=1}^{q} \alpha_k \left| x_{ik} - x_{jk} \right|^r \right)^{1/r}, \quad r \geq 1, \tag{2}$$

where x_{ik} is the value of dimension k for object i and α_k is a weight factor. When $\alpha_k = 1$, the Euclidean and the city-block distances are obtained for $r = 2$ and $r = 1$, respectively. Nevertheless, the MDS technique allows users to choose other metrics for the comparison of objects that can be better adequate for their data. In the sequel we will adopt the Canberra distance and the cosine correlation.

There are different stress measures, such as the normalized raw stress, which is σ^2 divided by the sum of squared dissimilarities. Possible alternatives are Kruskal's stress-1 and Kruskal's stress-2, which divide σ by the sum of squared distances, or by a function of the variances of distances, respectively. Another example is the S-stress measure given by the sum of squared errors between squared distances and squared dissimilarities [29,30].

The Shepard diagram is used to infer the quality of the MDS solution. Let p_{ij} denote the similarities between objects i and j. A Shepard diagram consists of pairs (p_{ij}, p_{ij}) and (p_{ij}, δ_{ij}). If a line connecting the pairs (p_{ij}, δ_{ij}) is drawn, then the approximation error, concerning dissimilarities of each object, is given by $d_{ij} - \delta_{ij}$. The Shepard diagram is thus useful for visualizing the residuals and outliers resulting from the MDS application to the data. A narrow scatter around the 45 degree line indicates a good fit between d_{ij} and δ_{ij}.

The stress plot represents σ^2 versus the number of dimensions q of the MDS maps. Usually, we get a monotonic decreasing chart and we choose q as a compromise between reducing σ^2 and having a low dimension for the MDS charts.

MDS can be divided according to the classification of data similarities, the number of similarity matrices and the nature of the MDS model. We thus have the non-metric, or metric MDS, if similarity data are qualitative or quantitative. In what concerns the number of similarity matrices and nature of the model we have classical MDS (i.e., with one matrix and unweighted models), replicated MDS (i.e., with several matrices and unweighted models) and weighted MDS (i.e., with several matrices and weighted models).

The MDS interpretation is based on the emerging clusters and distances between points in the map, rather than on their absolute coordinates, or the geometrical form of the locus. Thus, we can rotate or translate the MDS chart since the distances between points remain identical. Usually, two or three dimensional charts are selected, because they allow a direct graphical representation.

MDS has advantages over other methods, such as principal component analysis (PCA), since MDS can follow similarity/dissimilarity matrices based on several distinct metrics. MDS uses the inter-object distances rather than the coordinates of

the objects and, therefore, it turns out that the MDS is a more general method than PCA [18,31].

3.   Multidimensional scaling analysis

In this section we use MDS tools to visualize the relationships between $s = 22$ infectious diseases caused by viruses, namely Bird Flu (BFlu), Chicken Pox (CPox), Chikungunya (Chi), Dengue Fever (Den), Ebola (Ebo), Hepatitis B (HepB), HIV (HIV), HIV—untreated (HIV Un), Marburg virus disease (Mar), Measles (Mea), MERS (MERS), Mumps (Mum), Norovirus (Nor), Polio (Pol), Rabies (Rab), Rabies—untreated (Rab Un), Rhinovirus (Rhi), Rotavirus (Rot), Rubella (Rub), SARS (SARS), Seasonal Flu (SFlu) and Smallpox (Sma).

To the ith disease, $i = 1, \ldots, 22$, we associate $m_1 = 5$ quantitative and $m_2 = 2$ qualitative features. Therefore, we have a $m = m_1 + m_2 = 7$ dimensional space of attributes. We start by preprocessing the quantitative and the qualitative data, yielding a new equivalent μ-dimensional space (to be defined in the sequel) for disease comparison.

Table 1 lists the quantitative data, $\mathbf{U} = [u_{ig}]$, $i = 1, \ldots, 22$, $g = 1, \ldots, 5$. We consider the disease fatality rate, the average basic reproductive number, the average serial interval, the incubation period and the virus survival time outside the host.

Tables 2 and 3 summarize the qualitative features, namely the transmission mode, and the main symptoms of the disease.

The transmission mode dimension is represented as binary data (Table 2), $\mathbf{V} = [v_{ih}]$, $i = 1, \ldots, 22$, $h = 1, \ldots, 7$. This means that we consider $t = 7$ conditions, specifically animal–human, airborne droplet, bites, body fluids, sexual contact, surfaces and

faecal–oral. The value $v_{ih} = 1$ means that disease i has the transmission mode h, and $v_{ih} = 0$ means that disease i does not have the transmission mode h.

In a similar way, for the symptoms dimension we consider $y = 28$ indicators, represented as binary data $\mathbf{W} = [w_{il}]$, $i = 1, \ldots, 22$, $l = 1, \ldots, 28$ (Table 3). The value $w_{il} = 1$ (or 0) means that disease i has (or has not) symptom l.

Before applying the MDS algorithm we start by "normalizing" the quantitative data to the interval $[0, 1]$, i.e., $\tilde{u}_{ig} = u_{ig}/\max_i\{u_{ig}\}$. For the qualitative data, we apply $\tilde{v}_{ih} = v_{ih}/t$ and $\tilde{w}_{il} = w_{il}/y$, meaning that $\sum_h \tilde{v}_{ih} \in [0, 1]$ and $\sum_l \tilde{w}_{il} \in [0, 1]$. In this way we avoid having some features saturating the numerical values.

We proceed by constructing the vectors of features $\mathbf{x}_i = \{\tilde{u}_{ig}, \tilde{v}_{ih}, \tilde{w}_{il}\}$, yielding a μ-dimensional space ($\mu = m_1 + t + y$) that embeds all quantitative and qualitative data. This is equivalent to the m-dimensional space defined previously for disease comparison.

In the next subsections we use two indices to compare the preprocessed data, namely the Canberra distance, c_{ij}, and the cosine correlation, cc_{ij}. Other indices can be adopted, but these two are sufficient to explain the working concepts. We then apply the MDS technique and interpret the generated maps. Fig. 1 depicts a synoptic diagram of the disease's characteristics and quantification method.

In constructing Tables 1–3, data were collected from the following sources: Influenza A virus subtype H5N1 commonly known as "Bird Flu" [32–35]; Chicken Pox (varicella-zoster virus infection) [36–39]; Chikungunya [40–42]; Dengue Fever [43,44]; Ebola [45–47]; Hepatitis B [48–50]; Human Immunodeficiency Virus (HIV) and HIV—untreated [51–54]; Marburg haemorrhagic fever [47,55]; Measles [56–60]; Middle East Respiratory

Table 1 – Quantitative attributes of the diseases considered in the study.						
	Disease	Acronym	Fatality rate (%)	Average basic reproductive number	Average serial interval (days)	Incubation period (days)	Survival outside host (days)
i		g	1	2	3	4	5
1	Bird flu	BFlu	59.00	2.00	3.00	3.0	30.0
2	Chicken pox	CPox	0.00	7.50	14.00	14.0	2.0
3	Chikungunya	Chi	0.40	4.00	23.00	2.5	–
4	Dengue fever	Den	5.00	3.00	16.00	7.0	63.0
5	Ebola	Ebo	75.00	2.50	15.30	11.4	50.0
6	Hepatitis B	HepB	0.75	6.00	25.00	75.0	28.0
7	HIV	HIV	2.10	3.50	–	60.0	42.0
8	HIV—untreated	HIV Un	80.00	3.50	–	60.0	42.0
9	Marburg virus disease	Mar	25.00	1.60	9.00	6.0	21.0
10	Measles	Mea	0.20	15.00	11.70	11.0	0.1
11	MERS	MERS	27.00	0.50	7.60	5.0	3.0
12	Mumps	Mum	0.01	5.50	18.00	17.0	0.3
13	Norovirus	Nor	0.08	3.70	1.86	1.5	24.0
14	Polio	Pol	22.00	6.00	–	13.0	160.0
15	Rabies	Rab	0.00	1.60	–	40.0	6.0
16	Rabies—untreated	Rab Un	100.00	1.60	–	40.0	6.0
17	Rhinovirus	Rhi	0.00	3.70	7.50	3.0	1.0
18	Rotavirus	Rot	0.00	3.50	7.00	1.5	60.0
19	Rubella	Rub	0.00	6.50	18.30	17.7	0.9
20	SARS	SARS	11.00	3.50	10.00	8.0	9.0
21	Seasonal flu	SFlu	0.01	1.30	3.30	2.0	2.0
22	Smallpox	Sma	15.00	6.00	17.70	14.0	1.5

Table 2 – Main transmission mode of the diseases considered in the study.								
	Disease	Acronym	Animal–human	Airborne droplet	Bites	Body fluids	Sexual contact	Surfaces	Faecal–oral
i		h	1	2	3	4	5	6	7
1	Bird flu	BFlu	1	0	0	0	0	0	0
2	Chicken pox	CPox	0	1	0	0	0	0	0
3	Chikungunya	Chi	0	0	1	0	0	0	0
4	Dengue fever	Den	0	0	1	0	0	0	0
5	Ebola	Ebo	0	0	0	1	0	0	0
6	Hepatitis B	HepB	0	0	0	1	1	0	0
7	HIV	HIV	0	0	0	1	1	0	0
8	HIV—untreated	HIV Un	0	0	0	1	1	0	0
9	Marburg virus disease	Mar	0	0	0	1	0	0	0
10	Measles	Mea	0	1	0	0	0	0	0
11	MERS	MERS	0	1	0	0	0	0	0
12	Mumps	Mum	0	1	0	0	0	1	0
13	Norovirus	Nor	0	0	0	0	0	1	1
14	Polio	Pol	0	0	0	0	0	0	1
15	Rabies	Rab	0	0	1	0	0	0	0
16	Rabies—untreated	Rab Un	0	0	1	0	0	0	0
17	Rhinovirus	Rhi	0	1	0	0	0	0	1
18	Rotavirus	Rot	0	0	0	0	0	1	1
19	Rubella	Rub	0	1	0	0	0	0	0
20	SARS	SARS	0	1	0	0	0	0	1
21	Seasonal flu	SFlu	0	1	0	0	0	0	0
22	Smallpox	Sma	0	1	0	0	0	0	0

Syndrome (MERS) [61–63]; Mumps [64,65]; Norovirus [66,67]; Poliomyelitis [68–70]; Rabies and Rabies—untreated [71–74]; Rhinovirus [75–77]; Rotavirus [78–81]; Rubella [58,82]; Severe Acute Respiratory Syndrome (SARS) [61,83]; Influenza causing seasonal flu [34,84,85]; and Smallpox [86,87].

3.1. MDS analysis based on the Canberra distance

In this subsection we consider the construction of matrix $\mathbf{X}$ using a measure based on the Canberra distance, c_{ij}, between diseases i and j $(i, j = 1, \ldots, s)$:

$$c_{ij} = \sum_{k=1}^{m_1+t+y} \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|}. \tag{3}$$

Given this index, the $s \times s$ symmetric matrix, $\mathbf{C} = \left[1 - c_{ij} / \max_{i,j}\{c_{ij}\}\right]$, is then computed and the MDS tool applied. While several MDS criteria were tested, the Sammon criterion revealed good results and was adopted in all calculations. It should be noted that this criterion tries to optimize a cost function that describes how well the pairwise distances in a data set are preserved [88,89].

Fig. 2 depicts the 2- and 3-dimensional (2D and 3D) maps produced by MDS. Each point represents a disease, denoted by the corresponding label as shown in Tables 1–3. We can observe that the Canberra index leads to poor clustering. Nonetheless, we should note that MDS is merely a mathematical clustering and visualization tool and that a physical perspective of the reported results must be found in the light of the comparison index [90]. Therefore, a further explanation about physical mechanisms associated with the results must be envisaged by standard complementary procedures.

Figs. 3 and 4 depict the Shepard and stress plots, respectively, which represent standard tools for the assessment of the MDS results. The Shepard diagram shows a good distribution of points around the 45 degree line, particularly for the 3D representation, which means a good fit of the distances to the dissimilarities. The stress plot reveals that a three dimensional space describes well the locus of the $s = 22$ diseases. In fact, the stress diminishes strongly until $q = 2$, moderately towards $q = 3$ and weakly. The maximum curvature point of the stress plot is often adopted as the criterion for deciding the dimensionality of the MDS maps. This means that, although four or more dimensions would represent the data slightly more accurately, 3D maps represent a good compromise between accuracy and easiness of visualization.

3.2. MDS analysis based on the cosine correlation

In this subsection we adopt the cosine correlation, cc_{ij}, to construct the matrix $\mathbf{X}$. For each disease pair i and j $(i, j = 1, \ldots, s)$ we have:

$$cc_{ij} = \frac{\sum_{k=1}^{m_1+t+y} \alpha_k^2 \left(x_{ik} \cdot x_{jk}\right)}{\sqrt{\sum_{k=1}^{m_1+t+y} \alpha_k^2 x_{ik}^2 \cdot \sum_{k=1}^{m_1+t+y} \alpha_k^2 x_{jk}^2}}, \tag{4}$$

where $\alpha_k = \{\alpha_1, \ldots, \alpha_{m_1} | \alpha_{m_1+1}, \ldots, \alpha_{m_1+t} | \alpha_{m_1+t+1}, \ldots, \alpha_{m_1+t+y}\}$ represents weights, specified by the user, which are usually chosen to favour adequate clustering. Given expression (4) the $s \times s$ symmetric matrix, $\mathbf{C} = [cc_{ij}]$, is computed and the MDS is applied.

Fig. 5 represents the 2D and 3D maps resulting from the MDS analysis for $\alpha_k = \{5 \quad 2 \quad 1 \quad 1 \quad 1 | \alpha_p \cdot (1 \ldots 1) | \alpha_q \cdot (1 \ldots 1)\}$, $\alpha_p = 0.5$, $\alpha_q = 2$. The Shepard and stress plots are identical to the maps shown in Figs. 3 and 4, revealing, as before, that the MDS results

– Main symptoms of the diseases considered in the study.

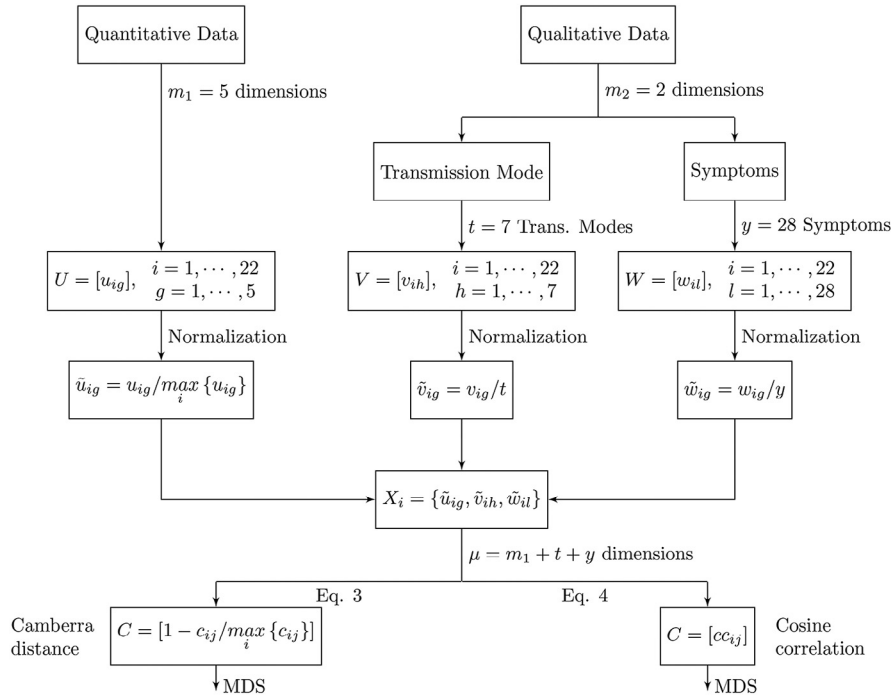| i	Disease	Acronym	Fever	Cough	Sore throat	Muscle and body aches	Nausea	Abdominal pain	Diarrhoea	Vomiting	Rash	Itchy	Fluid blisters	Tiredness, fatigue, weakness	Loss of appetite	Headache	Joint pain	Joint swelling	Bleeding	Jaundice	Swollen glands	Chills	Chest pain	Runny nose	Conjunctivitis	Red spots	Shortness of breath	Confusion	Agitation	Sneezing
i		l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	Bird flu	BFlu	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Chicken pox	CPox	1	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Chikungunya	Chi	1	0	0	1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	Dengue fever	Den	1	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
5	Ebola	Ebo	1	0	0	1	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	Hepatitis B	HepB	1	0	0	0	1	1	0	1	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
7	HIV	HIV	1	0	1	1	0	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
8	HIV—untreated	HIV Un	1	0	1	1	0	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
9	Marburg virus disease	Mar	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	0	1	1	1	0	0	0	0	0	0	0
10	Measles	Mea	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
11	MERS	MERS	1	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
12	Mumps	Mum	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
13	Norovirus	Nor	1	0	0	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	Polio	Pol	1	0	1	0	1	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	Rabies	Rab	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
16	Rabies—untreated	Rab Un	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
17	Rhinovirus	Rhi	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
18	Rotavirus	Rot	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	Rubella	Rub	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
20	SARS	SARS	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	Seasonal flu	SFlu	1	1	1	1	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
22	Smallpox	Sma	1	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 1 – Synoptic diagram of the disease's characteristics and quantification method. The indices denote: i—disease; g—quantitative data; h—qualitative data, transmission mode; l—qualitative data, symptoms.

are trustworthy. We observe now the emergence of a different pattern, but the main idea of clustering remains. This observation is usual in MDS charts, where alternative indices capture different characteristics of the phenomena and lead to distinct plots, but allowing the same conclusions. The "best" index is simply the one that produces a MDS map where clusters reflect real-world in a more direct perspective.

3.3. Clustering analysis

The standard MDS analysis is based on the object groups in the final map. We can rely either in the direct visualization of the plot, or in the implementation of some extra algorithm to extract the clusters.

Bearing this idea in mind, in subsection 3.3.1 we adopt the non-hierarchical clustering algorithm K-means to identify clusters in the MDS map. In subsection 3.3.2 we use hierarchical clustering to confirm the results obtained. In subsection 3.3.3 we analyse the sensitivity of the MDS maps. In subsection 3.3.4 we discuss the results.

We restrict the analysis to the cosine correlation metric since it revealed better results.

3.3.1. K-means clustering

Clustering consists on grouping objects that are, in some sense, similar to each other. The K-means is a non-hierarchical clustering method commonly used in machine learning and data mining [91]. The algorithm starts with a collection of s objects, where each object is a point in a q-dimensional space, and a given number of clusters, K, specified in advance by the user. The K-means groups the s objects into $K \le s$ clusters, so as to

minimize the objective function given by the sum of distances between the points and the centres of their clusters. The K-means arrives at a solution in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible.

A key issue in K-means is how to determine the correct number of clusters, K. It should be noted that the very notion of "good clustering" is subjective and is a question of point of view. However, we can rely on different indices to measure the quality of the clustering, namely the Davies–Bouldin, the Caliński–Harabasz and the silhouette indices [92–94].

In this work we adopt the silhouette to compare different solutions. The silhouette value, S, for each object, is a measure of how well each object lies within its cluster. Silhouette values vary in the interval $S \in [-1, 1]$. Silhouette values closer to $S = 1$ correspond to objects that are very distant from neighbouring clusters and, therefore, they are assigned to the right cluster. If $S = 0$, then the objects could be assigned to another cluster. When $S = -1$, then the objects are correctly assigned.

Given the coordinates of the $s = 22$ objects in the $q = 3$ dimensional space generated by the MDS, we evaluate the clusters identified by the K-means algorithm when varying the number of clusters in the interval $K \in [2, 7]$.

Fig. 6 depicts the silhouette average values versus the number of clusters, K, leading to the optimum value $K = 4$. Fig. 7 illustrates the shape of the silhouettes obtained for $K = 3, \ldots, 6$, where we can see that the best shape is obtained for $K = 4$.

For $K = 4$, the K-means generates the clusters $\mathcal{A} = \{$CPox, Chi, Mea, Mum, Nor, Rhi, Rot, Rub, SFlu$\}$, $\mathcal{B} = \{$HepB, HIV, Rab$\}$, $\mathcal{C} = \{$Den, Pol, SARS, Sma$\}$ and $\mathcal{D} = \{$BFlu, Ebo, HIV Un, Mar, MERS, Rab Un$\}$, to be discussed in subsection 3.3.4.
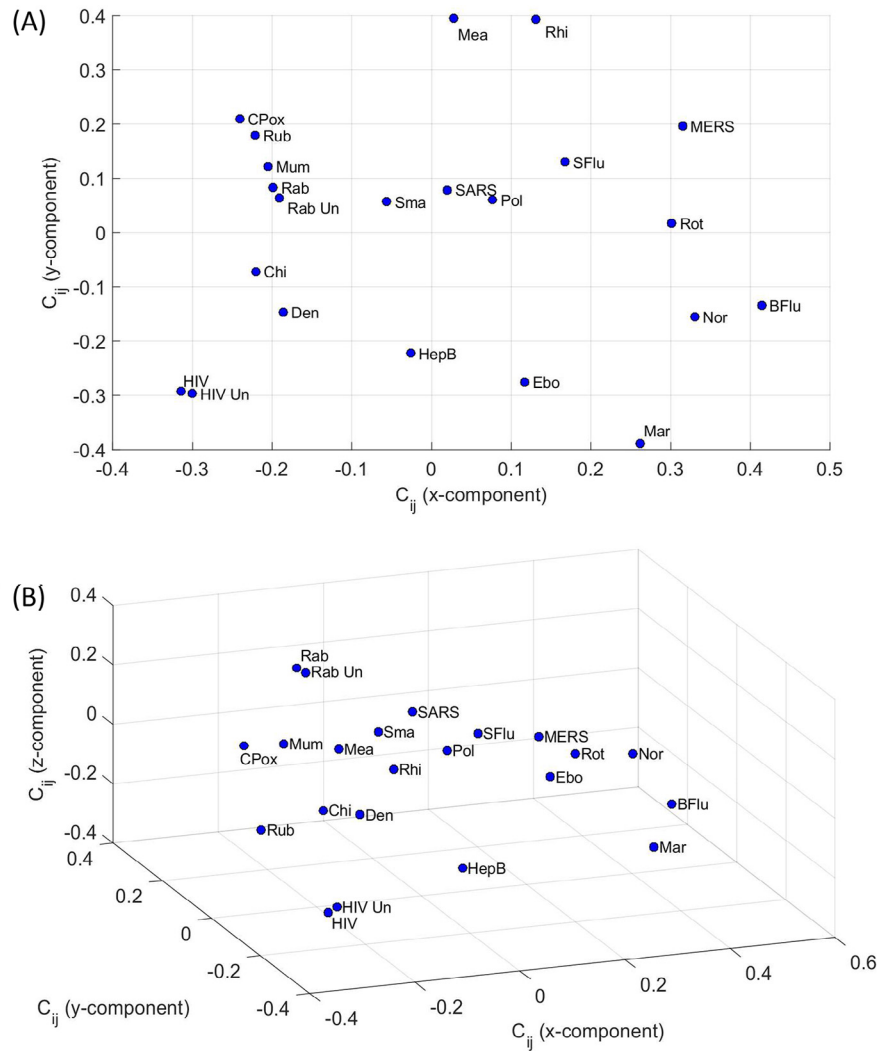
Fig. 2 – MDS maps for the Canberra index c_{ij} with representations: (A) 2D; (B) 3D.

3.3.2. Hierarchical clustering

As an alternative approach, not involving the MDS, we use a hierarchical clustering algorithm that is fed directly with matrix $\mathbf{C} = [cc_{ij}]$.

Fig. 8 depicts the dendrogram generated by successive (agglomerative) clustering and average-linkage method [95,96]. We cut the tree at the level 0.27, since below this value we see that the clusters became too close from each other. We see that four clusters emerge with this method, confirming the results obtained in the previous subsection. Nevertheless, MDS uses more efficiently the space and produces charts with a more fruitful map for the objects.

Fig. 3 – Shepard plots for the Canberra index c_{ij} for representations: (A) 2D; (B) 3D.

Fig. 4 – Stress versus q plot for the Canberra index c_{ij}.

3.3.3. Sensitivity analysis

The $s = 22$ virus diseases were compared based on quantitative and qualitative features. As qualitative characteristics are subjective, their influence upon the final results needs to be analysed, so as to prevent biased conclusions.

In this line of thought, we vary the weights, $\{\alpha_p, \alpha_q\}$, of the two qualitative features in the interval $\{\alpha_p, \alpha_q\} \in \{\alpha_p, \alpha_q\} \pm 100\%$ and we check their influence in the generated MDS map. Each weight is discretized into r distinct values evenly spaced in the intervals and then $r \times r$ instances of the $s \times s$ dimensional similarity matrix $\mathbf{C} = [cc_{ij}]$ are calculated. We use these matrices as the input for the MDS algorithm that generates $r \times r$ intermediate maps of "points" (i.e., one map per $\{\alpha_p, \alpha_q\}$ pair). Finally, the charts are processed by means of Procrustes analysis in order to obtain a single global plot of "shapes", where the "points" of the original maps are optimally superimposed [97].

Procrustes analysis performs linear transformations, namely translation, reflection, orthogonal rotation and scaling, with the objective of minimizing a measure of the difference between the "points" in the original maps. The algorithm (i) chooses one MDS map for reference (by selecting one of the available instances); (ii) superimposes all other MDS instances into the current reference; (iii) computes the mean form of the current set of superimposed maps; (iv) compares the distance between the mean and the reference instances to a given threshold value and, if above, sets the reference to the mean form and continues to step (ii).
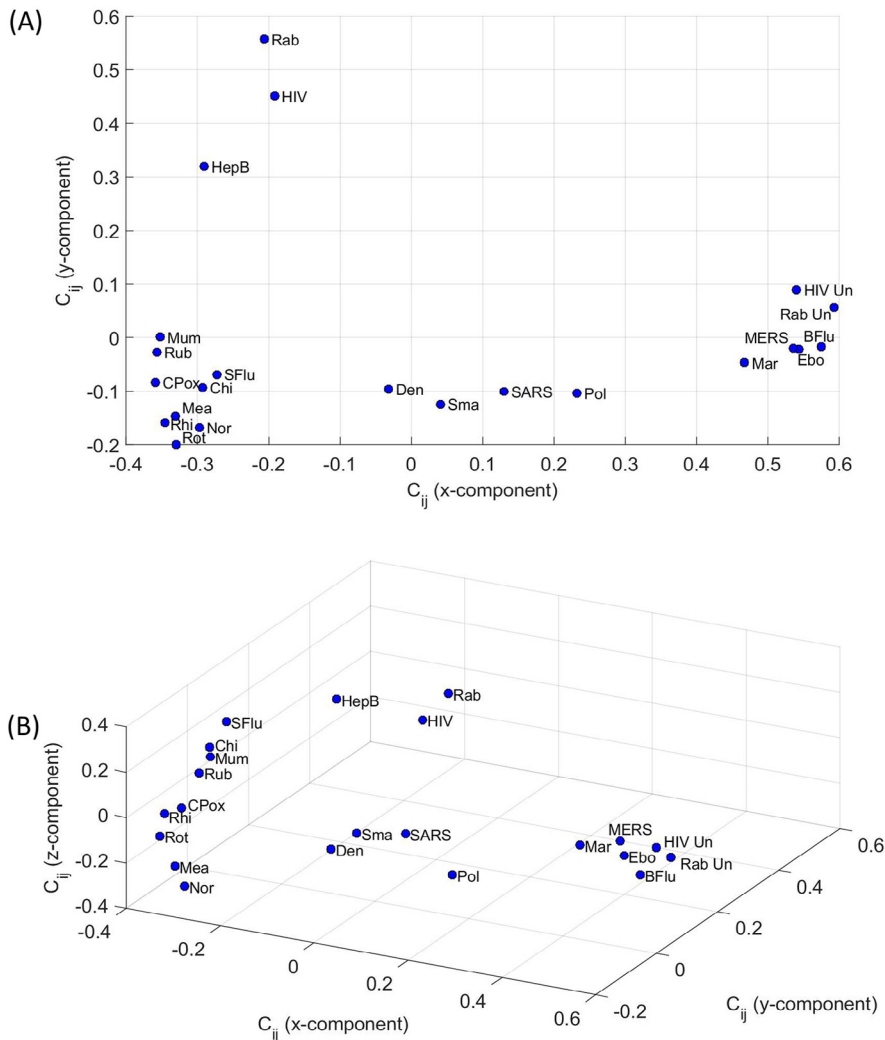

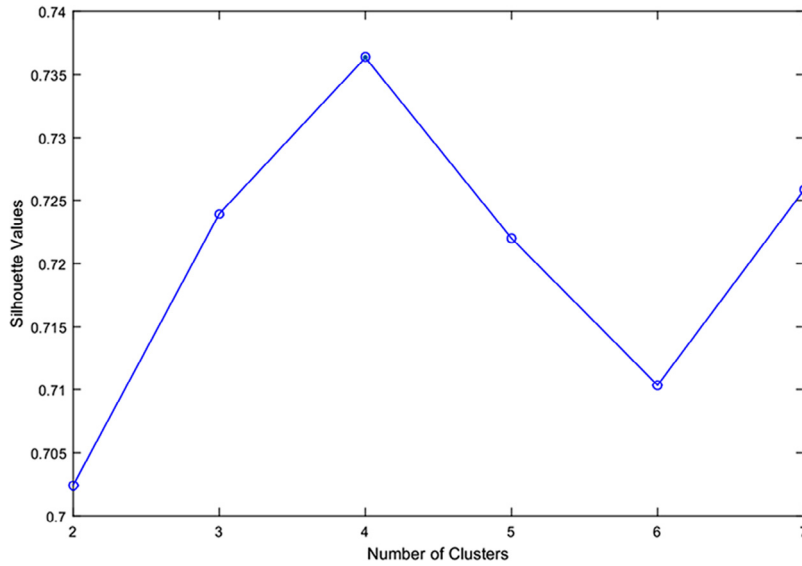**Fig. 5 – MDS maps for the cosine correlation cc_{ij} with representations: (A) 2D; (B) 3D.**

Fig. 6 – Silhouette average values versus number of clusters K, for K ∈ [2, 7].

Fig. 9 shows the 3D MDS global map obtained by the Procrustes algorithm, as well as the clusters $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$ identified previously. As can be seen, the results are quite robust to large variations in the qualitative features, since the $r \times r$ points corresponding to each original object (disease) deviate somehow, but the clusters remain.

In a complementary perspective we address in the sequel the sensitivity of the MDS results to quantitative features. In fact, the quantitative values found in the literature diverge slightly, since they depend on the time of the study and on the conditions observed in each particular case, namely environmental conditions (e.g., temperature, humidity, pressure), geographic region, human development, medical assistance, among others.

To assess the sensitivity we add random noise to the values of the quantitative features, with amplitude in the interval ±10% of the values shown in Table 1 (values are limited to zero to avoid negative numbers). In these conditions, we perform ten experiments, generating one MDS map per trial, and then the MDS individual maps are combined using Procrustes.
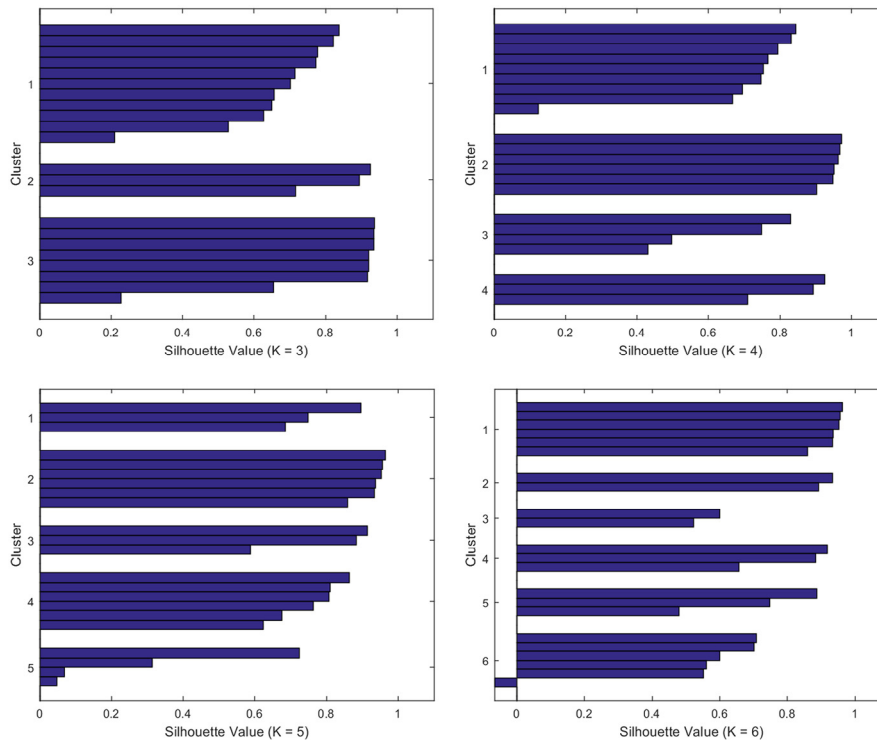

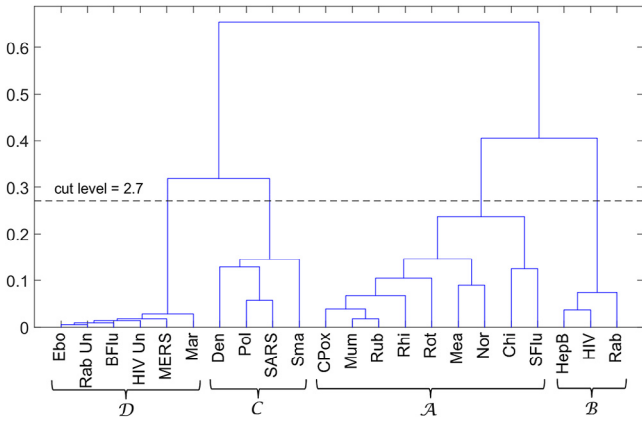**Fig. 7 – Silhouette shape for K = 3, … , 6.**

Fig. 8 – Dendrogram generated by the hierarchical clustering algorithm for the cosine correlation cc_{ij} index.

Fig. 10 illustrates the 3D MDS global map obtained and the corresponding superimposed clusters $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$. We conclude that the method is robust to variations in the values of the quantitative features used in the study.

3.3.4. Discussion

The clusters $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$ have interesting medical and epidemiological value. In cluster $\mathcal{D}$, we find the Ebola and the Marburg viruses, which can cause serious, or most of the times, a lethal human disease, even with therapeutic interventions. According to the National Institutes of Health (NIH) Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules (http://osp.od.nih.gov/sites/default/files/NIH_Guidelines.html#_Toc351276292) of the USA, from November 2013, those viruses are classified as Risk Group 4 pathogens. There are no vaccines or effective treatments for their
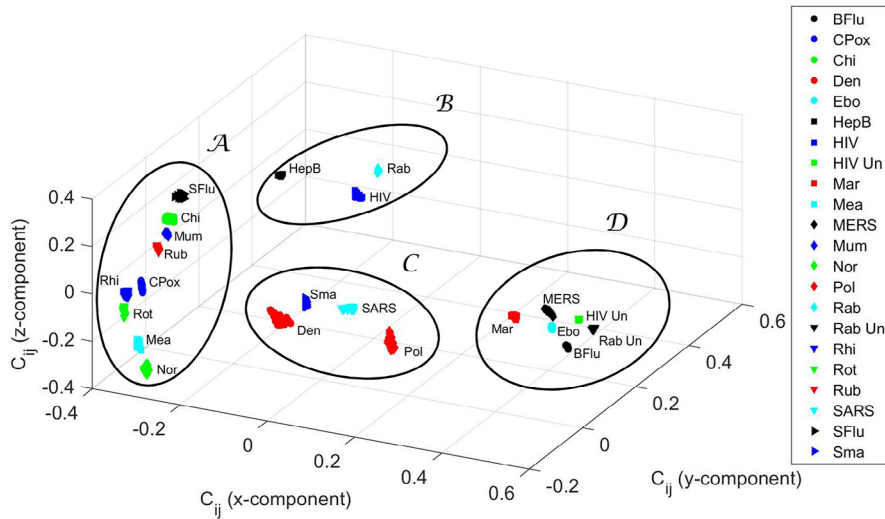

Fig. 9 – Three-dimensional MDS global map for the cosine correlation cc_{ij}, obtained by Procrustes with $\{\alpha_p, \alpha_q\} \in \{\alpha_p, \alpha_q\} \pm \mathbf{100\%}$.
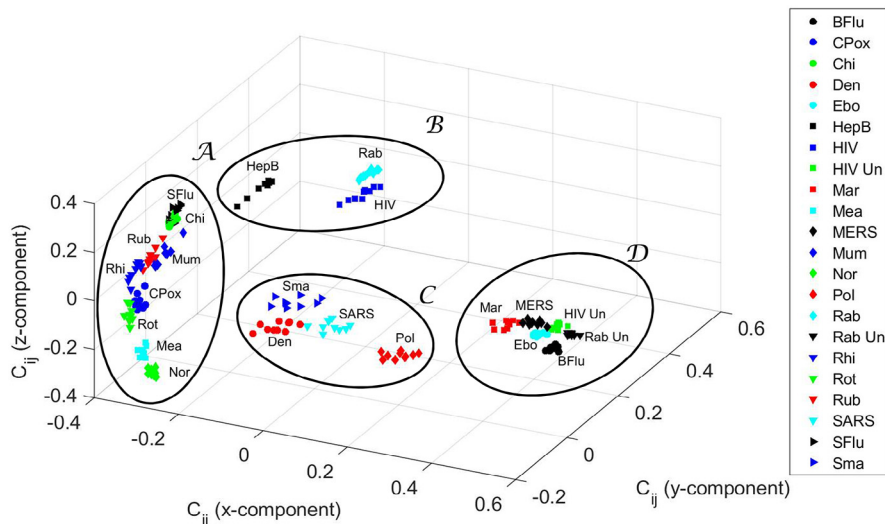

Fig. 10 – Three-dimensional MDS global map for the cosine correlation cc_{ij}, obtained by Procrustes with random variations in the values of the quantitative features.

infections. Accordingly, these viruses are manipulated only in Biosafety Level 4 conditions due to their high individual and community risk. They are also considered biological agents with material threat determinations in the scope of bioterrorism, in the USA [98]. Also, the highly infectious agents responsible for MERS (from the Coronaviridae Subfamily) and Bird flu (Influenza A virus subtype H5N1), classified as Risk Group 3, are mapped in $\mathcal{D}$. These viruses are recommended to be manipulated with Biosafety Level 3 precautions, indicated for agents that may cause serious or potentially lethal disease. In this cluster, untreated HIV and untreated rabies infections are also present.

In contrast, and distant in the MDS map, we have cluster $\mathcal{A}$. Most of the viruses are in Risk Group 2, because they generally do not cause serious or life threatening illness and most of them are readily treated or prevented easily with vaccines. They are manipulated, as most viruses, in Biosafety Level 2 environments. There is the exception of the Chikungunya, an important cause of febrile illness in the world, and now re-emerging as cause of large outbreaks of human disease [99]. The arbovirus (arthropod-borne) alphavirus responsible for this pathology is considered a Risk Group 3 pathogen and requires Biosafety Level 3 precautions [100].

Cluster $\mathcal{C}$ presents several virus species of different Risk Groups. Polio virus is a Risk Group 2 pathogen as well as the dengue fever virus, an arbovirus. On the other hand, the SARS-associated coronavirus is a pathogen of the Risk Group 3. Furthermore, smallpox, the disease caused by the variola virus, is also present in $\mathcal{C}$. Variola is considered a life-threatening disease posing the highest risk to national security due to its potential use as a biological weapon due to the high mortality rates and the major public health impact [98]. The reason is that smallpox was declared eradicated by the World Health Organization (WHO) in 1980, and vaccination, once widely practised, stopped in the same year [98]. Therefore, cluster $\mathcal{C}$ can be considered as a transition cluster from $\mathcal{A}$ to $\mathcal{D}$. In other words, it is located in the MDS map "equidistant" from $\mathcal{A}$ and $\mathcal{D}$.

Identical reasoning can be applied to cluster $\mathcal{B}$. In this cluster are the lentivirus (a subgroup of retrovirus) that causes HIV infection and acquired immunodeficiency syndrome (AIDS), a Risk Group 3 pathogen but generally manipulated with Biosafety Level 2 precautions, and the hepatitis B virus, considered belonging to Risk Group 2, equally transmitted by body fluids, and requiring a Biosafety Level 2 environment. Also considered of the Risk Group 2 is the third pathogen present in $\mathcal{B}$, i.e., the virus of Lyssavirus genus of the Rhabdoviridae family that causes rabies. Note that if the diseases caused by HIV and rabies are non-treated, then there is a high lethality in humans. They emerge in cluster $\mathcal{D}$ of the MDS map.

In conclusion, the MDS map resulted in a new visualization of the complex quantitative and qualitative data of several diseases caused by viruses, and several clusters were organized having some medical and epidemiological interest. In particular, a cluster emerged with viruses like Ebola and MERS, which are responsible for some recent viral outbreaks. In contrast, in the same MDS map, and distant from the previous group, there is a cluster of viruses associated with human diseases that present generally preventive and therapeutic interventions. The development of this methodology may help in understanding the dynamics of viral diseases.

4. Conclusion

This paper addressed the clinical characteristics of 22 viruses. A significant number of quantitative and qualitative characteristics were considered. When handling a large volume of information, we are confronted with the problem of comparing all details, but highlighting the most important properties. Discharging information a priori may lead to incomplete or even biased results. Therefore, embedding all details requires adequate statistical, computational and visualization techniques capable of revealing the main aspects while "filtering" the information with low relevance. The MDS technique adopted in this study proved to produce solid results in accordance with present day knowledge about those infectious agents.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

[1] J.E. Bennett, R. Dolin, M.J. Blaser, Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, Elsevier Health Sciences, Philadelphia, 2014.

[2] P.R. Murray, K.S. Rosenthal, M.A. Pfaller, Medical Microbiology, Elsevier Health Sciences, Philadelphia, 2013.

[3] I. Borg, P.J. Groenen, Modern Multidimensional Scaling: Theory and Applications, Springer Science & Business Media, Springer-Verlag, New York, 2005.

[4] T.F. Cox, M.A. Cox, Multidimensional Scaling, CRC Press, Boca Raton, 2000.

[5] M.L. Davison, Multidimensional Scaling, vol. 85, Wiley, New York, 1983.

[6] M. Fenton, P. Pearce, Multidimensional scaling and tourism research, Annal. Tour. Res. 15 (2) (1988) 236–254.

[7] H.T. Lawless, N. Sheng, S.S. Knoops, Multidimensional scaling of sorting data applied to cheese perception, Food Qual. Prefer. 6 (2) (1995) 91–98.

[8] W.S. Torgerson, Theory and Methods of Scaling, Wiley, New York, 1958.

[9] I.S. Caballero, J.Y. Yen, L.E. Hensley, A.N. Honko, A.J. Goff, J.H. Connor, Lassa and Marburg viruses elicit distinct host transcriptional responses early after infection, BMC Genomics 15 (1) (2014) 960.

[10] J. Kim, Y. Ahn, K. Lee, S.H. Park, S. Kim, A classification approach for genotyping viral sequences based on multidimensional scaling and linear discriminant analysis, BMC Bioinformatics 11 (1) (2010) 434.

[11] D.A. Lacher, P.F. Lehmann, Application of multidimensional scaling in numerical taxonomy: analysis of isoenzyme types of Candida species, Ann. Clin. Lab. Sci. 21 (2) (1991) 94–103.

[12] P.R. Vukea, S. Willows-Munro, R.F. Horner, T.H. Coetzer, Phylogenetic analysis of the polyprotein coding region of an infectious South African bursal disease virus (IBDV) strain, Infect. Genet. Evol. 21 (2014) 279–286.

[13] J. Tenreiro Machado, F.B. Duarte, G.M. Duarte, Analysis of financial data series using fractional Fourier transform and multidimensional scaling, Nonlin. Dynam. 65 (3) (2011) 235–245.

[14] J.A. Tenreiro Machado, A.M. Lopes, A.M. Galhano, Multidimensional scaling visualization using parametric similarity indices, Entropy 17 (4) (2015) 1775–1794.

[15] I. Cil, Consumption universes based supermarket layout through association rule mining and multidimensional scaling, Expert Syst. Appl. 39 (10) (2012) 8611–8625.

[16] R. Corten, Visualization of social networks in Stata using multidimensional scaling, Stata J. 11 (1) (2011) 52.

[17] K. Hollemeyer, W. Altmeyer, E. Heinzle, C. Pitra, Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry combined with multidimensional scaling, binary hierarchical cluster tree and selected diagnostic masses improves species identification of Neolithic keratin sequences from furs of the Tyrolean Iceman Oetzi, Rapid Commun. Mass Spectrom. 26 (16) (2012) 1735–1745.

[18] A.M. Lopes, J.A. Tenreiro Machado, Analysis of temperature time-series: embedding dynamics into the MDS method, Commun. Nonlinear Sci. Numer. Simul. 19 (4) (2014) 851–871.

[19] A.M. Lopes, J.A. Tenreiro Machado, C.M. Pinto, A.M. Galhano, Fractional dynamics and MDS visualization of earthquake phenomena, Comput. Math. Appl. 66 (5) (2013) 647–658.

[20] J.A. Tenreiro Machado, A.M. Lopes, Analysis and visualization of seismic data using mutual information, Entropy 15 (9) (2013) 3892–3909.

[21] F. Chan, H.-C. So, Efficient weighted multidimensional scaling for wireless sensor network localization, IEEE Trans. Signal Process. 57 (11) (2009) 4548–4553.

[22] X. Ji, H. Zha. Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling. In INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, volume 4, pages 2652–2661. IEEE, 2004.

[23] C.M. Ionescu, The Human Respiratory System: An Analysis of the Interplay between Anatomy, Structure, Breathing and Fractal Dynamics, Springer Science & Business Media, 2013.

[24] C. Tzagarakis, T.A. Jerde, S.M. Lewis, K. Uğurbil, A.P. Georgopoulos, Cerebral cortical mechanisms of copying geometrical shapes: a multidimensional scaling analysis of fMRI patterns of activation, Exp. Brain Res. 4 (3) (2009) 369–380.

[25] J.A. Tenreiro Machado, A.C. Costa, M.F. Lima, Dynamical analysis of compositions, Nonlin. Dynam. 65 (4) (2011) 399–412.

[26] J.A. Tenreiro Machado, A.M. Lopes, The persistence of memory, Nonlin. Dynam. 79 (1) (2014) 63–82.

[27] M.M. Bronstein, A.M. Bronstein, R. Kimmel, I. Yavneh, Multigrid multidimensional scaling, Numer. Linear Algebra. 13 (2–3) (2006) 149–171.

[28] J.B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika 29 (1) (1964) 1–27.

[29] S.S. Schiffman, M. Lance Reynolds, F.W. Young, J. Douglas Carroll, Introduction to Multidimensional Scaling: Theory, Methods, and Applications, Academic Press, New York, 1981.

[30] F.W. Young, Multidimensional Scaling: History, Theory, and Applications, Psychology Press, 2013.

[31] D.B. Stephenson, F.J. Dolas-Reyes, Statistical methods for interpreting Monte Carlo ensemble forecasts, Tellus A 52 (3) (2000).

[32] A.D.M.E. Osterhaus, J.C. De Jong, G.F. Rimmelzwaan, E.C.J. Claas, H5N1 influenza in Hong Kong: virus characterizations, Vaccine 20 (2002) S82–S83.

[33] P.S. Pandit, D.A. Bunn, S.A. Pande, S.S. Aly, Modeling highly pathogenic avian influenza transmission in wild birds and poultry in West Bengal, India, Sci. Rep. 3 (2013).

[34] J.J. Treanor, Influenza (Including Avian Influenza and Swine Influenza), in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1815–1839.

[35] J.P. Wood, Y.W. Choi, D.J. Chappie, J.V. Rogers, J.Z. Kaye, Environmental persistence of a highly pathogenic avian influenza (H5N1) virus, Environ. Sci. Technol. 44 (19) (2010) 7515–7520.

[36] I.G. Helmuth, A. Poulsen, C.H. Suppli, K. Mølbak, Varicella in Europe—a review of the epidemiology and experience with vaccination, Vaccine 33 (21) (2015) 2406–2413.

[37] A.H. Ross, E. Lenchner, G. Reitman, Modification of chicken pox in family contacts by administration of gamma globulin, NEJM 267 (8) (1962) 369–376.

[38] M. Weeks Wells, W.A. Holla, Ventilation in the flow of measles and chickenpox through a community: progress report, Jan. 1, 1946 to June 15, 1949, airborne infection study, Westchester County Department of Health, J. Am. Med. Assoc. 142 (17) (1950) 1337–1344.

[39] R.J. Whitley, Chickenpox and herpes zoster (varicella-zoster virus), in: Mandell, Douglas, and Bennett's principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1731–1737.

[40] R.N. Charrel, X. de Lamballerie, D. Raoult, Chikungunya outbreaks—the globalization of vectorborne diseases, NEJM 356 (8) (2007) 769.

[41] L. Markoff, Alphavirus, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1865–1874.

[42] S.C. Weaver, N.L. Forrester, Chikungunya: evolutionary history and recent epidemic spread, Antiviral Res. (2015).

[43] T.P. Endy, Dengue human infection model performance parameters, J. Infect. Dis. 209 (Suppl. 2) (2014) S56–S60.

[44] A. Khan, M. Hassan, M. Imran, Estimating the basic reproduction number for single-strain dengue fever epidemics, Infect. Dis. Poverty 3 (1) (2014) 12.

[45] C.L. Althaus. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. arXiv preprint arXiv:1408.3505, 2014.

[46] C.L. Althaus, Ebola: the real lessons from HIV scale-up, Lancet Infect. Dis. 15 (5) (2015) 507–508.

[47] T.W. Geisbert, Marburg and Ebola Hemorrhagic Fevers (Filovirus), in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1995–1999.

[48] H.-F. Hung, Y.-C. Wang, A.M.-F. Yen, H.-H. Chen, Stochastic model for hepatitis B virus infection through maternal (vertical) and environmental (horizontal) transmission with applications to basic reproductive number estimation and economic appraisal of preventive strategies, Stoch. Environ. Res. Risk Assess. 28 (3) (2014) 611–625.

[49] J. Mann, M. Roberts, Modelling the epidemiology of hepatitis B in New Zealand, J. Theor. Biol. 269 (1) (2011) 266–272.

[50] C.L. Thio, C. Hawkins, Hepatitis B virus and hepatitis delta virus, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1815–1839.

[51] A.L. de Goede, A.G. Vulto, A.D.M.E. Osterhaus, R.A. Gruters, Understanding HIV infection for the design of a therapeutic vaccine. Part I: epidemiology and pathogenesis of HIV infection, in: Annales Pharmaceutiques Francaises, Elsevier, 2014.

[52] J. Fettig, M. Swaminathan, C.S. Murrill, J.E. Kaplan, Global epidemiology of HIV, Infect. Dis. Clin. North Am. 28 (3) (2014) 323–337.

[53] S. Reid, O.A. Juma, Minimum infective dose of HIV for parenteral dosimetry, Int. J. STD AIDS 20 (12) (2009) 828–833.

[54] M.S. Reitz, R.C. Gallo, Human immunodeficiency virus, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 2054–2065.

[55] V. Rougeron, H. Feldmann, G. Grard, S. Becker, E. Leroy, Ebola and Marburg haemorrhagic fever, J. Clin. Virol. 64 (2015) 111–119.

[56] D.N. Durrheim, N.S. Crowcroft, P.M. Strebel, Measles—the epidemiology of elimination, Vaccine 32 (51) (2014) 6880–6883.

[57] A.A. Gershon, Measles virus (Rubeola), in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1967–1973.

[58] A.A. Gershon, Rubella virus (German measles), in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1875–1880.

[59] P.A. Rota, S.L. Liffick, J.S. Rota, R.S. Katz, S. Redd, M. Papania, et al., Molecular epidemiology of measles viruses in the United States, 1997–2001, Emerg. Infect. Dis. 8 (9) (2002) 902–908.

[60] T.L. Schlenker, C. Bain, A.L. Baughman, S.C. Hadler, Measles herd immunity: the association of attack rates with immunization rates in preschool children, JAMA 267 (6) (1992) 823–826.

[61] K. McIntosh, S. Perlman, Coronaviruses, including Severe Acute Respiratory Syndrome (Sars) and Middle East Respiratory Syndrome (MERS), in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1731–1737.

[62] Z.A. Memish, A. Assiri, R. Alhakeem, S. Yezli, M. Almasri, A. Zumla, et al., Middle East respiratory syndrome corona virus, MERS-CoV. Conclusions from the 2nd Scientific Advisory Board Meeting of the WHO Collaborating Center for Mass Gathering Medicine, Riyadh, Int. J. Infect. Dis. 24 (2014) 51–53.

[63] N. van Doremalen, T. Bushmaker, W.B. Karesh, V.J. Munster, Stability of Middle East respiratory syndrome coronavirus in milk, Emerg. Infect. Dis. 20 (7) (2014) 1263–1264.

[64] N. Litman, S.G. Baum, Mumps virus, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1942–1947.

[65] S. Rubin, M. Eckhaus, L.J. Rennick, C.G. Bamford, W. Paul Duprex, Molecular biology, pathogenesis and pathology of mumps virus, J. Pathol. 235 (2) (2015) 242–252.

[66] S.M. Ahmed, A.J. Hall, A.E. Robinson, L. Verhoef, P. Premkumar, U.D. Parashar, et al., Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis, Lancet Infect. Dis. 14 (8) (2014) 725–730.

[67] R. Dolin, J.J. Treanor, Noroviruses and Sapoviruses (Caliciviruses), in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 2122–2127.

[68] E.C. Pirtle, G.W. Beran, Virus survival in the environment, Rev. Sci. Tech. 10 (1991) 733–748.

[69] J.R. Romero, J.F. Modlin, Poliovirus, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 2073–2079.

[70] L. Weinstein, A. Shelokov, R. Seltser, G.D. Winchell, A comparison of the clinical features of poliomyelitis in adults and in children, N. Engl. J. Med. 246 (8) (1952) 297–302.

[71] K. Hampson, J. Dushoff, S. Cleaveland, D.T. Haydon, M. Kaare, C. Packer, et al., Transmission dynamics and prospects for the elimination of canine rabies, PLoS Biol. 7 (3) (2009) e1000053.

[72] O. Matouch, J. Jaros, P. Pohl, [Survival of rabies virus under external conditions], Vet. Med. (Praha) 32 (11) (1987) 669–674.

[73] D.L. Noah, C.L. Drenzek, J.S. Smith, J.W. Krebs, L. Orciari, J. Shaddock, et al., Epidemiology of human rabies in the United States, 1980 to 1996, Ann. Intern. Med. 128 (11) (1998) 922–930.

[74] K. Singh, C.E. Rupprecht, T. Bleck, Rabies (Rhabdoviruses), in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1984–1994.

[75] E.C. Dick, L.C. Jennings, K.A. Mink, C.D. Wartgow, S. LInborn, Aerosol transmission of rhinovirus colds, J. Infect. Dis. 156 (3) (1987) 442–448.

[76] J.O. Hendley, J.M. Gwaltney, Mechanisms of transmission of rhinovirus infections, Epidemiol. Rev. 10 (1) (1988) 242–258.

[77] R.B. Turner, Rhinovirus, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 2113–2121.

[78] P.R. Dormitzer, Rotaviruses, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 1854–1864.

[79] M.M. Patel, V. Pitzer, W.J. Alonso, D. Vera, B. Lopman, J. Tate, et al., Global seasonality of rotavirus disease, Pediatr. Infect. Dis. J. 32 (4) (2013) e134.

[80] V.E. Pitzer, K.E. Atkins, B.F. de Blasio, T. Van Effelterre, C.J. Atchison, J.P. Harris, et al., Direct and indirect effects of rotavirus vaccination: comparing predictions from transmission dynamic models, PLoS ONE 7 (8) (2012) e42320.

[81] J.E. Tate, A.H. Burton, C. Boschi-Pinto, A. Duncan Steele, J. Duque, U.D. Parashar, 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis, Lancet Infect. Dis. 12 (2) (2012) 136–141.

[82] S.E. Reef, S.B. Redd, E. Abernathy, L. Zimmerman, J. PIcenogle, The epidemiological profile of rubella and congenital rubella syndrome in the United States, 1998–2004: the evidence for absence of endemic transmission, Clin. Infect. Dis. 43 (Suppl. 3) (2006) S126–S132.

[83] L.L.M. Poon, Y. Guan, J.M. Nicholls, K.Y. Yuen, J.S.M. Peiris, The aetiology, origins, and diagnosis of severe acute respiratory syndrome, Lancet Infect. Dis. 4 (11) (2004) 663–671.

[84] W. Shi, F. Lei, C. Zhu, F. Sievers, D.G. Higgins, A complete analysis of HA and NA genes of influenza A viruses, PLoS ONE 5 (12) (2010) e14454.

[85] R.A. Weinstein, C.B. Bridges, M.J. Kuehnert, C.B. Hall, Transmission of influenza: implications for control in health care settings, Clin. Infect. Dis. 37 (8) (2003) 1094–1101.

[86] D.K. Milton, What was the primary mode of smallpox transmission? Implications for biodefense, Front. Cell. Infect. Microbiol. 2 (2012).

[87] C. Theves, P. Biagini, E. Crubezy, The rediscovery of smallpox, Clin. Microbiol. Infect. 20 (3) (2014) 210–218.

[88] B. Ahrens, Distance in spatial interpolation of daily rain gauge data, Hydrol. Earth Syst. Sci. 10 (2) (2006) 197–208.

[89] F.B. Duarte, J.A. Tenreiro Machado, G. Monteiro Duarte, Dynamics of the Dow Jones and the NASDAQ stock indexes, Nonlin. Dynam. 61 (4) (2010) 691–705.

[90] A.M. Lopes, J.A. Tenreiro Machado, C.M.A. Pinto, A.M.S.F. Galhano, Multidimensional scaling visualization of earthquake phenomena, J. Seismol. 18 (1) (2014) 163–179.

[91] A.K. Jain, Data clustering: 50 years beyond k-means, Pattern Recognit. Lett. 31 (8) (2010) 651–666.

[92] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Commun. Stat. Theory Methods 3 (1) (1974) 1–27.

[93] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. 2 (1979) 224–227.

[94] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65.

[95] A.M. Lopes, J.A. Tenreiro Machado, Dynamic analysis of earthquake phenomena by means of pseudo phase plane, Nonlin. Dynam. 74 (4) (2013) 1191–1202.

[96] F. Murtagh, A survey of recent advances in hierarchical clustering algorithms, Comput. J. 26 (4) (1983) 354–359.

[97] J.C. Gower, Generalized Procrustes analysis, Psychometrika 40 (1) (1975) 33–51.

[98] L.L. Borio, D.A. Henderson, N.A. Hynes, Bioterrorism: an overview, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 178–190.

[99] F.J. Burt, M.S. Rolph, N.E. Rulli, S. Mahalingam, M.T. Heise, Chikungunya: a re-emerging virus, Lancet 379 (9816) (2012) 662–671.

[100] R. Khabbaz, B.P. Bell, A. Schuchat, S.M. Ostroff, R. Moseley, A. Levitt, et al., Emerging and reemerging infectious disease threats, in: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. eighth ed., Saunders, 2014, pp. 158–177.