



Projection Clustering Unfolding: a new algorithm for clustering individuals or items in a preference matrix

Mariangela Sciandra · Antonio D'Ambrosio · Antonella Plaia

Received: date / Accepted: date

Abstract In the framework of preference rankings, the interest can lie in clustering individuals or items in order to reduce the complexity of the preference space for an easier interpretation of collected data. The last years have seen a remarkable flowering of works about the use of decision tree for clustering preference vectors. As a matter of fact, decision trees are useful and intuitive, but they are very unstable: small perturbations bring big changes. This is the reason why it could be necessary to use more stable procedures in order to clustering ranking data. In this work, a Projection Clustering Unfolding (PCU) algorithm for preference data will be proposed in order to extract useful information in a low-dimensional subspace by starting from an high but mostly empty dimensional space. Comparison between unfolding configurations and PCU solutions will be carried out through Procrustes analysis.

Keywords Projection pursuit · Preference data · Clustering rankings

Riassunto *Nell'ambito delle classifiche di preferenza, uno degli obiettivi potrebbe consistere nel raggruppare gli individui o gli oggetti al fine di ridurre la complessità dello spazio di preferenza, per un'interpretazione più semplice dei dati raccolti. Gli ultimi anni hanno visto una notevole fioritura di articoli sull'uso degli alberi decisionali per il raggruppamento dei vettori*

M. Sciandra, A. Plaia

Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo

Viale delle Scienze, Edificio 13

Tel.: +39-09123895235

E-mail: mariangela.sciandra@unipa.it, antonella.plaia@unipa.it

A. D'Ambrosio

Via Cinthia, M.te S. Angelo, Napoli, Italy

E-mail: antdambr@unina.it

di preferenze. Tuttavia va detto che, se da un lato gli alberi decisionali sono utili e intuitivi, dall'altro sono molto instabili: piccole perturbazioni possono condurre a grandi cambiamenti. Questo è il motivo per cui potrebbe essere necessario utilizzare procedure più stabili al fine di raggruppare i dati di ranking. In questo lavoro, verrà proposto un algoritmo di *Projection Clustering Unfolding (PCU)* per i dati delle preferenze per estrarre informazioni utili in un sottospazio a bassa dimensione partendo da uno spazio dimensionale alto ma per lo più vuoto. Il confronto tra le configurazioni di proiezione e le soluzioni PCU sarà effettuato attraverso l'analisi procustiana.

Parole chiave *Tecniche di proiezione - Dati di preferenza - Tecniche di raggruppamento per graduatorie*

1 Introduction

Projection pursuit includes a lot of techniques for finding interesting projections of multivariate data in low dimensional spaces (Friedman and Tukey, 1974). One particular structure is that of clusters in the data. Projection Pursuit Clustering (PPC) is a synthesis of projection pursuit and nonhierarchical clustering methods that simultaneously attempts to cluster the data and to find a low-dimensional representation of this cluster structure. As introduced by Huber (1985), a Projection Pursuit (PP) algorithm consists of two components: an index function $I(\alpha)$ that measures the “usefulness” of projection and a search algorithm that varies the projection direction so as to find the optimal projections, given the index function $I(\alpha)$ and the data set X . In this work we propose an iterative strategy that combine suitable clustering methods for preference rankings with Multidimensional unfolding techniques. We call our proposal *Projection Clustering Unfolding*. All the methodology is illustrated and evaluated on one real and well known dataset.

2 Preference data

In every day life ranking and classification are basic cognitive skills that people use in order to graduate everything that they experience. Grouping and ordering a set of elements is considered easy and communicative, so often it happens to observe rankings of sport-teams, universities, countries and so on. A particular case of ranking data is represented by preference data, in which individuals show their preferences over a set of alternatives, items from now on. Since preference rankings can be considered as indicators of individual behaviours, when subject-specific characteristics are available, an important issue relies on the identification of profiles of respondents giving same/similar rankings.

Ranking data arise when a group of judges is asked to rank a fixed set of objects (*items*) according to their preferences. When ranking k items, labeled $1, \dots, k$, a ranking π is a mapping function from the set of items $\{1, \dots, k\}$ to the set of ranks $\{1, \dots, k\}$, endowed with the

natural ordering of integers, where $\pi(i)$ is the rank given by the judge to item i ¹. When all k items are ranked in k distinct ranks, we observe a complete ranking or *linear ordering* (Cook et al., 1986). Yet, it is also possible that a judge fails to distinguish between two or more objects and assigns them equally, thus resulting in a tied ranking or *weak ordering*. Besides complete and tied rankings, *partial* and *incomplete rankings* exist: the first occurs when only a specific subset of $q < k$ objects are ranked by judges, while incomplete ranking occurs when judges are free to rank different subsets of k objects (Cook et al., 1986). Obviously, different types of ordering will generate different sample space of ranking data. With k objects there are $k!$ possible complete rankings; this number gets even larger when ties are allowed (for the cardinality of the universe when ties are allowed refer to Good (1980) and Marcus (2013)). From a methodological point of view, preference analysis often models the probability for certain preference structures, finally providing the probabilities for choosing one single object. Many models have been proposed over the years, such as order statistics models, distance-based models and Bradley-Terry models. Moreover, in order to incorporate subject specific covariates, extension of the above mentioned models have been proposed, such as distance based tree models (Lee and Philip, 2010), decision tree models with ad-hoc impurity functions (Yu et al., 2011; Plaia and Sciandra, 2017), distance-based multivariate trees for rankings (D'Ambrosio and Heiser, 2016) and some log-linear version of standard Bradley-Terry models (Dittrich et al., 1998). Recently, model-based clustering algorithms to analyse and explore ranking data have been proposed in literature (Jacques and Biernacki, 2014; Biernacki and Jacques, 2013). Yet, it is important to note that the classical cluster algorithm not always can be extended to preference data, because the concept of clustering for this type of data is not unique: in presence of preference data, clustering can be done over the individuals or over the objects. Often rank data can reveal simultaneous clusters of both individuals and items.

3 Projection pursuit

Projection pursuit includes a lot of techniques for finding interesting projections of multivariate data in low dimensional projections (Friedman and Tukey, 1974). One particular structure is that of clusters in the data. Projection Pursuit Clustering (PPC) is a synthesis of projection pursuit and nonhierarchical clustering methods that simultaneously attempts to cluster the data and to find a low-dimensional representation of this cluster structure.

One of the most important features of PP is that it is one of the few multivariate methods able to bypass the “curse of dimensionality” problem. Many of the methods of classical multivariate analysis turn out to be special cases of PP, for example principal components and discriminant analysis.

How does PP work? When PP is performed on a small number of dimension, it is possible to examine essentially all such projections to select those of interest: the appearance of the

¹ Preference rankings can be represented through either rank vectors (as in this paper) or order vectors D'Ambrosio et al. (2015).

projected data set does not change abruptly as the projection direction changes, and the space of projection directions will be of low dimensionality. When the projection is made up in higher dimensions, the appearance of the projected data will still smoothly change, but it becomes increasingly impractical to explore possible projections exhaustively (Tukey and Tukey, 1981). Projection pursuit works by associating a function value to each and every low-dimensional projection. This function value must be a measure of "interestingness" so it should be large for projections revealing interesting structure, and small for uninteresting ones. Then, PP could be defined as the process of making such selections by the local optimisation over projection directions of some index of "interestingness". The notion of "interesting projection" has usually been defined referring to departure from normality (Huber, 1985), but several alternatives have been proposed also looking for multimodality (Nason and Sibson, 1992) or clustering. Once an objective function I , called *projection index* and depending on a normalized projection vector α , is assigned to every projection characterizing the structure present in the projection, interesting projection are then automatically picked up through a numerical optimization of the projection index. One of the most common problems in PP is the oscillating behaviour of the projection indices: often procedures looking for the most interesting projection stop at the nearest local optimum from the starting point. So several authors devoted their works to avoid the local property of the optimization algorithm (Hall, 1988; Posse, 1995). A way for catching all and only all significantly interesting projections is to extract them in a monotonic way from the most structured projection to the less but still useful solution. In its classical notation a PP can be summarized as follows. Let X be either a P -dimensional random vector (distributional) or some $P \times N$ data matrix (sample). To form a univariate linear projection of X onto the real line we require a P -vector a . This vector might as well be of unit length, since it is only the direction of projection that is of interest. The projected data, Z , are formed by $Z = a^\top X$. For a linear projection onto K ($K < P$) dimensions we require a $P \times K$ matrix A , and the projected data, Z , are formed by $Z = A^\top X$. If the columns of A form an orthonormal set then the projection will be orthogonal.

The measure of "interestingness" evaluated by the projection index I , then will be expressed as

$$I(Z) = I(A^\top X) = I(A).$$

These interesting projections will be evidence of structure within the multivariate set and may form the basis of hypotheses which may be confirmed by more traditional statistical methods.

3.1 Projection Indexes

The aim of projection pursuit is to reveal possible non-linear and therefore interesting structures hidden in the high-dimensional data. As introduced before, to what extent these structures are "interesting" is measured by an index. Principal components analysis, for example, can be seen as a projection pursuit method in which the index of "interestingness" $I(a)$ is in this case the proportion of the total variance accounted for by a linear combination $a^\top X$ subject to the

normalizing constraint $a^\top a = 1$. In this particular case, this projection index is simple to maximize and has an algebraic solution; however this is the exception rather than the rule. Most projection indexes require an algorithm that will calculate I at values of a and maximize I according to some numerical optimization routine.

Several projection indexes have been proposed in literature. Since the work of Huber (1985), and more recently Hall and Li (1993), the notion of an “interesting projection” has been clearly defined as one exhibiting departure from normality. Consequently, test for non-normality were thought as suitable projection indexes. But it has also been shown that in order a projection index to be considered efficient, must satisfy basic requirements, namely affine invariance (Huber, 1985), consistency (Hall, 1988), simplicity and sensitivity to departure from normality in the core rather than in the tails of the distribution. Friedman and Tukey (1974) developed a hill-climbing optimisation methods to find interesting projections. The index they used for 1-dimensional projection pursuit can be written as a combination of two components $I(a) = s(a)d(a)$, where $s(a)$ measures the general spread of the data, and $d(a)$ measures the local density of the data after projection onto a projection vector a . In defining a projection index, Friedman and Tukey (1974) thought was interesting within a projection and tried to optimise a projection index to maximise this; as an alternative, Jones and Sibson (1987) defined a measure of un-interesting projections and attempted to maximise divergence away from it. Other projection indexes were based on some measure of entropy (Yenyukov, 1989); Jones and Sibson (1987) developed an approximation to the entropy index, called the moment index, which is based on summary statistics of the data (more precisely the third and fourth outer product tensors). Very few projection pursuit indices incorporate class or group information in the calculation. Hence, they cannot be adequately applied in *supervised* classification problems to provide low dimensional projections revealing class differences in the data. Aim of the projection Pursuit Clustering (PPC) is to recover clusters in lower dimensional subspaces of the data by simultaneously performing dimension reduction and clustering. Therefore, results from a PPC algorithm could make possible to use them as a first step of *unsupervised* classification problems.

4 Projection pursuit clustering

Bolton and Krzanowski (2003) define Projection Pursuit Clustering as the synthesis of a projection pursuit algorithm and nonhierarchical clustering methods that simultaneously attempts to cluster the data and to find a low-dimensional representation of this cluster structure. Aim of the PPC is to seek, in high dimensional data, a few interesting low-dimensional projection that reveal differences between classes. PPC works as follows: iteratively it finds both an optimal clustering for a subspace of given dimension and an optimal subspace for this clustering. Some authors have already associated PP with clustering; for example Eslava and Marriott (1994) proposed the use in a PP algorithm of projection indexes with class information able to uncover low-dimensional cluster structure; Lee et al. (2005) proposed the LDA (linear dis-

criminant analysis) projection pursuit index using class information through an extension of the linear discriminant analysis idea. Lee and Cook (2010) developed a Penalized Discriminant Analysis (PDA) index useful when data exhibit high correlation data or for situations with a small number of observations over a large number of variables.

Other contributions looked at MDS (multidimensional scaling) in terms of projection pursuit by identifying the *stress* function with the projection index and constrain the multidimensional configuration to orthogonal projections of the data Borg and Groenen (1997). In a more recent work Lee and Philip (2010) developed a projection pursuit classification tree, a new approach to build a classification tree using projection pursuit indices with class information. A PP step is performed at each node so that the best projection is used to separate two groups of classes using various projection pursuit indices with class information. One class is assigned to only one final node and the depth of the projection pursuit classification tree cannot be greater than the number of classes. The projection coefficients of each node can be interpreted as the importance of the variables to the class separation of each node; then the way in which these coefficients change should be useful to explore how classes are separated in a tree.

5 Clustering preference data

In dealing with preference rankings, one of the main issues is to identify homogeneous sub-populations of judges when heterogeneity among them is assumed. This is exactly the goal of clustering methods. Projection pursuit-based clustering methods have been proposed over the year in order to deal with a large variety of data (Friedman and Tukey, 1974; Bock, 1987; Heiser and Groenen, 1997; Miasnikov et al., 2004). As a matter of fact, there are no proposals that allow to deal with preference data. Preference rankings are characterized by a set of items, or objects, and a set of judges, or individuals, that have to rank the items according to their preference. Clustering methods for preference rankings can be done over the individuals (Murphy and Martin, 2003; Jacques and Biernacki, 2014), or over the objects (Marden, 2014). Often rank data can reveal simultaneous clusters of both individuals and items. Multidimensional unfolding techniques can graphically show such a situation (De Soete and Heiser, 1993). Here we combine suitable clustering methods for preference rankings with Multidimensional unfolding techniques. Our approach is similar to the Cluster Differences Unfolding (CDU) (Vera et al., 2013), which can be considered as the natural extension to Unfolding of the Cluster Difference Scaling (CDS) (Heiser, 1993). The main difference is that CDU, which is devoted to metric Unfolding, performs a cluster analysis over both the sets of individuals and objects, producing a configuration plot that shows the bari-centers of the sets retaining their preference relationship.

Here we propose an iterative strategy that performs a non-hierarchical cluster analysis on only one set, typically the individuals, leaving the other set free to be configured in the reduced geometrical space in such a way that the relationships between the preference order of the individuals with respect to the items remain unchanged. We call our proposal *Projection Clustering Unfolding* (PCU).

5.1 The *Projection Clustering Unfolding* (PCU)

Unfolding can be seen as a particular Multidimensional Scaling technique for rectangular data, in general showing preference of n persons for m objects. The most accepted formulation of the problem in terms of a badness-of-fit function is given in a least squares framework by the minimization of the *stress* function (Kruskal, 1964), defined as

$$\sigma^2(\mathbf{A}, \mathbf{B}, \hat{\mathbf{\Delta}}) = \sum_{i=1}^n \sum_{j=1}^m (\hat{\delta}_{ij} - d_{ij})^2, \quad (1)$$

where $\hat{\mathbf{\Delta}}$ is a $n \times m$ matrix in which each entry $\hat{\delta}_{ij}$ represents the disparity or monotonically transformed dissimilarity between the i th individual and the j th item, and $d_{ij} = d_{ij}(\mathbf{A}, \mathbf{B})$ represents the Euclidean distance between the individuals' (\mathbf{A}) and items (\mathbf{B}) configuration points in P -dimensional space, $i = 1, \dots, n$, $j = 1, \dots, m$ (Borg and Groenen, 1997).

Here we assume that $\mathbf{A} = \mathbf{G}\mathbf{X}$, where \mathbf{G} is a $n \times K$ indicator matrix whose elements g_{ik} , $k = 1, \dots, K$, are equal to one if the i th individual belongs to the k th cluster, and zero otherwise. We assume that $g_{ik} \cap g_{il} = \emptyset$ for $k \neq l = 1, \dots, K \forall i = 1, \dots, n$. \mathbf{X} is the $K \times P$ matrix of the bari-centers of the K clusters, where P indicates the dimension of the Unfolding solution. We propose an alternating optimization strategy that, given a configuration of both the individuals and the items, searches the optimum partition of the individuals in K clusters. Then, given the optimal partition of the individuals, the configuration of both individuals and items are updated. The first step consists in a first unfolding configuration with a random assignment of the individuals to the K clusters. As Unfolding model, we use the PREFSCAL algorithm (Busing et al., 2005), which is particularly feasible when dealing with ordinal Unfolding, that penalizes the stress function and uses the SMACOF-3 algorithm (Heiser, 1981) as optimization engine.

5.2 The *Projection Clustering Unfolding*: a real example

As an example, we analyse the well-known breakfast data set (Green and Rao, 1972). Breakfast data contains the preferences of 42 individuals towards 15 breakfast items from the most preferred (1) to the least preferred (15). We set $K = 4$ clusters and the simplest approach to the ties, i.e. untie tied observations. As the final solution is sensitive to the random choice of the clusters at the first step, we repeated the analysis 20 times, obtaining the configuration shown in Figure 1.

The figure shows the configuration of the 4 cluster centers in the 2-dimensional solution. We expect that the closer the bari-centers are to the items, the higher is the preference of that cluster to that items. We ran the Unfolding analysis without any restrictions on the same data, and then we performed a Procrustes analysis (Borg and Lingoes, 1987) by considering the unrestricted solution as target configuration. Procrustes analysis allows to evaluate the ability to reproduce the configuration both graphically and with the L -statistic, which is the

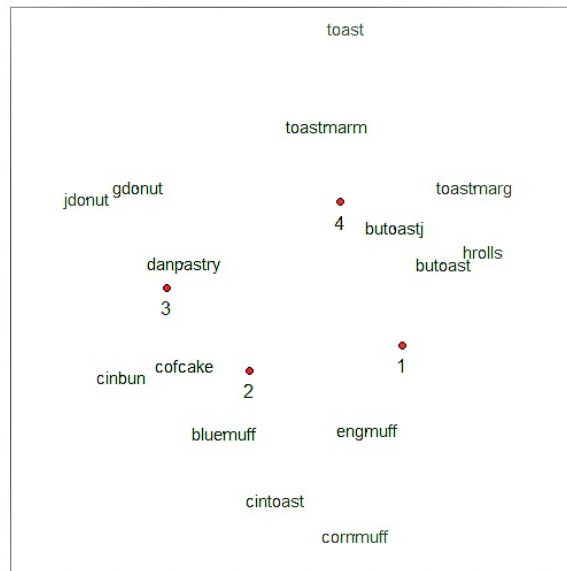


Fig. 1 Projection Clustering Unfolding solution. Items are coded as: toast=toast pop-up; butoast=buttered toast; engmuff=English muffin and margarine; jdonut=jelly donut; cintoast=cinnamon toast; bluemuff=blueberry muffin and margarine; hrolls=hard rolls and butter; toastmarm=toast and marmalade; butoastj=buttered toast and jelly; toastmarg=toast and margarine; cinbun=cinnamon bun; danpastry=Danish pastry; gdonut=glazed donut; cofcake=coffee cake; cornmuff=corn muffin and butter.

sum of the squared differences between the true and the fitted configuration after that both configurations are set into optimal correspondence through Procrustean transformation. The lower the Procrustes statistic, the better the fitted configuration. We used a normalized version of the Procrustes statistic as suggested by Deun et al. (2007):

$$L(\mathbf{X}, \hat{\mathbf{X}}) = \frac{\text{tr}((\mathbf{X} - \hat{\mathbf{X}})^T(\mathbf{X} - \hat{\mathbf{X}}))}{\text{tr}(\mathbf{X}^T \mathbf{X})}, \quad (2)$$

where \mathbf{X} is the true configuration and $\hat{\mathbf{X}}$ is the fitted one.

Figure 2 shows the Procrustes configuration plot limited to only the objects points. The recovery is excellent, also confirmed by a value of $L = 0.013$.

Figure 3 shows the Procrustes plot for the individuals' point configurations ($L = 0.161$). This figure gives an idea of the composition of the clusters in terms of the allocation of the individuals around the cluster centers.

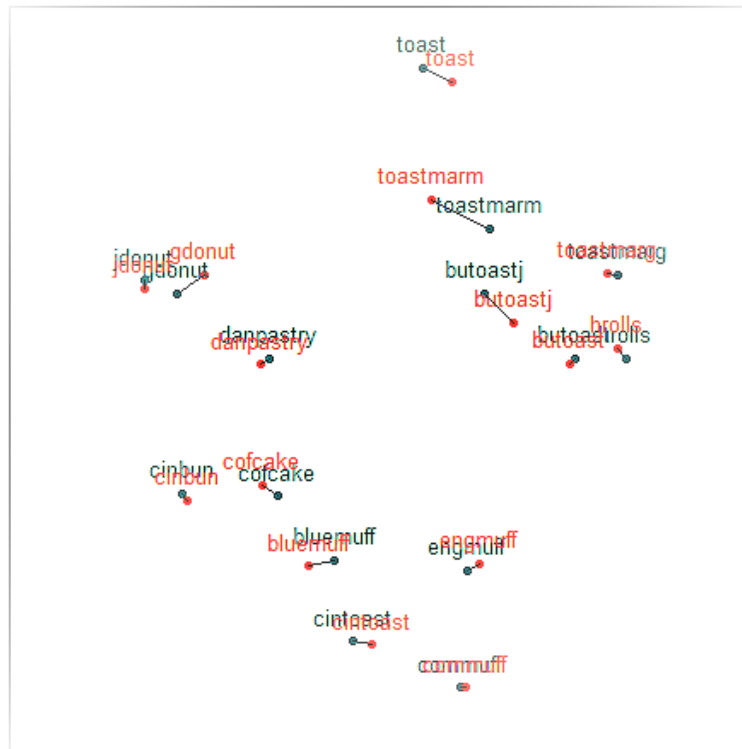


Fig. 2 Procrustes configuration plot: objects' points unfolding configuration (green) vs objects' points PCU solution (red).

Figure 4 shows the overall Procrustes configuration plot ($L = 0.091$). This graphical representation shows that the PCU solution (red) is really similar to the unrestricted Unfolding analysis (red) in terms of interpretation.

Both the graphical representation and the L -statistic confirm that the PCU procedure does not distort the original Unfolding analysis. Of course, the technical settings have been set equal for both Unfolding and PCU.

After the Procrustes analysis we performed a last allocation step. We first computed the squared Euclidean distance between the unrestricted unfolding solution and the (fitted and scaled) PCU configuration, and then we assigned the individuals to the clusters with a procedure similar to the one of the K-means. We obtained the following confusion matrix:

| | | Unfolding Alignment | | | |
|-----|-----------|---------------------|-----------|-----------|-----------|
| | | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
| PCU | Cluster 1 | 8 | 0 | 0 | 0 |
| | Cluster 2 | 0 | 12 | 0 | 0 |
| | Cluster 3 | 0 | 0 | 16 | 0 |
| | Cluster 4 | 0 | 0 | 1 | 5 |

Table 1 Median ranking within cluster

| Item | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Median ranking Breakfast data |
|-----------|-----------|-----------|-----------|-----------|-------------------------------|
| toast | 13 | 13 | 15 | 7 | 15 |
| butoast | 1 | 10 | 11 | 1 | 11 |
| engmuff | 2 | 3 | 7 | 8 | 6 |
| jdonut | 15 | 12 | 3 | 6 | 7 |
| cintoast | 5 | 5 | 8 | 13 | 8 |
| bluemuff | 6 | 2 | 6 | 9 | 3 |
| hrolls | 4 | 9 | 14 | 2 | 12 |
| toastmarm | 8 | 8 | 10 | 2 | 10 |
| butoastj | 7 | 7 | 9 | 2 | 9 |
| toastmarg | 3 | 11 | 12 | 3 | 13 |
| cinbun | 12 | 4 | 5 | 10 | 4 |
| danpastry | 10 | 4 | 1 | 4 | 1 |
| gdonut | 14 | 6 | 4 | 5 | 5 |
| cofcake | 9 | 1 | 2 | 11 | 2 |
| cornmuff | 11 | 6 | 13 | 12 | 14 |
| τ_X | 0.554 | 0.520 | 0.647 | 0.514 | 0.306 |

References

- Biernacki, C. and J. Jacques (2013). A generative model for rank data based on insertion sort algorithm. *Computational Statistics & Data Analysis* 58, 162 – 176.
- Bock, H. H. (1987). *On the Interface between Cluster Analysis, Principal Component Analysis, and Multidimensional Scaling*, pp. 17–34. Dordrecht: Springer Netherlands.
- Bolton, R. J. and W. J. Krzanowski (2003). Projection pursuit clustering for exploratory data analysis. *Journal of Computational and Graphical Statistics* 12(1), 121–142.
- Borg, I. and P. Groenen (1997). *Modern Multidimensional Scaling: Theory and Applications*. Springer, Statistic Series, New York.
- Borg, I. and J. Lingoes (1987). *Multidimensional similarity structure analysis*. Springer-Verlag New York.
- Busing, F. M. T. A., P. J. K. Groenen, and W. J. Heiser (2005, Mar). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika* 70(1), 71–98.
- Cook, W., M. Kress, and L. M. Seiford (1986). An axiomatic approach to distance on partial orderings. *Revue Francaise d'Automatique, d'Informatique et De Recherche Operationnelle* 20(2), 115–122.
- D'Ambrosio, A., S. Amodio, and C. Iorio (2015). Two algorithms for finding optimal solutions of the kemeny rank aggregation problem for full rank. *Electronic Journal of Applied Statistical Analysis* 8(2).

- D'Ambrosio, A. and W. J. Heiser (2016, Sep). A recursive partitioning method for the prediction of preference rankings based upon kemeny distances. *Psychometrika* 81(3), 774–794.
- De Soete, G. and W. J. Heiser (1993, Dec). A latent class unfolding model for analyzing single stimulus preference ratings. *Psychometrika* 58(4), 545–565.
- Deun, K. V., W. J. Heiser, and L. Delbeke (2007). Multidimensional unfolding by nonmetric multidimensional scaling of spearman distances in the extended permutation polytope. *Multivariate Behavioral Research* 42(1), 103–132.
- Dittrich, R., R. Hatzinger, and W. Katzenbeisser (1998). Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Journal of the Royal Statistical Society. Series C* 47(4), 511–525.
- Emond, E. J. and D. W. Mason (2002). A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis* 11(1), 17–28.
- Eslava, G. and F. H. C. Marriott (1994, Mar). Some criteria for projection pursuit. *Statistics and Computing* 4(1), 13–20.
- Friedman, J. and J. Tukey (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers* 23, 881–889.
- Good, I. J. (1980). The number of orderings of n candidates when ties and omissions are both allowed. *Journal of Statistical Computation and Simulation* 10(2), 159–160.
- Green, P. E. and V. Rao (1972). *Applied multidimensional scaling*. Hinsdale, IL: Dryden.
- Hall, P. (1988, Dec). Estimating the direction in which a data set is most interesting. *Probability Theory and Related Fields* 80(1), 51–77.
- Heiser, W. (1981). Unfolding analysis of proximity data. *Unpublished doctoral dissertation, University of Leiden, The Netherlands*.
- Heiser, W. J. (1993). Clustering in low-dimensional space. In O. Opitz, B. Lausen, and R. Klar (Eds.), *Information and Classification*, Berlin, Heidelberg, pp. 162–173. Springer Berlin Heidelberg.
- Heiser, W. J. and P. J. F. Groenen (1997, Mar). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika* 62(1), 63–83.
- Huber, P. J. (1985). Projection pursuit. *Ann. Statist.* 13(2), 435–475.
- Jacques, J. and C. Biernacki (2014). Model-based clustering for multivariate partial ranking data. *Journal of Statistical Planning and Inference* 149, 201 – 217.

- Jones, M. C. and R. Sibson (1987). What is projection pursuit? *Journal of the Royal Statistical Society. Series A (General)* 150(1), 1–37.
- Kruskal, J. B. (1964, Mar). Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika* 29(1), 1–27.
- Lee, E. and D. Cook (2010, Jul). A projection pursuit index for large p small n data. *Statistics and Computing* 20(3), 381–392.
- Lee, E., D. Cook, S. Klinke, and T. Lumley (2005). Projection pursuit for exploratory supervised classification. *Journal of Computational and Graphical Statistics* 14(4), 831–846.
- Lee, P. H. and L. Y. Philip (2010). Distance-based tree models for ranking data. *Computational Statistics & Data Analysis* 54(6), 1672 – 1682.
- Marcus, P. (2013). Comparison of heterogeneous probability models for ranking data. *Master thesis.*
- Marden, J. I. (2014). *Analyzing and modeling rank data*. Chapman and Hall/CRC.
- Miasnikov, A. D., J. E. Rome, and R. M. Haralick (2004). A hierarchical projection pursuit clustering algorithm. *ICPR 2004. Proceedings of the 17th International Conference 1*, 268–271.
- Murphy, T. B. and D. Martin (2003). Mixtures of distance-based models for ranking data. *Computational Statistics & Data Analysis* 41(3), 645 – 655.
- Nason, G. P. and R. Sibson (1992, Sep). Measuring multimodality. *Statistics and Computing* 2(3), 153–160.
- Plaia, A. and M. Sciandra (2017). Weighted distance-based trees for ranking data. *Advances in Data Analysis and Classification*.
- Posse, C. (1995). Tools for two-dimensional exploratory projection pursuit. *Journal of Computational and Graphical Statistics* 4(2), 83–100.
- Tukey, P. A. and J. W. Tukey (1981). *Preparation; prechosen sequences of views*, pp. 189–213. Number 3. V. Barnett, editor, *Interpreting Multivariate Data*.
- Vera, J. F., R. Macías, and W. J. Heiser (2013, Oct). Cluster differences unfolding for two-way two-mode preference rating data. *Journal of Classification* 30(3), 370–396.
- Yenyukov, I. S. (1989). *Indices for projection pursuit*, pp. 181–188. Nova Science Publishers, New York.
- Yu, P. L. H., W. M. Wan, and P. H. Lee (2011). *Decision Tree Modeling for Ranking Data*, pp. 83–106. Berlin, Heidelberg: Springer Berlin Heidelberg.