SIS-FENStatS 2026

valgrAI

Archetypal analysis of European 10-year government bond yields with multidimensional scaling of two-mode three-way asymmetric dissimilarities

Aleix Alcacer, Rafael Benítez, Vicente J. Bolós, Irene Epifanio

UNIVERSITAT JAUME I

VNIVERSITAT DE VALÈNCIA

1 BACKGROUND: Multidimensional scaling for 1-mode 2-way asymmetric data by h-plotting

Let $\mathbf{\Delta}$ be an $n \times n$ dissimilarity matrix with elements δ_{ij} representing the dissimilarity from object "i" to object "j". When $\mathbf{\Delta}$ is asymmetric, we define $\mathbf{D} := (\mathbf{\Delta}|\mathbf{\Delta}')$, by concatenating $\mathbf{\Delta}$ and its transpose $\mathbf{\Delta}'$ column-wise, where $'$ denotes transposition.

$\mathbf{D}$ is treated as a data matrix. The first n columns correspond to variables of the form $d_{\cdot j}$, that is, the dissimilarities from all other objects to object "j". The second block of n columns represents variables of the form $d_{j\cdot}$, that is, the dissimilarities from object "j" to all other objects.

H-plot is applied to $\mathbf{D}$, i.e., one performs an eigendecomposition of the variance–covariance matrix $\mathbf{S}$ of $\mathbf{D}$: $\mathbf{H}_2 := (\sqrt{\lambda_1}\mathbf{q}_1, \sqrt{\lambda_2}\mathbf{q}_2)$, where λ_1, λ_2 are the two largest eigenvalues with corresponding eigenvectors $\mathbf{q}_1, \mathbf{q}_2$. The Euclidean distance between two rows, $\mathbf{h}_i$ and $\mathbf{h}_j$, is approximately equal to the sample standard deviation of the difference between variables "i" and "j" (see [5]).

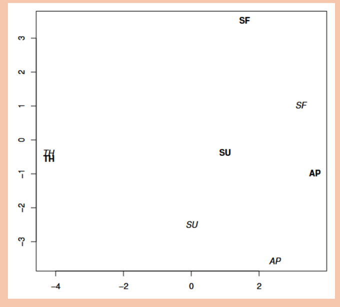
Advantages of h-plots for mutidimensional scaling:

1. Intuitive interpretability: Profiles are directly comparable since all profiles are displayed within a single, unified Euclidean representation.
2. Explicit solution.
3. Scale invariance.
4. Computational efficiency (scalability).
5. Simple goodness-of-fit evaluation: $\dfrac{\lambda_1^2 + \lambda_2^2}{\sum_j \lambda_j^2}$
6. Robustness.

TOY EXAMPLE

Citing/Cited	AP	SF	SU	TH
AP	1	10	5	10
SF	4	2	5	10
SU	4	9	3	6
TH	10	10	7	1

Dissimilarities between journals.

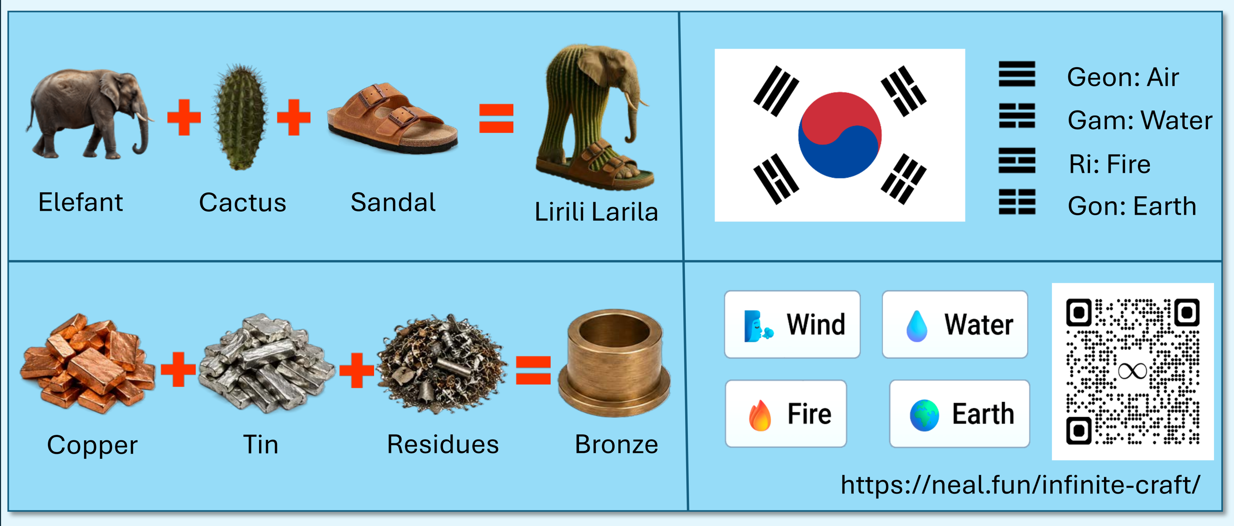


H-plot projection of one-mode two-way journal citation data. Bold font represents the dissimilarity variables of cited ($d_{\cdot j}$), while italic font denotes citing ($d_{j\cdot}$).

2 BACKGROUND: Archetypoid analysis

Archetypoid Analysis (ADA) is and unsupervised statistical learning method that represents data as mixtures of "extreme" or pure types (archetypoids), which are actual observations (see [3,8]).

Archetype (Wikipedia): from Greek ἀρχή *arkhḗ*, "beginning", "origin", and τύπος *týpos*, "pattern", "model", or "type"; **original pattern from which copies are made** (see [3,4]).

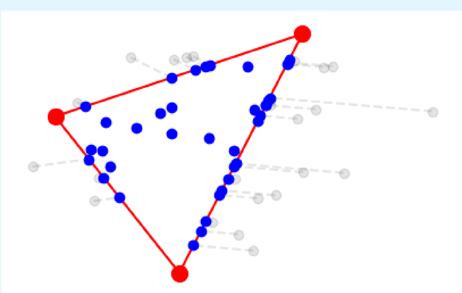


Given data $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$, select k archetypoids $\mathbf{z}_1, \ldots, \mathbf{z}_k$ from the observed data. Then, each observation is approximated by a convex combination of these archetypoids:

$$\mathbf{x}_i \approx \hat{\mathbf{x}}_i = \sum_{g=1}^{k} \alpha_{ig}\mathbf{z}_g, \qquad \sum_{g=1}^{k} \alpha_{ig} = 1, \quad \alpha_{ig} \geq 0, \qquad i = 1, \ldots, n.$$

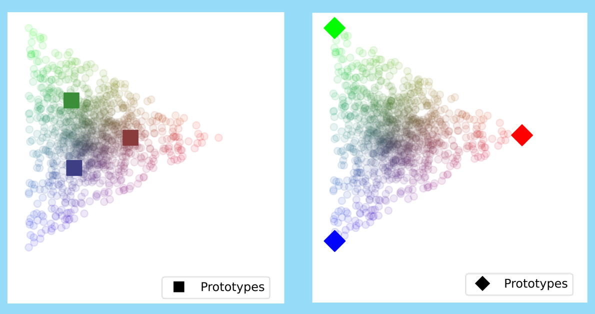
Under these constraints, the coefficients α_{ig} are chosen to minimize the residual sum of squares:

$$RSS = \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{g=1}^{k} \alpha_{ig}\mathbf{z}_g \right\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_{ij} - \sum_{g=1}^{k} \alpha_{ig} z_{gj} \right)^2.$$



Data approximated as convex combinations of the archetypoids (red points).

Clustering versus ADA



Archetypal patterns are more readily interpretable than central or average points, as people tend to understand data more intuitively through contrasting or opposing examples.

3 PROPOSAL: Archetypal profiles for 2-mode 3-way asymmetric data

NEW

Assume that $\mathbf{\Delta}_l$ $(l = 1, \ldots, L)$ denotes a square dissimilarity matrix, where each element δ_{ijl} corresponds to the pairwise dissimilarity from object i to object j $(i, j = 1, \ldots, n)$ observed at occasion or condition l. If $\delta_{ijl} = \delta_{jil}$, the dissimilarity matrix is considered symmetric; otherwise, it is asymmetric.

For the asymmetric case, we define a $n \times 2Ln$ matrix concatenating the matrices $\mathbf{\Delta}_l$ and their transposes $\mathbf{\Delta}_l'$ column-wise:

$$\mathbf{D} := (\mathbf{\Delta}_1|\mathbf{\Delta}_1'| \ldots |\mathbf{\Delta}_L|\mathbf{\Delta}_L').$$

Then, $\mathbf{D}$ is h-plotted in two dimensions, obtaining the matrix $\mathbf{H}_2$ with $2Ln$ rows and 2 columns. We reverse the concatenation and partition the matrix $\mathbf{H}_2$ into individual blocks, each consisting of n rows. Each of the $2L$ blocks is visualized in a two-dimensional plot, with L distinct colors indicating the different conditions, and with two different font styles: **bold** and *italic* for $d_{\cdot jl}$ and $d_{j\cdot l}$ profiles, respectively. This new matrix, denoted as $\mathbf{Y}$, consists of n rows and $4L$ columns.

Unsupervised statistical learning techniques for two-mode three-way asymmetric data:

Once the h-plot projections in the Euclidean space are represented in the matrix $\mathbf{Y}$, different statistical methodologies can be applied:

- In **clustering**, k-medoids can be applied, which is a promising alternative to the scarce techniques designed for asymmetric two-mode three-way data, according to the comparisons in [1] .
- **ADA can be applied**, allowing, **for the first time**, the extraction of **archetypal profiles** from this class of data.

TOY EXAMPLE 2

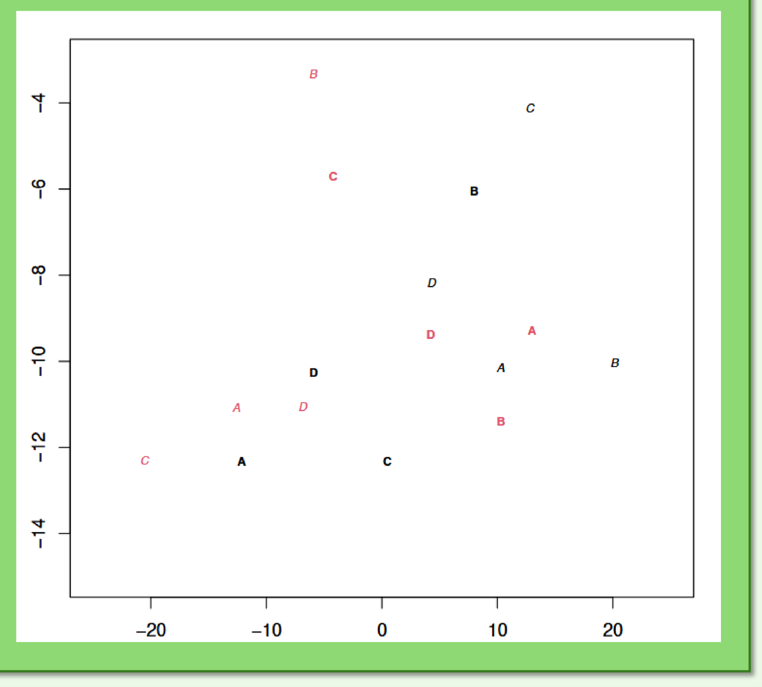
	Occasion 1				Occasion 2			
	A	B	C	D	A	B	C	D
A	50	25	50	25	50	50	19	23
B	50	0	50	25	18	25	10	10
C	20	20	50	10	50	50	0	20
D	20	10	20	0	27	22	5	0

Artificial three-way proximity data.

	Received 1				Sending 1				Received 2				Sending 2			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
A	0	25	0	25	0	0	30	30	0	0	31	27	0	32	0	23
B	0	50	0	25	25	50	30	40	32	25	40	40	0	25	0	28
C	30	30	0	40	0	0	0	30	0	0	50	30	31	40	50	45
D	30	40	30	50	25	25	40	50	23	28	45	50	27	40	30	50

Matrix $\mathbf{D}$ for the unconditional two-mode three-way artificial example. "Received" means $d_{\cdot j}$ profiles, while "Sending" denotes $d_{j\cdot}$ profiles, and numbers indicate the occasions 1 and 2.

H-plot projection of unconditional two-mode three-way artical example. The first condition is represented in **black**, and the second in **red**. There are two different font styles: **bold** and *italic* for $d_{\cdot jl}$ and $d_{j\cdot l}$ profiles, respectively.



4 APPLICATION

We use a data set corresponding to monthly time series of **10-year government bond yields** of 23 European countries obtained from Eurostat: **AT** (Austria), **BE** (Belgium), **CY** (Cyprus), **CZ** (Czech Republic), **DE** (Germany), **DK** (Denmark), **EL** (Greece), **ES** (Spain), **FI** (Finland), **FR** (France), **HU** (Hungary), **IE** (Ireland), **IT** (Italy), **LT** (Lithuania), **LU** (Luxembourg), **LV** (Latvia), **MT** (Malta), **NL** (Netherlands), **PL** (Poland), **PT** (Portugal), **SE** (Sweden), **SK** (Slovakia), **UK** (United Kingdom).

The time interval runs from **January 2001** to **April 2025** because many countries do not have data prior to 2001 and, on the other hand, the United Kingdom does not have data available after April 2025 at the time of this analysis. Other countries such as Bulgaria, Estonia and Romania have not been considered because some data for this period is missing.

The time interval is divided into three subintervals, taking **April 2011** (when the effects of the European debt crisis became apparent, see [2]) and **March 2020** (the start of the COVID-19 crisis) as reference points.

Asymmetric dissimilarities (bounded between 0 and 1) are computed for each time subinterval using the **oriented wavelet squared coherence** proposed by [6]. This measure builds on wavelet squared coherence and incorporates the wavelet phase difference (see [7]). The resulting dissimilarities are defined in a causal framework: when "i" influences "j", the dissimilarity from "i" to "j" is lower than that from "j" to "i".

Code and data: https://epifanio.uji.es/RESEARCH/asyhploteuro.zip

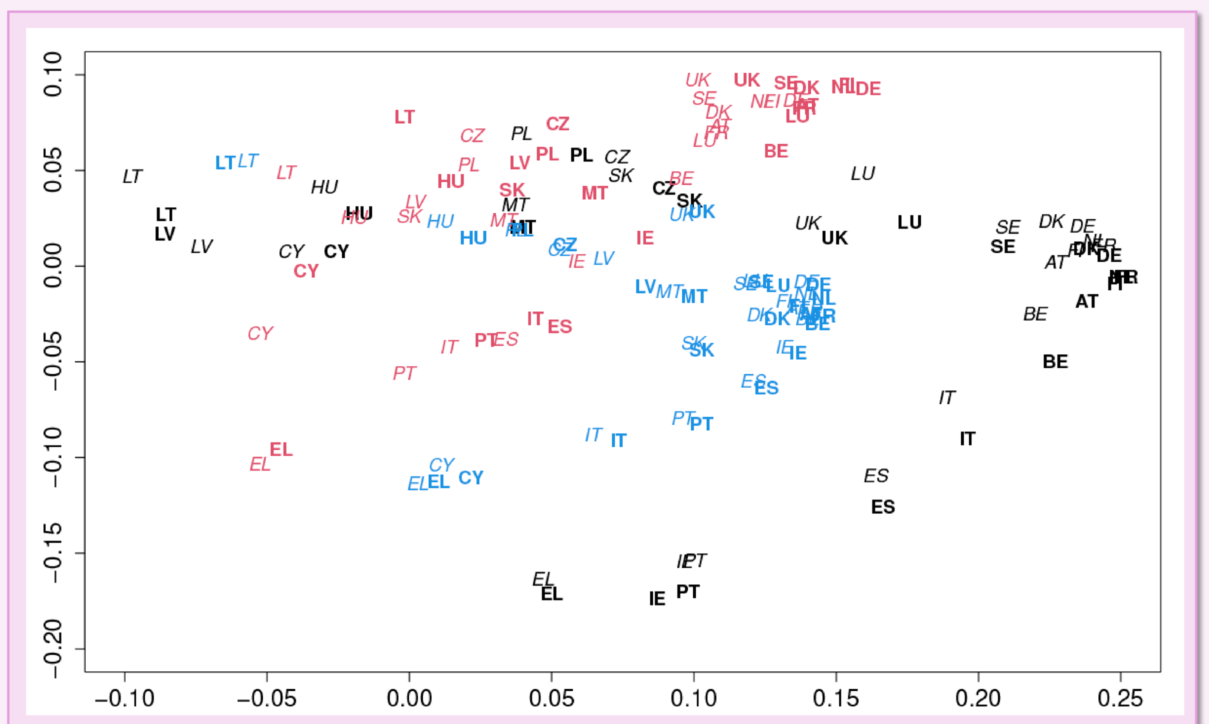
5 RESULTS

The structure formed by the countries in each time subinterval is preserved in the h-plot, exhibiting only some shifts.

EL (**archetypoid 1**), together with other Southern European countries such as **PT**, **ES**, and **IT**, occupies a branch located at the lower part of the structure.

Countries with a predominance of **archetypoid 3** (**NL**, **AT**, **BE**, **DE**, **DK**, **FI**, **FR**, **SE**), form a central group.

Finally, Baltic countries **LT** and **LV**, and Central European countries, such as **HU**, **PL** and the **CZ**, form another branch at the top characterized by a major weight of **archetypoid 2**.



H-plot projection of European countries. **Bold** font represents influenced $d_{\cdot j}$ profiles, while *italic* font denotes influencing $d_{j\cdot}$ profiles. The first time subinterval (from January 2001 to March 2011) is represented in **black**, the second (from April 2011 to February 2020) in **red**, and the third (from March 2020 to April 2025) in **blue**.

Some temporal movements can nonetheless be observed. **IE** is positioned within the **EL** branch during the **first period**, but transitions to the central group in the subsequent two periods. **ES** also moves closer to this central group during these later periods. Conversely, **CY**, which is located in the upper branch in the **first period**, shifts to a position adjacent to **EL** by the **third period**.

With respect to asymmetry (see [1] for how to assess it from h-plot), the **second period** exhibits the highest degree of asymmetry, suggesting a greater prevalence of cause–effect relationships during this interval. By contrast, the **third period** is the least asymmetric. At the country level, **EL** displays the lowest degree of asymmetry, which may be interpreted as a more independent evolution of its time series. In contrast, the Baltic countries **LT** and **LV**, together with **LU**, show the highest asymmetry, indicating stronger cause–effect relationships in the evolution of their time series.

	AT	BE	CY	CZ	DE	DK	EL	ES	FI	FR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	SE	SK	UK
1	6	15	36	0	0	1	100	51	2	4	0	54	51	0	0	1	4	0	0	71	0	9	0
2	0	0	64	64	1	7	0	0	1	0	100	0	0	100	25	99	65	0	75	0	14	52	34
3	94	85	0	36	99	92	0	49	97	96	0	46	49	0	75	0	31	100	25	29	86	39	66

3 archetypoids are calculated according to the screeplot: **EL**, **HU**, and **NL**. The table shows the parameters α_{ig} multiplied by 100. This represents the percentage of each archetypoid (**1**, **2**, or **3**) in each country approximation from 2001 to 2025.

Similar results to next poster, see [2].

6 CONCLUSIONS

- We extended the h-plot methodology for multidimensional scaling to the analysis of three-way proximity data, encompassing both symmetric and asymmetric as well as conditional and unconditional frameworks.
- The proposed methodology facilitates the identification of archetypal profiles and clustering structures for three-way asymmetric proximities, enhancing interpretability and analytical flexibility.
- We apply the proposed methodology to 10-year government bond yields of European countries over the period 2001–2025. The results identify three country archetypoids: Greece, Hungary, and the Netherlands. All other countries can be represented as mixtures of these archetypoids.

REFERENCES

[1] Alcacer, A., Benítez, R., Bolós, V.J., Epifanio, I.: Multidimensional scaling of two-mode three-way asymmetric dissimilarities: finding archetypal profiles and clustering. arXiv:2511.15813 (2025).
[2] Alcacer, A., Benítez, R., Bolós, V.J., Epifanio, I.: Biarchetype analysis for univariate functional data. An application to macroeconomic financial time series. In: Proceedings of the SIS-FENStatS 2026 Conference. Springer (2026).
[3] Alcacer, A., Epifanio, I., Mair, S., Morup, M.: A survey on archetypal analysis. arXiv:2504.12392 (2025).
[4] Cutler, A., Breiman, L.: Archetypal Analysis. Technometrics 36(4), 338–347 (1994).
[5] Epifanio, I.: H-plots for displaying nonmetric dissimilarity matrices. Statistical Analysis and Data Mining 6(2), 136–143 (2013).
[6] Ferrer, R., Benítez, R., Bolós, V.J.: Interdependence between green financial instruments and major conventional assets: a wavelet-based network analysis. Mathematics 9(8), 900–919 (2021).
[7] Torrence, C., Webster, P.J.: Interdecadal Changes in the ENSO–Monsoon System. Journal of Climate 12(8), 2679–2690 (1999).
[8] Vinué, G., Epifanio, I., Alemany, S.: Archetypoids: A new approach to define representative archetypal data. Computational Statistics & Data Analysis 87, 102–115 (2015).

ACKNOWLEDGEMENT

This work was partially supported by the Spanish Ministry of Science and Innovation PID2023-153128NB-I00, PID2022-141699NB-I00, PID2020-118763GA-I00, and Generalitat Valenciana CIPROM/2023/66.