Editorial for Special Issue on Analysis of Asymmetric Relationships

Giuseppe Bove¹ · Akinori Okada²

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

This Special Issue of the journal *Advances in Data Analysis and Classification* collects seven papers that deal with the Analysis of Asymmetric Relationships. In many disciplines such as psychology, sociology, marketing research, behavioral sciences, and so on, direct or indirect measures of relationship between pairs of objects are frequently studied. Examples are proximity ratings, socio-matrices, brand switching or import-export data.

While most of these relationships are asymmetric in practice, asymmetry had been ignored for long time. But in 1960's some researchers began to pay attention to the asymmetry arising, for instance, in confusion experiments or in the frequency of citations among journals, or in switching among brands. So various procedures which can analyze and represent the asymmetry have been introduced for one-mode two-way data, and later extended to one-mode and two-mode three-way data. In the first article of this issue, entitled “Methods for the Analysis of Asymmetric Pairwise Relationships”, *Giuseppe Bove* and *Akinori Okada* provide a brief review of models and methods of multidimensional scaling and cluster analysis able to deal with asymmetric proximities, taking a ‘data-analytic’ approach and emphasizing data visualization.

A key role in the development of models and methods for multidimensional scaling and cluster analysis has been played by John Gower. In two pioneering papers

✉ Giuseppe Bove
bove@uniroma3.it

Akinori Okada
okada@rikkyo.ac.jp

¹ Department of Education, Università Roma Tre, Via Milazzo 11/b, 00185 Rome, Italy

² Research Institute, Tama University, Tokyo, Japan

published in the Seventies, Gower presented new tools for visual inspection of square asymmetric matrices based on least-squares approximation. In the second article of this issue, entitled “Skew Symmetry in Retrospect”, *John Gower* himself provides a short account of how he became interested in analysing asymmetry, when he spent the year 1974 in Adelaide with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), collaborating with Graham Constantine. The way they have achieved results in the study of rank approximation and singular value decomposition (SVD) of skew-symmetric matrices is described in the first part of the paper. Relevant in this context the discovery of a new geometrical interpretation of the algebra of the SVD in terms of areas of triangles. In the central part of the article properties of the rank-two linear form of skew-symmetry are described. It is shown how in certain situations substantive models could be diagnosed by opportunely linking linear skew-symmetry and symmetry, an aspect that has interested Gower also recently. Other research lines regarding the structure of various skew-symmetric matrices are mentioned in the final part of the article, and in the Appendix a simple proof of the singular value decomposition of skew-symmetric matrices is given.

The seminal work of Lawrence Hubert on Psychometrika in 1973 stimulated the development of cluster analysis methods for different types of data representing asymmetric relationships. A lack of proposals is registered with regard to skew-symmetric matrices, despite the interest in the study of skew-symmetry is long ago dated. In the third article, entitled “CLUSKEXT: CLUstering model for SKew-symmetric data including EXTernal information”, *Donatella Vicari* proposes a clustering model which relies on the decomposition of a skew-symmetric matrix into within and between cluster effects. The model is able to incorporate external information in order to explain imbalances between the objects, and a graphical representation can be displayed for the fitted imbalances within and between the clusters. Parameters are estimated by an alternating least-squares algorithm. The external variables are assumed as covariates in accounting for the total sum of squares, so the loss function can be partitioned into components that represent different aspects of the lack of fit due to clustering and regression. Accordingly, it is possible to evaluate the adequacy of the model in a confirmative perspective if it is compared with the version not including the external variables. Artificial data describe intuitively the proposed methodology at the beginning of the paper. Finally, applications to Ekman colour data and to cola brand switching data illustrate the performance and usefulness of the proposed model.

Relationships referred to asymmetrically weighted networks are considered by *Gunnar Carlsson, Facundo Mémoli, Alejandro Ribeiro* and *Santiago Segarra* in the fourth article of this issue, entitled “Hierarchical clustering of asymmetric networks”. Methods based on the directed dissimilarity structure of a network are proposed by considering the concept of admissibility. Any reasonable clustering method should satisfy some behaviors that can be imposed as axioms to characterize the space of methods that are admissible. The axioms of value and transformation are studied in detail and some their extensions and alternatives are considered. Accordingly, clustering methodologies are proposed and their properties in the space of the admissible methods are shown. The developed clustering methods are exemplified through their application to the network of internal migration between states of the U.S. for the year 2011.

Multilinear and distance models have been used complementary to (or incorporated in) statistical models to provide an easy interpretation of the association in contingency tables. This is possible because with these approaches plots of the rows and columns of the tables are obtained where scalar products or distances approximate association parameters and residuals. Examples are generalized correspondence analysis used complementary to log-linear models or RC(M) model reparametrization that provide expressions for the odds ratios in terms of distances. In the fifth article of this issue, entitled “Transitional modelling of experimental longitudinal data with missing values”, *Mark de Rooij* extends these ideas to the analysis of longitudinal multinomial data in a distance framework. A randomized context where a treatment is compared to a control group on several time points is assumed. A transitional approach is considered, so the main interest is whether the transition probabilities for the treatment group differ from those of the control group. A generalization of the ideal point classification (IPC) model previously proposed by de Rooij is adopted to detect the different transition profiles. The proposed distance approach enhances interpretation by graphical representation; moreover, sparse data are treated more efficiently compared to a multinomial regression model. A Monte Carlo simulation study is provided to investigate the capability to deal with missing values. Finally, the new approach is illustrated by a real data example concerning the McKinney Homeless Research Project in San Diego.

Asymmetric square data matrices with similarly classified rows and columns can be analysed by any method appropriate to a rectangular table, such as Biplot or Multidimensional Unfolding. Rows and columns are represented by separate points by these methods, so it is possible to depict directly all the entries in the data matrix by row-column and column-row scalar products or distances. Following this approach, the sixth article of this issue, entitled “Assessing the asymmetric effects on branch rivalry of Spanish financial sector restructuring”, by *Marti Sagarra*, *Frank M.T.A. Busing*, *Cecilio Mar-Molinero* and *Josep Rialp*, examines the changes in market structure and competition among Spanish banks caused by the banking crisis that began in 2008. The analysis focuses in the rivalry structure of the Spanish banking sector derived from Chen’s indexes of intercompany rivalry. These indexes were collected in two asymmetric rivalry matrices between 15 financial groups, respectively for years 2008 (before the crisis) and 2012 (the last available data set). The asymmetry in the competitive pressure between banks is analysed separately for the two years by two-way multidimensional unfolding. The two configurations obtained appeared very similar, and a Procrustes Analysis confirmed the strong similarity. Finally, the common map obtained for all the financial groups by three-way multidimensional unfolding allows to conclude that, despite the many changes that have taken place during the period studied, the basic structure of the banking sector in Spain remains unchanged.

Confusion experiments were mentioned at the beginning of this editorial as one of the first contexts where asymmetry has been analysed. In particular, in marketing research many methods were proposed to deal with brand confusion data. In the seventh article, entitled “Relating brand confusion to ad similarities and brand strengths through image data analysis and classification”, *Daniel Baier* and *Sarah Frost* propose a new data collection approach that replaces the consumer experiments, usually demanding and costly, by a much cheaper content-based dissimilarity analysis of advertisements for brands. Brand advertisements are collected from the internet and

used to learn to classify via a multinomial logit approach. Predictors are feature-wise distances between advertisements. Results allow to say that the new approach can be profitably integrated with the standard methodology based on confusion experiment.

The Editors gratefully acknowledge the assistance of the experts and colleagues in the process of reviewing the manuscripts that were submitted for this Special Issue.