AN APPLICATION OF ASYMMETRIC MULTIDIMENSIONAL SCALING TO THE VQR 2015-2019 DATA

Giuseppe Bove[1]

[1] Dipartimento di Scienze della Formazione, Università degli Studi Roma Tre,
(e-mail: giuseppe.bove@uniroma3.it)

ABSTRACT: Proximity matrices are frequently asymmetric and analysed by using the additive decomposition into symmetric and skew-symmetric components. A preliminary graphical description of the two components can allows to detect interesting relationships in the data. An application to the matrix of flows of scientific products between GEV's in the VQR 2015-2019 is presented to emphasize the advantages of the graphical approach.

KEYWORDS: proximity data, asymmetric multidimensional scaling, research assessment.

1 Introduction

Asymmetric proximities between pairs of entities (e.g., import-export data, sociomatrices, brand switching, flows and migration data, etc.) are analysed in economics, sociology, marketing research and other behavioural sciences by using a variety of models and methods to detect meaningful relationships. When no substantive model is readily available and there is not a priori reason for preferring any particular model, we might look for graphical displays as preliminary investigation of the proximity data.

In this presentation, the approach analysing separately by diagrams the symmetric and the skew-symmetric components of the proximity matrix is applied to the matrix of flows of scientific products between groups of experts for evaluation (GEV), in the process of research quality assessment of the Italian research institutions regarding the years 2015-2019 conducted by the agency ANVUR (Italian Ministry of University and Research - MUR).

In the next sections, first the asymmetric multidimensional scaling (MDS) models for symmetry and skew-symmetry are presented, focalizing on the diagrams that are obtained and their interpretation. Follows the description of the proximity data published by ANVUR and the presentation of the diagrams obtained by the application of asymmetric MDS.

2 Models for symmetry and skew-symmetry

A ($n \times n$) proximity matrix $\boldsymbol{\Omega} = (\omega_{ij})$ between pairs of entities (i,j) ($i,j=1,2, \ldots, n$) can always be decomposed additively as $\boldsymbol{\Omega} = \mathbf{M} + \mathbf{N}$, with $\mathbf{M} = (m_{ij})$ a symmetric component and $\mathbf{N} = (n_{ij})$ a skew-symmetric component, defined respectively, $m_{ij} = (\omega_{ij} + \omega_{ji})/2$ and $n_{ij} = (\omega_{ij} - \omega_{ji})/2$. The elements n_{ij} are the deviations of the proximities from symmetry, because $n_{ij} = \omega_{ij} - m_{ij}$. Constantine and Gower (1978) remark that the symmetric component **M** and the skew-symmetric component **N** of any square matrix **Ω** are uncorrelated and the following orthogonal breakdown of the sum of squares (**SS**) holds, **SS**(**Ω**) = **SS**(**M**) + **SS**(**N**), reflecting the uniqueness of the additive decomposition (for a review, see Bove, Okada and Vicari (2021)). Thus, the two components can be viewed independently, and studied by separate models. An advantage of the separate analysis is that one can deal with the representation problem by adopting different kinds of models (e.g., spatial vs nonspatial) and different kinds of transformations (e.g., metric vs nonmetric) for the two components.

The symmetric component **M** can be represented by a Euclidean distance model in r dimensions

$$f(m_{ij}) = \hat{d}_{ij} = d_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j) + e_{ij} = \sqrt{\sum_{s=1}^{r} (x_{is} - x_{js})^2} + e_{ij}, \quad (1)$$

where f is a monotone transformation, mapping the proximities ω_{ij} into a set of transformed values $\hat{d}_{ij} = f(m_{ij})$, x_{is} and x_{js} are the coordinates of row (column) i and row (column) j on dimension s, respectively, and the symmetric property holds ($d_{ij} = d_{ji}$). The estimation of the model provides a map where entries m_{ij} are approximated by distances in r dimensions: when the entries m_{ij} are similarities, the larger the values, the smaller the distances. The model can be estimated using standard statistical software for symmetric multidimensional scaling.

The skew-symmetric matrix **N** can be represented in a diagram by the two-step method proposed in Bove and Vicari (2023). At the first step, the sizes of the skew-symmetries $|n_{ij}|$ are collected in a symmetric matrix $\mathbf{T} = (t_{ij}) = (|n_{ij}|)$, a standard dissimilarity matrix that can be approximated in a *low-dimensional* (usually two-dimensional) Euclidean space by the distance model (1) where m_{ij} is replaced with t_{ij}. At the second step, the signs of the skew-symmetries $sign(n_{ij})$ can be represented in the configuration obtained at the first step, by the drift vector method proposed in Borg and Groenen (2005, par. 23.5) applied to matrix $\mathbf{K}^{sign} = (k_{ij}^{sign}) = (sign(n_{ij}))$. From each point i an arrow is drawn to each other point j so that its length and direction correspond to the values in row i of the skew-symmetric matrix $\mathbf{K}^{sign}$. For each pair (i,j), if n_{ij} is positive, the arrow goes from point i to point j, while, when n_{ij} is negative, the arrow points in the opposite direction. When the number of rows of the proximity matrix is large, the representation of all arrows may result into cluttered pictures, and it can be convenient to draw only the average vector (drift vector) of the arrow bundle attached to each point. The length of the drift vector represents the homogeneity of the directions of the (n-1) vectors emitting from the point. The arrows representing the drift vectors go towards areas of the diagram containing points having more frequently negative skew-symmetries and they can exhibit systematic asymmetric trends.

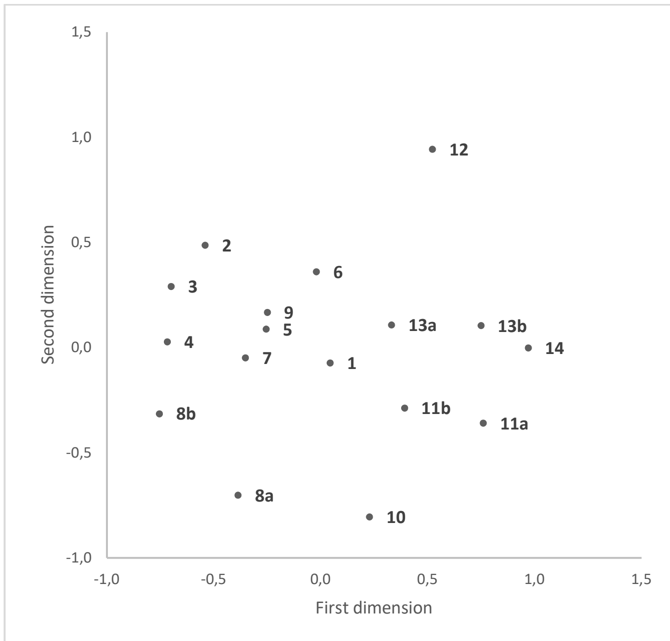

Fig. 1 - MDS of the symmetric component of the transformed 2015-2019 VQR data

3 Application to 2015-2019 VQR data

The methods for symmetry and skew-symmetry presented in section 2 are now applied to a proximity matrix published on-line by ANVUR in the VQR 2015-2019 Final Report - Statistics and summary results (Table 2.11). Entries of the matrix are the flows of scientific products between the seventeen GEV's (their labels and names are listed in Table 2.1 of the Final Report) and can be considered as measures of similarity between GEV's. Diagonal entries are strongly dominant (97.3% of the total flows) and represent the scientific products evaluated inside the GEV's. Nonetheless, off-diagonal entries seem interesting to detect affinity relationships between GEV's. A preliminary transformation was applied to remove the main effects, which reflects the influence of the row and column totals. A simple correction for such main effects is to equalize all self-similarities by dividing the entry in each cell (i,j) by the square root of the product of the entries in diagonal cells (*i*,*i*) and (*j*,*j*) (self-similarities). This transformation does not affect the asymmetry, the ratios between off-diagonal entries in cells (*i*,*j*) and (*j*,*i*) for ($i \neq j$) remain the same.

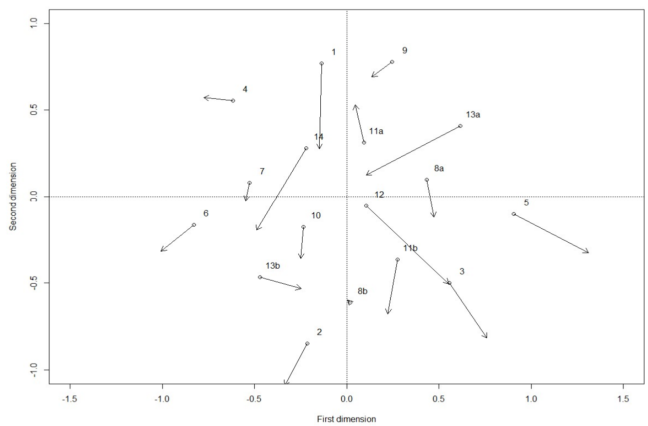

Figure 1 provides the representation of the GEV's average flows (symmetric component), small distances represent high average flows. GEV's positioned far from the origin (e.g., GEV 10 - Antiquity, Philologic.-Literary and Historical-Artistic sciences and GEV 12 – Legal sciences) have small or null flows with the other GEV's. GEV's close to the origin (e.g., GEV 1 - Mathematics and computer sciences and GEV 5 – Biological sciences) have high flows with several other GEV's. Figure 2 provides the representation of the skew-symmetric component, large distances indicate high imbalances (e.g., GEV 2 – Physical science and GEV 9 – Industrial and information engineering, in this case the direction of the arrows indicate that the skew-symmetry is positive from GEV 9 to GEV 2). Homogeneity of the directions characterize GEV's with long arrows (e.g., GEV 13a – Economics and statistics).
Fig. 2 – Drift vectors for skew-symmetry in the transformed 2015-2019 VQR data

References

BORG, I., GROENEN, P.J.F. 2005. *Modern Multidimensional Scaling. Theory and Applications*. (Second Edition). New York: Springer.

BOVE, G., OKADA, A. & VICARI, D. 2021. *Methods for the analysis of asymmetric proximity data*. Singapore: Springer Nature.

BOVE, G., VICARI, D. 2023. Graphical Analysis and Clustering of Asymmetric Proximities. In: Okada, A. et al. (eds.) *Facets of Behaviormetrics: The 50th Anniversary of the Behaviormetric Society*. Singapore: Springer Nature (to appear)

CONSTANTINE, A.G., GOWER, J.C. 1978. Graphical representation of asymmetric matrices. *Applied Statistics*, **3**, 297-304.