



A Kemeny Distance-Based Robust Fuzzy Clustering for Preference Data

Pierpaolo D'Urso¹ · Vincenzina Vitale¹

Accepted: 30 August 2022 / Published online: 7 October 2022
© The Author(s) 2022

Abstract

We propose two robust fuzzy clustering techniques in the context of preference rankings to group judges into homogeneous clusters even in the case of contamination due to outliers or, more generally, noisy data. The two fuzzy C-Medoids clustering methods, based on the same suitable exponential transformation of the Kemeny distance, belong to two different approaches and differ in the way they introduce the fuzziness in the membership matrix, the one based on the “m” exponent and the other on the Shannon entropy. As far as the Kemeny distance is concerned, it is equivalent to the Kendall distance in the case of complete rankings but differs from the latter in the way of handling tied rankings. Simulations prove that our methods are able to recover the natural structure of the groups neutralizing the effect of possible noises and outliers. Two applications to real datasets are also provided.

Keywords Preference data · Kemeny distance · Fuzzy clustering · Robust method

1 Introduction

In the last decades, many research studies focus their attention on how a group of subjects ranks a list of objects according to each personal preference (the rank vector). Hence, a rank vector is a permutation of the positive integers $1, \dots, k$. More rigorously, given k items to rank, labelled $1, \dots, k$, a ranking π is a mapping function from the set of items $1, \dots, k$ to the set of ranks $1, \dots, k$, where $\pi(i)$ is the rank given to item i . If $\pi(i) < \pi(i')$, then item i is said to be preferred to item i' . An alternative format for ranking data is the ordering vector that is, more simply, the inverse function of ranking π so that the generic $\pi^{-1}(j)$ denotes the item ranked in the j th position.

A judge can assign a distinct value to each of the k items defining a complete ranking or he can specify only $m < k$ items, defining a partial ranking. This is the case where the judge specifies only his most-liked $m < k$ items, leaving the remaining ones undefined. Different

✉ Pierpaolo D'Urso
pierpaolo.durso@uniroma1.it

¹ Department of Social Sciences and Economics, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Rome, Italy

is the case of *ties*, when a judge assigns the same integer value to two or more items (He evaluates them equally).

Therefore, we could be interested, for example, to rank candidates in an election or to rank shopping items in order to identify the pattern of preferences as well as the most representative permutation. More generally, involving the concept of preference, this kind of data has a broad field of application: from Marketing, Risk and Credit scoring to Psychology, Sociology, Politics, and so on.

Preference analysis literature mainly focuses on modelling the probability for certain preference structures under the assumption of a homogeneous population. The relative statistical models can be classified in four classes (Critchlow et al., 1991; Marden, 1995): order statistics models (Thurstone, 1927), rankings induced by paired comparisons (Bradley, 1984; Bradley & Terry, 1952), distance-based models (Smith, 1950; Mallows, 1957), and multistage models (Luce, 1959; Plackett, 1975; Fligner & Verducci, 1988).

In this work, we assume the heterogeneity in the population of judges and, therefore, the goal is to identify subgroups of judges that are homogeneous in terms of their preferences following a model-free clustering approach in the fuzzy framework.

We use the Kemeny distance (Kemeny & Snell, 1962) to enhance flexibility and model applicability. Unlike Kendall distance, it is a metric also in the more general case of tied rankings while the Kendall distance satisfies the triangular inequality only with full rankings (Emond & Mason, 2002). However, as pointed out by Heiser and D'Ambrosio (2013), in the latter case, the two distances become equivalent being the Kemeny distance twice the Kendall distance. But handling preference data under the assumption of a heterogeneous population is only one of the main goals of this work; in this sense, our methodology has a twofold scope: to adapt the fuzzy C-Medoids (FCMd) method to the case of ranking data and to account for noisy data and outliers.

To the best of our knowledge, there is no extensive literature about clustering techniques for preference data using a model-free approach; the most similar approach is that proposed, recently, by D'Ambrosio and Heiser (2019) in which a K-Median Cluster Component Analysis based on the Kemeny distance is presented. In the work of Müllensiefen et al. (2018), there is an application of the Partitioning Around Medoids (PAM) algorithm for clustering rankings based on a weighted footrule distance. Other interesting works based on the distance between rankings in hierarchical cluster analysis can be found in Brentari et al. (2016) and Bonanomi et al. (2017, 2019). In contrast, most works concern a mixture model-based approach. In the class of distance-based models, the extension to mixture models for heterogeneous populations is due to Murphy and Martin (2003), Meila and Chen (2010), and Lee and Yu (2012). Other mixture model-based clustering algorithms to analyse ranking data have been proposed by Biernacki and Jacques (2013), Jacques and Biernacki (2014), and Mollica and Tardella (2017).

Franczak et al. (2016) well demonstrated the advantages of a model-based imputation procedure that simultaneously accounts for heterogeneity while imputing. They proved that their model-based approach is able to recover the group structure and key features of the datasets also in presence of a huge amount of outliers.

It is worth noting that, to the best of our knowledge, there is no literature in the field of clustering and fuzzy clustering techniques that proposes a robust metric for rankings. In order to identify homogeneous groups of judges even when some contamination, due to outliers or more generally noises, is present in the data, we propose two robust fuzzy clustering techniques both based on the same suitable exponential transformation of the Kemeny distance. They differ in the specification of the fuzziness.

In detail, the first proposal is an extension of the well-known FCMD method, which uses the “ m ” exponent to transform the matrix of the crisp assignment to that of membership degree (Krishnapuram et al., 1999; Krishnapuram et al., 2001) while the second one is an extension of the fuzzy C-Means clustering technique with entropy regularization (Li and Mukaidono 1995, 1999; Miyamoto & Mukaidono 1997), i.e. the Shannon entropy is introduced in the objective function to obtain a fuzzy partition avoiding the use of the exponent “ m ”. Both techniques are extended to deal with rankings (full and tied) and, moreover, noisy data; indeed, robustness is achieved by defining a suitable exponential transformation of the Kemeny distance. As far as the latter model is concerned, it is worth noticing that the extension to the fuzzy medoids-based version is also introduced.

We named, henceforth, the first proposal as the “Exponential Kemeny-based FCMD method” while the second one as the “Exponential Kemeny-based FCMD method with entropy regularization”.

We argue that the proposed clustering methods could be very useful in many contexts, above all in marketing research; we believe they can be a valid instrument for market segmentation, whose goal is that of producing different strategies for different customer segments in order to offer them appropriate products and/or services. Moreover, the detection of clustered outliers could help to identify niche markets, customer segments with very particular needs and tastes.

The outline of the article is as follows. After a brief introduction to the literature about the fuzzy clustering techniques in Section 2, our methodological proposals are introduced and described in Section 3; simulation results are shown in detail in Section 4 while the application to two real datasets is proposed in Section 5. In the last section (Section 6), we address some conclusions and open research problems.

2 Fuzzy Clustering Techniques

Cluster analysis based on fuzzy theory (Zadeh, 1965) allows units to belong to more than one cluster simultaneously as opposed to the classic clustering approach whose partitions are characterized by non-empty and pairwise disjoint subsets, providing a crisp assignment of the units to the clusters.

In more detail, when an object is almost equally distant from two or more clusters, the fuzzy approach relaxes the requirement of a crisp assignment, replaced by the notion of degree of membership in a cluster¹, particularly useful in case of overlapping or when the goal is to group complex objects characterized by an unavoidable and intrinsic “imprecision”.

The interest for fuzzy clustering methods quickly grew over time so that many fuzzy clustering algorithms were proposed; seminal papers were due to Bellman et al. (1966) and Ruspini (1969, 1970, 1973) even if the most representative one was the fuzzy C-Means (FCM) clustering method (Dunn 1973; Bezdek 1974, 1981), widely applied to this day. The “fuzzification” is obtained by rising each unknown degree of membership to an “ m ” exponent which is named *fuzziness coefficient* because it controls the extent of fuzziness degree of the partition.

¹ In general, one assumes that the membership degrees associated to the object and the selected clusters sum to one.

Another variant of the C-Means clustering method in a fuzzy perspective was introduced by Li and Mukaidono (1995, 1999) and Miyamoto and Mukaidono (1997) as an alternative to the fuzzy methods based on the “ m ” exponent; they answer to the criticism moved by some researchers towards the role of the fuzziness coefficient often considered an artificial device, unnatural and with no physical meaning.

To overcome this limit, in the latter class of methods, fuzziness is controlled by including in the objective function² an entropy regularization term (Miyamoto & Mukaidono, 1997), i.e. the Shannon entropy which, when applied to the degrees of membership, can be called fuzzy entropy (Coppi & D’Urso, 2006).

Following this approach, the total functional is optimized by both maximizing the internal cohesion and the given measure of entropy, thus maximizing the total amount of information.

Other interesting prototype-based models, which are variants of the FCM, have also been proposed in the literature and we refer to the medoids-based clustering technique that, in a non-fuzzy framework, was due to Vinod (1969), Church (1978), Mulvey and Crowder (1979), Rao (1971), and Kaufman and Rousseeuw (1987) (the latter proposed the well-known Partitioning Around Medoids method (PAM)) while, staying within a fuzzy context, we can refer to the seminal papers of Krishnapuram et al. (1999, 2001).

The fuzzy C-Medoids (FCMd) clustering techniques group objects around representative prototypes observed in the dataset, i.e. the medoids, that synthesize the structural information of each cluster. Therefore, the medoids are objects whose overall distance from all other objects in the same cluster is minimal: this prevents prototypes from being “virtual”, as in the case of the centroids in the C-Means algorithm. While a first advantage certainly relates to practical applications for which the identification of representative non-fictitious prototypes can be very interesting and useful for the interpretation of the clusters, the main strength comes from the consideration that the FCMd method is more robust than the FCM if noise or outliers occur in the data, the medoid being less influenced by such extreme values than the mean. Readers interested in a deeper and more detailed dissertation on fuzzy clustering may refer to D’Urso (2015) and references therein.

In the next section, we will move on to the description of the proposed methods, which are methodological extensions of the fuzzy medoids-based methods previously described; the purpose is to adapt them to the case where one is interested in the group’s structure of particular data such as preference data, considering the possibility that outliers or noisy data may mask the real composition of the groups themselves, both in terms of their number and in terms of membership of the units to the groups.

3 Robust FCMd Clustering for Preference Data

Clustering ranking data aims at identifying groups of individuals characterized by similar preferences or choices with respect to a set of items.

In clustering preference data, the features of permutations themselves suggest the use of a fuzzy approach as a natural way to cope with the uncertainty of assigning a judge (i.e. a permutation) to each cluster.

² with $m = 1$

In this context, the FCMd clustering technique seems the most appropriate choice since the identification of an observed permutation acting as the cluster prototype could improve the interpretation of the selected cluster.

As far as the distance matrix is concerned, we argue that the choice of the appropriate metric is one of the most important issues of any fuzzy and non-fuzzy clustering method.

In this context, we propose the use of the Kemeny distance (Kemeny and Snell, 1962) that could be seen as a suitable extension of the Kendall distance (Kendall 1938, 1948) for the case of two tied rankings π and π^* over the set $[k]$ (Li et al., 2017). Therefore, given two rankings π and π^* and using the representation of each ranking by means of a $k \times k$ matrix as proposed by Kendall (1938) according to each element

$$\pi(i, j) = \begin{cases} 1 & \text{if object } i \text{ is ranked before the object } j \\ -1 & \text{if object } i \text{ is ranked behind object } j \\ 0 & \text{if the objects are tied or } i=j, \end{cases}$$

the Kemeny distance between two rankings π and π^* is defined as:

$$d(\pi, \pi^*) = \frac{1}{2} \sum_{i,j=1}^k |\pi(i, j) - \pi^*(i, j)|,$$

where the maximum value is equal to $k \times (k - 1)$. As previously said, the Kemeny distance is equivalent to the Kendall distance for full rankings, i.e. the former is twice the latter (Heiser and D'Ambrosio, 2013) but this relation does not hold in the case of tied rankings. Moreover, while the Kemeny distance is a metric also with tied rankings, the Kendall one does not satisfies the third condition of the triangular inequality.

In this study, we propose a suitable exponential transformation of the above metric to both define a robust version of the standard FCMd and the FCMd with entropy regularization.

Both models have a twofold scope: to handle preference data by grouping individuals according to their personal choices and to improve the robustness of the standard methods when dealing with outlying permutations and noisy data.

In the next paragraph, we will provide the mathematical formalization of the exponential transformation of the Kemeny distance.

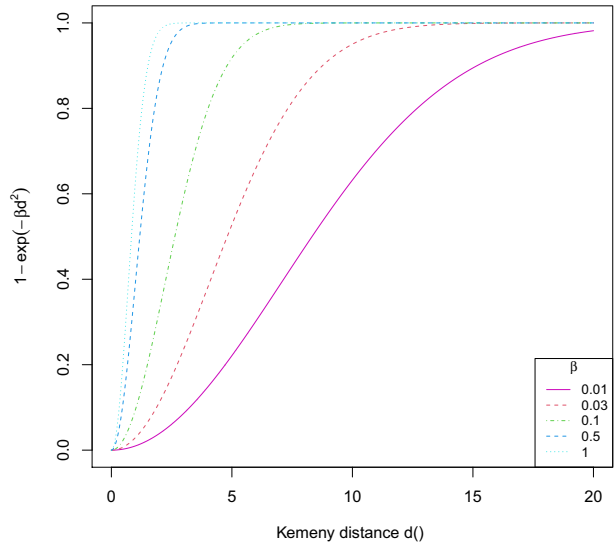
3.1 The Exponential Transformation of the Kemeny Distance

As already mentioned, one of the relevant issues in clustering is to neutralize the negative effects of noisy data as well as the outliers, that are, the latter, units that markedly deviate from the rest of the data.

Moreover, as well pointed out by García-Escudero et al. (2008), “the precise detection of the outliers is an important task due to the serious troubles they introduce in standard clustering procedures as well as the appealing interest that outliers could have by themselves after explaining why they depart from general behaviour”.

The outliers could be a group of observations, smaller than the natural clusters, that differ markedly from them (*i.e. clustered outliers*) or, alternatively, could be represented by isolated points, each forming its own group (*i.e. radial outliers*) (García-Escudero et al., 2003). In this context, the outlier is identified essentially according to its distance with respect to the centers of the clusters, both embedding the radial and clustered types while

Fig. 1 The Effect of β parameter on $d_{exp}^2(\boldsymbol{\pi}_l, \boldsymbol{\pi}_t)$



noisy datum is intended as the permutation generated by the uniform distribution, given a rankings generative scheme.

As pointed out by García-Escudero and Gordaliza (2005), the FCMd method is only a “timid” robustification of clustering in presence of outliers and, therefore, we propose two “robust” versions of the FCMd method based on a suitable exponential transformation of the Kemeny distance.

Therefore, as suggested in literature by Wu and Yang (2002) and D’Urso and De Giovanni (2014), the following exponential transformation, which lies in the interval $[0,1]$, has been applied to the Kemeny distance:

$$d_{exp}^2(\boldsymbol{\pi}_l, \boldsymbol{\pi}_t) = 1 - \exp\{-\beta \cdot d^2(\boldsymbol{\pi}_l, \boldsymbol{\pi}_t)\}_{l \neq t} \quad \text{with } l, t = 1 \dots n, \quad (1)$$

where $d^2(\boldsymbol{\pi}_l, \boldsymbol{\pi}_t)$ is the squared Kemeny distance between the ranking representation given by the l -th judge and the ranking representation given by the t -th judge, respectively.

The parameter β is a positive constant usually chosen as the inverse of a measure of data variability, as pointed out by Wu and Yang (2002). While the mathematical definition of β provided in this work will be described in Section 3.2.1, its effect on the exponential transformation (1) is shown in Fig. 1: it reaches more rapidly 1 the higher is the value of β .

When using the exponential transformation in the fuzzy clustering, the membership degrees associated with the outliers are approximately equal to $1/C$ (where C is the number of clusters), thus dealing with outlying units as fuzzy ones.

It is worth noting that, in the case of well separated groups, i.e. in presence of low variability, the behaviour of the clustering algorithm based on the exponential transformation tends to assign approximately equal membership degree values to all units far from the medoids. Each unit not close to medoids becomes a candidate outlier.

On the contrary, in case of overlapping clusters or well-separated clusters but with a large amount of outlying units, the algorithm tends to assign approximately equal membership degree values to the units that are only slightly separated from the bulk of data (D’Urso et al., 2018).

3.2 Exponential Kemeny-Based FCMD Method

The exponential transformation of the Kemeny distance (1) is used to develop the so-called Exponential Kemeny-based fuzzy C-Medoids clustering method (*Exp-FCMDK*); the goal is to find the C prototypes, i.e. the subset of medoids (π_1, \dots, π_C) , where C is the number of clusters, and the $U_{n \times C}$ matrix of fuzzy coefficients, by minimizing the following objective function:

$$\left\{ \begin{array}{l} \min : \quad \sum_{l=1}^n \sum_{c=1}^C u_{lc}^m d_{exp}^2(\pi_l, \tilde{\pi}_c) = \\ \quad \sum_{l=1}^n \sum_{c=1}^C u_{lc}^m \left\{ 1 - \exp \left[-\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} \\ \text{s.t.} \quad \sum_{c=1}^C u_{lc} = 1, u_{lc} \geq 0. \end{array} \right. \quad (2)$$

In detail, $\left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)|\right)^2$ is the squared Kemeny distance between the l -th unit and the medoid of the c -th cluster.

$u_{lc} \in [0,1]$ denotes the membership degree of the l -th unit to the c -th cluster while $\tilde{\pi}_c$ is the permutation medoid for the cluster c . The “ m ” parameter (with $m > 1$) controls the fuzziness of the partition and therefore usually named as “fuzziness parameter” (for further insight on the role of m see D’Urso (2015)).

The solution for each u_{lc} is:

$$u_{lc} = \frac{1}{\sum_{c'=1}^C \left[\frac{1 - \exp \left[-\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_{c'}(i,j)| \right)^2 \right]}{1 - \exp \left[-\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right]} \right]^{\frac{1}{m-1}}} \quad (3)$$

The proofs of the iterative solutions (3) are provided in the Appendix A while the computational steps of the proposed robust clustering method are described in the Algorithm 1.

In the next paragraph, we describe the beta’s formulation used in the above method.

```

1: Fix  $C, m, \text{nstart}$  and  $\text{max.iter}$ ;
2: Set  $iter = 0$ ;
3: Randomly initialize  $U = \{u_1, \dots, u_n\}$  with  $u_l = (u_{l,1}, \dots, u_{l,C}) (l = 1, \dots, n)$ ;
4: Store the  $C$  initial medoids:  $\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C$ ;
5: repeat
6:   Store the current medoids  $(\tilde{\pi}_{i_1}, \tilde{\pi}_{i_2}, \dots, \tilde{\pi}_{i_C})_{OLD} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C)$ ;
7:   Compute  $u_l (l = 1, \dots, n)$  by using (3);
8:   For  $c = 1, \dots, C$ , determine the index  $q_c \in \{1, \dots, n\}$  satisfying:
      
$$q_c = \arg \min_{1 \leq q \leq n} \sum_{l=1}^n u_{lc}^m \left\{ 1 - \exp \left[ -\beta \left( \frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_q(i,j)| \right)^2 \right] \right\}$$

9:   return  $\tilde{\pi}_c = \pi_{q_c}$  for  $c = 1, \dots, C$ 
10:   $iter \leftarrow iter + 1$ ;
11: until  $(\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C)_{OLD} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C)$  or  $iter = \text{max.iter}$ 
12: Repeat initialization: start  $\text{nstart}$  times from step 2 and provide as final output of the algorithm the  $U$  matrix and the medoids  $\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C$  yielding the smallest value of the objective function out of these  $\text{nstart}$  random initializations.

```

Algorithm 1 Exponential Kemeny-based FCMD algorithm

3.2.1 The Choice of β Parameter

Following Wu and Yang (2002) and D’Urso et al. (2018), we choose β as the inverse of a suitable measure of variability. In particular, we propose the following β ’s formulation:

$$\beta \equiv \beta_1 = \left[\frac{\sum_{l=1}^n d^2(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_q)}{n} \right]^{-1} \tag{4}$$

where $\tilde{\boldsymbol{\pi}}_q : q = \text{argMedian}_{1 \leq l \leq n} \sum_{l'=1}^n d^2(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_{l'})$.

3.3 Exponential Kemeny-Based FCMD Method with Entropy Regularization

Here, the Exponential transformation of the Kemeny distance (1) is used to develop the so-called Exponential Kemeny-based fuzzy C-Medoids clustering method with entropy regularization (*Exp-FCMD_{ent}*); the goal is to find the C prototypes, i.e. the subset of medoids $(\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_C)$, where C is the number of clusters, and the $\mathbf{U}_{n \times C}$ matrix of fuzzy coefficients, by minimizing the following objective function:

$$\begin{cases} \min : & \sum_{l=1}^n \sum_{c=1}^C u_{lc} d_{exp}^2(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_c) + p \sum_{l=1}^n \sum_{c=1}^C u_{lc} \log(u_{lc}) = \\ & \sum_{l=1}^n \sum_{c=1}^C u_{lc} \left\{ 1 - \exp \left[-\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} + p \sum_{l=1}^n \sum_{c=1}^C u_{lc} \log(u_{lc}) \\ \text{s.t.} & \sum_{c=1}^C u_{lc} = 1, u_{lc} \geq 0, \end{cases} \tag{5}$$

where $\left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2$ is the squared Kemeny distance between the l -th unit and the medoid of the c -th cluster.

u_{lc} denotes the membership of the l -th unit to the c -th cluster while $\tilde{\boldsymbol{\pi}}_c$ is the permutation medoid for the cluster c . The second addend in the objective function represents the entropy regularization term: the Shannon entropy is weighted by a factor p , called *degree of fuzzy entropy* given that the higher is p the higher is the degree of fuzziness. It controls the contribution of the regularization function to the clustering criterion and acts as the “m” exponent in the previous fuzzy method so that $u_{lc} \in [0,1]$.

Since the aim is to both maximize the Shannon entropy and the internal cohesion, the total function is optimized by subtracting the regularization term to the clustering criterion leading to (5).

The solution for each u_{lc} is:

$$u_{lc} = \frac{1}{\sum_{c'=1}^C \left[\frac{\exp \left(\frac{1}{p} \left[1 - \exp \left\{ -\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right\} \right) \right)}{\exp \left(\frac{1}{p} \left[1 - \exp \left\{ -\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_{c'}(i,j)| \right)^2 \right\} \right) \right)} \right]} \tag{6}$$

The proofs of the iterative solutions (6) are provided in the Appendix B while the computational steps of the proposed robust clustering method are described in the Algorithm 2.

By substituting in the objective function (5) $d_{exp}^2(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_c)$ with its square root, that is:

- 1: Fix C, p, nstart and max.iter ;
- 2: Set $\text{iter} = 0$;
- 3: Randomly initialize $\mathbf{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ with $\mathbf{u}_l = (u_{l,1}, \dots, u_{l,C})$ ($l = 1, \dots, n$)
- 4: Store the C initial medoids: $\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C$
- 5: **repeat**
- 6: Store the current medoids $(\tilde{\pi}_{i_1}, \tilde{\pi}_2, \dots, \tilde{\pi}_C)_{OLD} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C)$;
- 7: Compute \mathbf{u}_l ($l = 1, \dots, n$) by using (6);
- 8: For $c = 1, \dots, C$, determine the index $q_c \in \{1, \dots, n\}$ satisfying:

$$q_c = \arg \min_{1 \leq q \leq n} \sum_{l=1}^n u_{lc} \left\{ 1 - \exp \left[-\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_q(i,j)| \right)^2 \right] \right\} + p \sum_{l=1}^n u_{lc} \log(u_{lc})$$
- 9: **return** $\tilde{\pi}_c = \pi_{q_c}$ for $c = 1, \dots, C$
- 10: $\text{iter} \leftarrow \text{iter} + 1$;
- 11: **until** $(\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C)_{OLD} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C)$ or $\text{iter} = \text{max.iter}$
- 12: *Repeat initialization*: start nstart times from step 2 and provide as final output of the algorithm the \mathbf{U} matrix and the medoids $\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_C$ yielding the smallest value of the objective function out of these nstart random initializations.

Algorithm 2 Exponential Kemeny-based FCMd algorithm with entropy regularization

$$d_{exp}(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_c) = \sqrt{1 - \exp \left[-\beta \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right]},$$

we also define the model henceforth called $Exp\text{-FCMdK}_{ent,root}$.

For these entropy-based methods, β is defined as D’Urso et al. (2018):

$$\beta \equiv \beta_2 = \left[\frac{\sum_{l=1}^n d^2(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_q)}{n} \right]^{-1} \tag{7}$$

where $\tilde{\boldsymbol{\pi}}_q : q = \text{argmin}_{1 \leq l \leq n} \sum_{l'=1}^n d^2(\boldsymbol{\pi}_{l'}, \tilde{\boldsymbol{\pi}}_{l'})$.

Remark 1 By replacing $d^2_{exp}(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_c)$ with $d^2(\boldsymbol{\pi}_l, \tilde{\boldsymbol{\pi}}_c)$ in the objective functions (2) and (5), the methods reduce to the FCMd methods based on the squared Kemeny distance, i.e. to their non-robust versions; henceforth, we denote them with $FCMdK$ and $FCMdK_{ent}$, respectively.

3.4 The Choice of Number of Clusters

In general, the partitional clustering algorithms require the researcher to set the number of clusters C to be generated.

In this study, to select the optimal number of clusters C we adopt the *Fuzzy Silhouette index* (Campello and Hruschka, 2006), one of the most known cluster internal validity criteria, the mathematical definition of which is given below.

Consider a data object $j \in \{1, 2, \dots, n\}$ belonging to cluster $p \in \{1, \dots, C\}$. Let a_{pj} be the average distance of object j to all other objects belonging to the same cluster p and let d_{qj} be the average distance of j to all objects belonging to another cluster q , different than p . Denote with b_{pj} the minimum d_{qj} over $q = 1, \dots, C, (q \neq p)$, that is the dissimilarity of object

j with respect to its closest cluster. Campello and Hruschka (2006) defined the silhouette of object j as:

$$s_j = \frac{(b_{pj} - a_{pj})}{\max(b_{pj} - a_{pj})}. \quad (8)$$

The Fuzzy Silhouette (FS) is, then, defined as:

$$FS = \frac{\sum_{j=1}^N (\mu_{pj} - \mu_{qj})^\alpha s_j}{\sum_{j=1}^N (\mu_{pj} - \mu_{qj})^\alpha}. \quad (9)$$

μ_{pj} and μ_{qj} correspond to the first and second largest element of the j -th column of the fuzzy partition matrix, respectively, while $\alpha \geq 0$ is a weighting coefficient.

Therefore, it is a weighted average of the individual silhouettes s_j . A higher value of FS means a better assignment of the units to the clusters which implies that, simultaneously, the intra-cluster distance is minimized while the inter-cluster distance is maximized.

3.5 Advantages of the Proposed Clustering Methods

The proposed clustering approaches inherit all the theoretical advantages related to the fuzzy framework, the Partitioning Around Medoids (PAM) technique, the robustness of clustering procedures and the Kemeny distance as briefly described in the following paragraphs.

3.5.1 Advantages Connected to Fuzzy Clustering

Fuzzy clustering is particularly suitable in many real applications since identifying unambiguous boundaries between clusters is, often, very difficult (McBratney & Moore, 1985; Wedel & Kamakura, 1998). Moreover, compared with hard clustering, the membership degrees matrix arising from fuzzy methods also provides a useful indication of the existence of a second best cluster to consider (Everitt & Landau, 2001). In addition, fuzzy clustering is computationally efficient and works well with distribution-free methods (Hwang et al., 2007).

3.5.2 Advantages Connected to Fuzzy Partitioning Around Medoids (FPAM) Procedure

The FCMd clustering is the extension of the PAM technique in a fuzzy context. Unlike fuzzy C-Means, it is able to identify an observed prototype rather than a virtual one that is particularly suitable in the case of permutations. It also reduces computational complexity compared to the fuzzy C-Means algorithm because the distance matrix is computed once during the iterative clustering procedure (D'Urso et al., 2018).

3.5.3 Advantages Connected to Robustness of Clustering Process

The robustness against outliers is achieved by using the exponential transformation (Wu & Yang, 2002) of the Kemeny distance. It guarantees that clustering procedure is able to neutralize the disruptive effect of the outliers, by recovering the structure of natural groups.

3.5.4 Advantages Connected to Kemeny Distance

The Kemeny distance is one of the most common chosen distances, especially for rankings with ties; it is equivalent to the Kendall distance in the case of complete rankings but differs from the latter in the way of handling tied rankings. It allows great flexibility since tied rankings are widely used in most applications.

4 Simulation Study

4.1 Simulation Plan

To assess the performance of the proposed clustering methods *Exp-FCMdK*, *Exp-FCMdK_{ent}* and *Exp-FCMdK_{ent,root}*, a simulation plan was defined following the scheme below.

4.1.1 Generative Models for Rankings

As the first ranking generator process, we chose the well-known Mallows model (Mallows, 1957), which is an exponential model defined by a central permutation π_0 and a spread parameter θ . For positive values of θ , π_0 is the mode of the distribution, the most likely permutation. The probability associated with any other permutation decreases exponentially as a function of its distance from the central permutation and it is defined as

$$p(\pi) = \frac{\exp(-\theta d(\pi, \pi_0))}{\psi(\theta)},$$

where $\psi(\theta)$ is a normalization constant. When $\theta = 0$, all rankings have a uniform probability to be sampled. The *rmm()* function of the R package *PerMallows* (Irurozki et al., 2016) was used.

As the second ranking generator process, we used the Insertion Sorting Rank (ISR) model for which a ranking is a result of a sorting process of the k objects, based on paired comparisons and proposed by Biernacki and Jacques (2013) according to, given a ranking π , a central ranking π_0 and a parameter $v \in [0.5, 1]$:

$$p(\pi, \pi_0, v) = \frac{1}{k!} \sum_{\sigma \in \mathcal{P}_k} v(\pi | \sigma, \pi_0, v) = \frac{1}{k!} \sum_{\sigma \in \mathcal{P}_k} p^{G(\pi, \sigma, \pi_0)} (1 - v)^{A(\pi, \sigma) - G(\pi, \sigma, \pi_0)},$$

where the sum over $\sigma \in \mathcal{P}_k$ corresponds to all the possible initial presentation orders of the objects to rank while $G(\pi, \sigma, \pi_0)$ is equal to the number of good paired comparisons during the sorting process leading to return π when the presentation order is σ . With $A(\pi, \sigma)$, they denote the total number of paired comparisons. The closer v is to one, the more the distribution of rankings is tightened around π_0 while when $v = 0.5$ a uniform distribution is assumed. The *simulISR()* function of the R package *Rankcluster* (Jacques et al., 2014) was used.

4.1.2 Number of Clusters, Items, and Outliers

For each ranking generative model, we simulated six scenarios with increasing complexity due to the contamination of the natural groups with clustered and radial outliers and then noisy data as summarized in Table 1.

Table 1 The six considered simulated scenarios

Scenario	Type of noise	Objects	Number of noisy/outlying objects
I (Baseline)	None	60	0
II	Clustered outliers	63	3
III	Clustered outliers	66	6
IV	Radial outliers	66	6
V	Radial outliers	72	12
VI	Noisy data	80	20

We considered 2 and 3 groups, equally sized so that, in the former case, the group size is 30 while in the latter 20 and generated by the following central permutations:

- 2 groups: ${}_1\pi_0 = \{1,2,3,4,5\}$ and ${}_2\pi_0 = \{3,1,2,5,4\}$
- 3 groups: ${}_1\pi_0 = \{1,2,3,4,5,6,7\}$, ${}_2\pi_0 = \{1,2,4,7,6,5,3\}$ and ${}_3\pi_0 = \{1,5,2,6,3,4,7\}$.

When the generative model is the Mallows model with Kendall distance, $\theta = 1.5$ while, when it is the ISR model, $\nu = 0.9$. Outliers have been generated by the following central permutations:

- 2 groups: ${}_{ou}\pi_0 = \{5,4,3,2,1\}$
- 3 groups: ${}_{ou}\pi_0 = \{7,6,5,4,3,2,1\}$

but with different spread parameters to simulate the different types of outliers. Therefore, clustered outliers were generated by the above central permutations setting $\theta = 2$ for the Mallows model with two groups and $\theta = 1.5$ with three groups while we set $\nu = 0.9$ for the ISR model for both two and three groups.

To simulate radial outliers, we generated 100 permutations fixing $\theta = 0$ and $\nu = 0.5$ respectively, i.e. we considered a uniform distribution for rankings in the universe of possible permutations, and then we sampled only 12 among them for which the probability of being generated by the centers of the clusters was very low. For noisy data, we considered the first 20 permutations simulated by the same uniform distribution as above.

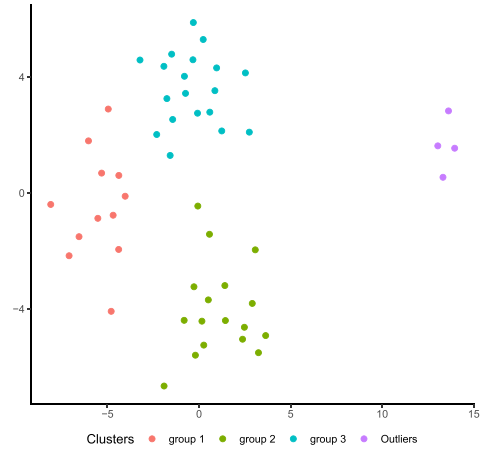
In Fig. 2, examples of clustered outliers with three well-separated natural groups, radial outliers with two less-separated natural groups and noisy data with three well-separated natural groups are shown using the Sammon projection in a two dimensional space.

For each combination of generative model and number of groups, we simulated 30 artificial dataset leading to 120 baseline datasets but considering all six scenarios for which a certain amount of outliers was added, the number of different datasets evaluated increases to 720. The simulated natural groups must be the same for comparative purposes in terms of robustness against outliers and noisy data.

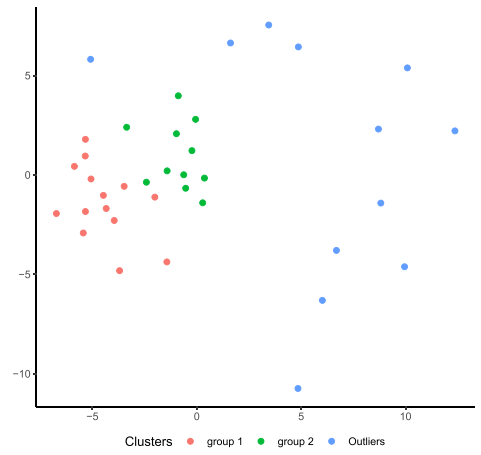
We argue that all modal permutations used to generate outliers were chosen so that their Kemeny distance from each modal permutation has been equal or close to the maximum value of the Kemeny distance (e.g. for two permutations of length 5 the maximum distance is equal to 20 while it is 42 for two permutations of length 7). Moreover, we simulated only full rankings since the R functions used to simulate rankings do not allow to generate from central permutations with ties.

The proposed robust methods were compared with the corresponding non robust baseline methods $FCMdk$ and $FCMdk_{ent}$ and the mixture model-based approach proposed by

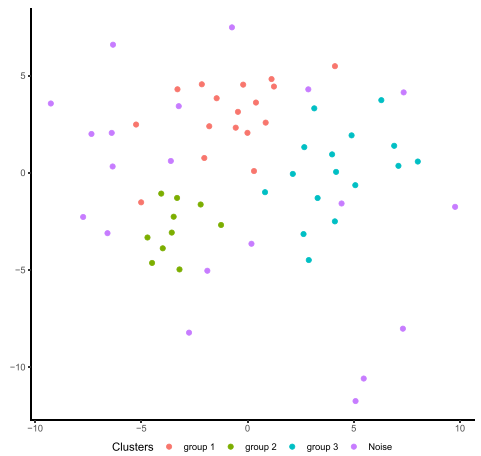
Fig. 2 An example of clustered and radial outliers and noisy data



(a) Clustered outliers with three natural groups



(b) Radial outliers with two natural groups



(c) Noisy data with three natural groups

Jacques and Biernacki (2014) which is the extension of the Insertion Sorting Rank (ISR) model. The heterogeneity of the rank population is modelled by a mixture of ISR. Maximum likelihood estimation is performed through a SEM-Gibbs algorithm, allowing partial rankings too. For the Mixture ISR model (henceforth, *M-ISR*), the membership degree matrix is represented by the conditional probabilities estimates for the observed ranks to belong to each cluster. The model is implemented in the R package `Rankcluster` (Jacques et al., 2014).

The additional model used for comparison purposes is the K-Median Cluster Component Analysis (henceforth, *CCA*) proposed by D’Ambrosio and Heiser (2019), the most similar to our models in that it is a probabilistic-distance clustering model based on the same Kemeny distance. As pointed out by D’Ambrosio and Heiser (2019), the method allows each ranking to be assigned to all C clusters by a membership probability, thus mimicking a fuzzy clustering with exponent $m = 2$. The model is implemented in the R package `ConsRankClass` (D’Ambrosio, 2021).

For each simulated setup, the 30 fuzzy partitions related to the 60 permutations (only those belonging to the natural groups excluding the outliers and the noisy data) are compared with the reference crisp partition using the Adjusted Concordance Index (ACI) proposed by D’Ambrosio et al. (2021); this is an external validation criterion that corrects the normalized degree of concordance (NDC) introduced by Hullermeier et al. (2011) for the agreement that may be due to chance. Lying in the range $[-1, 1]$, it is equal to 1 in the case of perfect correspondence between the two partitions; thus, the higher the value, the better the agreement between the two partitions.

4.2 Simulation Results

In this section, simulation results are provided for each scenario of interest. We set C according to the number of simulated groups in each configuration; then, we applied the clustering methods varying $m \in \{1.3, 1.5, 2\}$ for the *FCMdK* and *Exp-FCMdK* methods while $p \in \{0.05, 0.1, 0.2\}$ for the *Exp-FCMdK_{ent}* and *Exp-FCMdK_{ent,root}* ones. For the *FCMdK_{ent}* we considered $p \in \{0.01, 0.02, 0.04\}$ ³.

We ran all algorithms considering 100 random restarts with an execution time of a few seconds. The associated code has been implemented in R.

For the *CCA* method, we used the branch-and-bound algorithm to find the median ranking fixing the number of replications to 10 (as suggested by the authors) while for the *M-ISR* mixture model, we set the `maxTry` to 10 and the `run` to 2. Simulation results for both ranking generative models and all six scenarios are shown in Figs. C.1–C.8, in Appendix C, in which we plotted the violin plots of the Adjusted Concordance index for each method. The subscript in the label provides the value of the corresponding m and p parameters depending on the model specification.

As a first piece of evidence, simulation results are very promising and confirm what is already known in the literature: the FCMd method is only a “timid robustification” of clustering against outliers (García-Escudero and Gordaliza, 2005). It is worth specifying that the methods are not directly comparable in terms of performance as fuzziness increases since there is no direct correspondence between the value of the “ p ” parameter in the entropy-based method and that of the “ m ” exponent in the other fuzzy-type method.

³ The value of p depends on the scaling of the $d(\cdot)$ matrix so it must be set accordingly. Therefore, we used the normalized Kemeny distance so that the distance lies in $[0, 1]$ as for the exponential transformation.

Direct comparability in terms of performance as the fuzziness increases is possible within the methods based on “ m ” and between them and the CCA model as well as between the 2 robust entropy-based methods.

In general, as expected, the performances of all methods, in terms of ACI values, tend to decrease as the fuzziness increases because we compared a crisp partition with a fuzzy one.

We argue that the choice of the best value of the fuzziness parameter strictly depends on the scaling of the distance used as well as on the degree of separation among groups; thus, in the practical applications, we recommend setting it taking into account all these issues.

From the simulation results based on the ISR scheme for 2 and 3 groups, one can arise the following considerations. If it was reasonable to expect that, assuming the ISR generative scheme, the mixture ISR clustering model based on the same generative process would perform well, even in the presence of outliers, we cannot fail to emphasize the good performance of our robust clustering methods.

In particular, the $Exp-FCMdK_{ent_{root}}$ entropy-based robust method has the best performance; in the case of two simulated natural groups (see Figs. C.1 and C.2), its good behaviour, when $p \leq 0.10$ becomes evident especially when the amount of radial outliers is consistent or noisy data are present. Compared with the $Exp-FCMdK_{ent}$ robust method, the $Exp-FCMdK_{ent_{root}}$ is indeed more stable especially when radial outliers contaminates the natural clusters showing a ACI distribution that is tighter.

As far as the M -ISR model is concerned, it seems however less robust against radial outliers and noisy permutations when these increase in number (see Fig. C.2).

The contribution of our clustering methods is particularly evident when we go to compare their performance with that of the CCA model. Even when we consider the $Exp-FCMdK$ with $m = 2$, our fuzzy method always outperforms CCA. Looking at the results in more detail, since CCA is a median-based model, it seems, on the one hand, to be able to identify outliers by assigning them equal membership degrees ($1/C$), like the $Exp-FCMdK$ methods, but on the other hand, it produces, in general, too blurred partitions. This leads to a low performance in terms of values of the ACI, especially when outliers and noisy data occur. Then, it is fairly evident how the possibility to tune the level of fuzziness adds considerable flexibility to this class of methods.

When we consider the same sampling scheme but with three natural groups (see Figs. C.3 and C.4), the $M - ISR$ performs quite well again but in the case of clustered outliers.

We argue that, however, looking at its violin plots, its levels of performance decrease in general as the amount of outliers or noisy data increases.

Regarding our robust methods, the best performance is associated with the $Exp-FCMdK_{ent_{root}}$ if we compare the ACI distributions when a large number of radial outliers or noisy data are present (and p is low).

Looking at the violin plots of the CCA method, although they have more or less the same shape regardless of the type of contamination, they show very low performance due to the tendency of the method to provide partitions that are too blurred.

When a Mallows scheme and two groups are considered (see Figs. C.5 and C.6), the $Exp-FCMdK_{ent_{root}}$ outperforms the other methods especially if we look at the ACI distribution in the case of radial outliers and noisy data. We point out that a good performance is also associated with the $Exp-FCMdK$ method.

If we consider the Mallows scheme with three groups (see Figs. C.7 and C.8), the very good performances of our robust methods (except for high values of the p and m parameters) are more evident. Moreover, we notice that, in this simulated setup, the M -ISR model has lower performances, followed by CCA.

Summing up, this simulation plan proved the excellent behaviour of the $Exp-FCMdK_{ent,root}$ method in particular, which becomes evident especially in the case of radial outliers and noisy data. It is less sensitive to the value of the fuzziness parameter showing higher values of the Adjusted Concordance index as well as tightened distributions of the same index.

The $M-ISR$ is, however, a good competitor while the CCA method produces too blurred partitions thus drastically reducing its performance especially when the Mallows sample scheme is assumed.

5 Application to Real Data

In this section, we applied the above methods to two real datasets: the *Gaming Platforms* dataset due to Fok et al. (2012) and available in the R package `PLMIX` (Mollica and Tardella, 2017) and the *University rankings* dataset, including tied rankings and available in the R package `ConsRankClass` (D'Ambrosio, 2021).

5.1 The Gaming Platforms Dataset

The *Gaming Platforms* dataset contains the results of a survey conducted on a sample of 91 Dutch students who rank, from the most-liked (Rank 1) to the least-liked (Rank 6), 6 gaming platforms : 1 = X-Box, 2 = PlayStation, 3 = PSPortable, 4 = GameCube, 5 = GameBoy and 6 = Personal Computer (PC). The best solution for each robust method has been chosen based on the combination of C and the fuzziness parameter (m or p depending on the method) that maximizes the Fuzzy Silhouette index⁴.

In particular, we varied $C \in \{2, 3, 4, 5, 6, 7, 8\}$, $m \in \{1.3, 1.5, 1.7, 2\}$ and $p \in \{0.05, 0.10, 0.15\}$

For the $M-ISR$ model, the best solution has been chosen according to the BIC criterion while for the CCA model according to the Fuzzy Silhouette.

All clustering methods identified two groups but for the $Exp-FCMdK_{ent,root}$ and CCA which identified three groups. The centers are shown in Table 2 while the membership degrees matrices are shown in Table D.4.

As first evidence, it can be seen that, within each method, the centers of the groups differ mainly in the podium, assigned to the Personal Computer or the PlayStation depending on the case; the X-Box is almost always ranked as the second best gaming platform while the GameBoy and GameCube are the least preferred ones.

In detail, the $Exp-FCMdK$ and $Exp-FCMdK_{ent}$ methods differ only in the second medoid: their Kemeny distance is equal to 2, so they are very similar; indeed the preferences are the same except for the inversion of the ranking between the PSPortable and the Personal Computer. We also argue that the medoid 2 of the $Exp-FCMdK$ is also equal both to the medoid 3 of the $Exp-FCMdK_{ent,root}$ and the third median ranking of the CCA . The $M-ISR$ model has in common with the latter the second center while the center 1 identifies the group of subjects for which the best platform is the X-Box followed by the PlayStation and Personal Computer.

⁴ We point that the best method is also chosen according to the trade-off between maximizing the FS index and avoiding partitions that are too blurred.

Table 2 Gaming Platforms dataset: the centers associated with each clustering method

Ranking	$Exp-FCMdk_{ent}$			$M-ISR$			$Exp-FCMdk_{ent,not}$			CCA		
	Medoid 1	Medoid 2	Medoid 1	Medoid 2	Center 1	Center 2	Medoid 1	Medoid 2	Medoid 3	Medoid 1	Medoid 2	Medoid 3
1	PC	PlayStation	PC	PlayStation	X-Box	PC	PC	PC	PlayStation	GameCube	PC	PlayStation
2	X-Box	X-Box	X-Box	X-Box	PlayStation	PlayStation	X-Box	PlayStation	X-Box	X-Box	PlayStation	X-Box
3	PlayStation	PSPortable	PlayStation	PC	PC	X-Box	PlayStation	X-Box	PC	PSPortable	X-Box	PC
4	PSPortable	PC	PSPortable	PSPortable	PSPortable	PSPortable	GameCube	PSPortable	PSPortable	GameBoy	PSPortable	PSPortable
5	GameCube	GameCube	GameCube	GameCube	GameCube	GameBoy	PSPortable	GameBoy	GameCube	PlayStation	GameBoy	GameCube
6	GameBoy	GameBoy	GameBoy	GameBoy	GameBoy	GameCube	GameBoy	GameCube	GameBoy	PC	GameCube	GameBoy

The main interesting evidence is that the $Exp-FCMdk_{ent_{root}}$ and the CCA share the second and third center, but differ in the first one. Looking at the medoid 1 identified by the $Exp-FCMdk_{ent_{root}}$ algorithm, it can be seen that the preferred platform is again the Personal Computer, but with a rather different ranking of the other gaming platforms. The CCA model does not identify this group, while it identifies the group of subjects whose median ranking has an opposite liking to that of the other two groups, assigning to the PlayStation and Personal Computer the last positions. We argue that students belonging to this small group are the same as those to which the $Exp-FCMdk_{ent_{root}}$ method assigns equal membership degrees in all three clusters, thus those identified by the robust method as outliers (see Table D.4). The presence of the third natural group detected by the $Exp-FCMdk_{ent_{root}}$ but not by the others could be due to the ability of this method to neutralize the disruptive effect of outliers and discover an additional cluster.

The $M-ISR$ model seems to be the most affected by the presence of noisy data since all scattered permutations have been included in one group.

5.2 The University Rankings Dataset

The dataset on university rankings concerns a survey of 303 students attending the Vienna University of Economics. They were asked to indicate preferences for the following 6 universities: London, Paris, Milan, St. Gallen, Barcelona, and Stockholm. This dataset is interesting because it contains tied rankings. We applied the $Exp-FCMdk_{ent_{root}}$ robust method and, for comparison purposes, the CCA method based on the same distance. The Fuzzy Silhouette index suggested 2 groups for both methods and, as far as the $Exp-FCMdk_{ent_{root}}$ is concerned, we chose the value of p equal to 0.10. Table 3 shows the two centers for each method while their membership degrees matrices are reported in Table D.5. Looking at the centers, one can notice that the two clustering methods lead to the same centers but for the first one which only differs in assigning, in the case of the $Exp-FCMdk_{ent_{root}}$, rank 2 to Paris and rank 3 to St. Gallen while the CCA model assigns to them the same second best position.

Looking at the membership degree matrices in Table D.5, as expected by simulations, the main difference between the two methods is that the CCA method provides a partition fuzzier than the $Exp-FCMdk_{ent_{root}}$. Based on the latter method and the crisp assignment of the units to the clusters using a cut-off value of 0.7, the number of students per group is 71 and 78 respectively, while the fuzzy units are 63⁵.

The medoids associated with the $Exp-FCMdk_{ent_{root}}$ are representative of one group for which, after London and Paris, the best choice is St. Gallen followed by the remaining ones in last position and another group according to which, keeping the first two places unchanged, the third best study location is Barcelona and then Milan while St. Gallen is ranked last here together with Stockholm.

This dataset contains eight background binary covariates concerning the following aspects: **Stud** (main discipline of study with categories *commerce-other*), **Eng** (knowledge of English with categories *good-poor*), **French** (knowledge of French with categories *good-poor*), **Spanish** (knowledge of Spanish with categories *good-poor*), **Italian** (knowledge of

⁵ We argue that the number of fuzzy units can be reduced by setting $p = 0.05$ obtaining the same partition but with a reduced number of fuzzy units.

Table 3 University Rankings dataset: the centers associated with the $Exp-FCMdk_{ent_{row}}$ and the CCA clustering methods

	1	2	3	4	5
$Exp-FCMdk_{ent_{row}}$	Medoid 1	London	Paris	St. Gallen	Milan, Barcelona, Stockholm
	Medoid 2	London	Paris	Barcelona	Milan
CCA	Median ranking 1	London	Paris, St. Gallen	Milan, Barcelona, Stockholm	St. Gallen, Stockholm
	Median Ranking 2	London	Paris	Barcelona	Milan
					St. Gallen, Stockholm

Italian with categories *good-poor*), **Work** (full-time employment while studying with categories *no-yes*), **Degree** (intention to take an international degree with categories *no-yes*) and **Gender** (sex with categories *female-male*).

Figure 3 shows the relative frequency distribution of these variables in the groups (considering only the non-fuzzy units) compared with the same frequency distribution in the whole sample of students. The group of those preferring Barcelona to St. Gallen is characterized by a percentage of students studying *commerce* lower than that of the second group and the whole sample but with a better knowledge of Spanish and Italian. This group is also characterized by a higher share of worker-students and a lower percentage of students who are intentioned to take an international degree. As far as the gender is concerned, the first group seems characterized by a slightly higher percentage of female students.

Summing up, the language knowledge and the international profile of the degree are the main drivers conditioning the preference among St. Gallen and Barcelona (and Milan).

6 Concluding Remarks

This work focuses on the definition of two robust fuzzy clustering techniques with the two-fold aim of detecting homogeneous groups of judges, according to their preferences on a set of items, and neutralizing the effects of possible outliers or noisy data during the clustering procedure. The use of the FCMd method according to the two different approaches to deal with fuzziness and based on the exponential transformation of the Kemeny distance accomplishes all these goals simultaneously.

Moreover, the use of the Kemeny distance has the added value of making the method more flexible, extending its field of application thanks to its ability to handle tied rankings too.

Simulation results are very promising and show the good performance of our proposals, especially that of the $Exp-FCMdK_{ent_{root}}$, which is robust against the most insidious type of outliers, the radial ones, and against noisy permutations that often occur when handling real data. It seems less sensitive to the value of the fuzziness parameter showing higher values of the Adjusted Concordance index as well as a more tightened distribution of the same index.

Our proposals could be seen as valid alternatives to the *M-ISR* model and, in particular, to the *CCA* one. As already said, our models have the further advantage of tuning the fuzziness parameter allowing us to obtain the right centers of the clusters and at the same time a much less fuzzy partition, ensuring robustness against noisy data and anomalous permutations.

Furthermore, we also showed that our model-free approach works well under different ranking generative schemes.

As a further development of this work, to increase the flexibility of this class of methods, we will focus on defining a weighted distance-based fuzzy clustering method that allows to use different weights for different ranks, believing that the weights can significantly improve the performance of these methods (Lee & Yu, 2012).

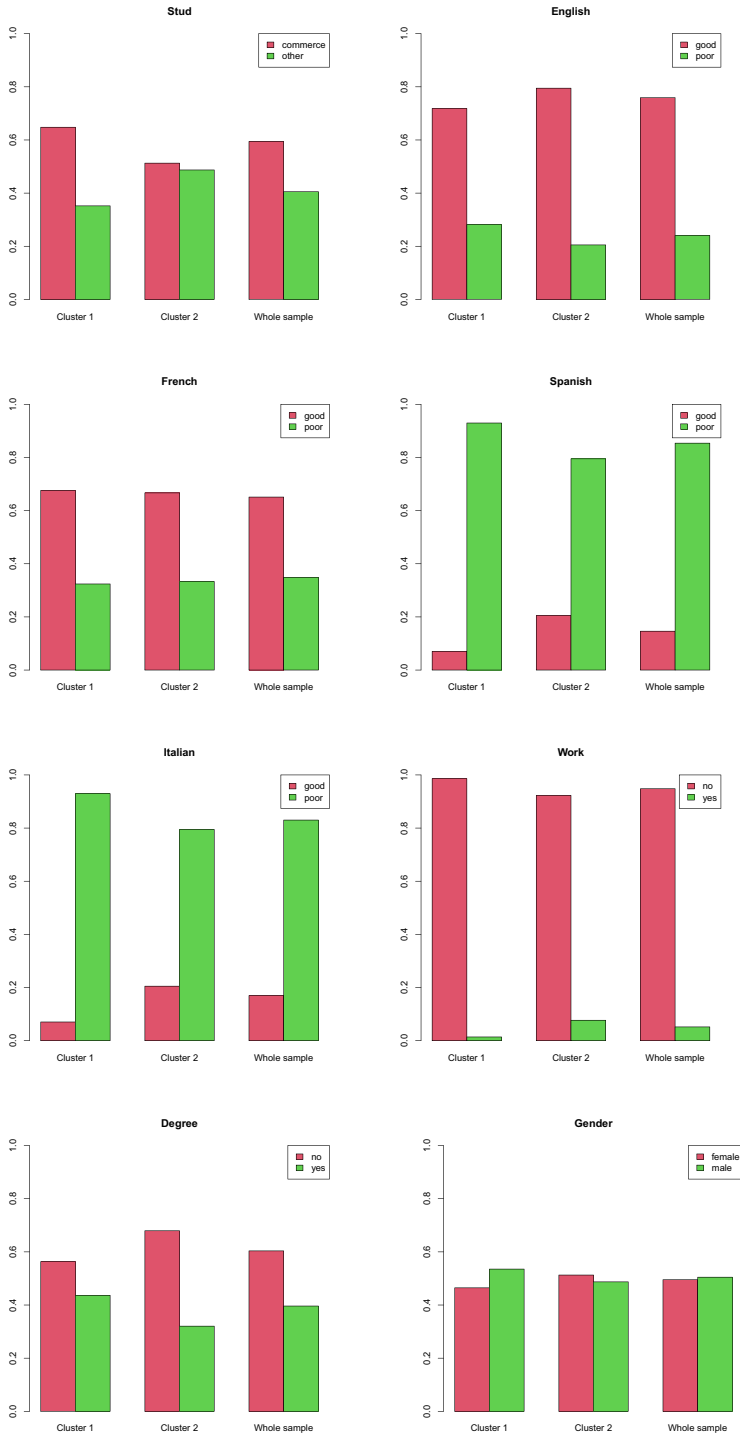


Fig. 3 Gaming Platforms dataset: the associated covariates

Appendix A

Proof We consider the Lagrangian function:

$$L_m(\mathbf{u}_l, \lambda) = \sum_{l=1}^n \sum_{c=1}^C u_{lc}^m \cdot \left\{ 1 - \exp \left[-\beta_1 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} - \lambda \left(\sum_{c=1}^C u_{lc} - 1 \right) \tag{10}$$

where $\mathbf{u}_l = (u_{l1}, \dots, u_{lc}, \dots, u_{lC})'$ and λ is the Lagrange multiplier. Therefore, we set the first derivatives of (10) with respect to u_{lc} and λ equal to zero, yielding:

$$\frac{\partial L_m(\mathbf{u}_l, \lambda)}{\partial u_{lc}} = 0 \Leftrightarrow m \cdot u_{lc}^{m-1} \left\{ 1 - \exp \left[-\beta_1 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} - \lambda = 0 \tag{11}$$

$$\frac{\partial L_m(\mathbf{u}_l, \lambda)}{\partial \lambda} = 0 \Leftrightarrow \sum_{c=1}^C u_{lc} - 1 = 0 \tag{12}$$

From (11), we obtain:

$$u_{lc} = \left(\frac{\lambda}{m \left\{ 1 - \exp \left[-\beta_1 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\}} \right)^{\frac{1}{m-1}} \tag{13}$$

and, by considering (12):

$$\frac{\lambda}{m} = \left(\frac{1}{\sum_{c=1}^C \left(\frac{1}{\left\{ 1 - \exp \left[-\beta_1 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\}} \right)} \right)^{\frac{1}{m-1}} \tag{14}$$

Finally, substituting (14) in (13), we obtain u_{lc} as

$$u_{lc} = \frac{1}{\sum_{c'=1}^C \left[\frac{\left\{ 1 - \exp \left[-\beta_1 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\}^{\frac{1}{m-1}}}{\left\{ 1 - \exp \left[-\beta_1 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_{c'}(i,j)| \right)^2 \right] \right\}^{\frac{1}{m-1}}} \right]} \tag{15}$$

□

Appendix B

Proof We consider the Lagrangian function:

$$L_m(\mathbf{u}_l, \lambda) = \sum_{l=1}^n \sum_{c=1}^C u_{lc} \left\{ 1 - \exp \left[-\beta_2 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} + p \sum_{l=1}^n \sum_{c=1}^C u_{lc} \log(u_{lc}) - \lambda \left(\sum_{c=1}^C u_{lc} - 1 \right) \tag{16}$$

where $\mathbf{u}_l = (u_{l1}, \dots, u_{lc}, \dots, u_{lC})'$ and λ is the Lagrange multiplier. Therefore, we set the first derivatives of (16) with respect to u_{lc} and λ equal to zero, yielding:

$$\frac{\partial L_m(\mathbf{u}_l, \lambda)}{\partial u_{lc}} = 0 \Leftrightarrow \left\{ 1 - \exp \left[-\beta_2 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} + p [\log(u_{lc}) + 1] - \lambda = 0 \tag{17}$$

$$\frac{\partial L_m(\mathbf{u}_l, \lambda)}{\partial \lambda} = 0 \Leftrightarrow \sum_{c=1}^C u_{lc} - 1 = 0 \tag{18}$$

From (17), we obtain:

$$\log(u_{lc}) = \frac{1}{p} \left\{ \lambda \left[1 - \exp \left[-\beta_2 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right] \right\} - 1 \tag{19}$$

and, then:

$$u_{lc} = \exp \left\{ \frac{\lambda}{p} - \frac{1}{p} \left\{ 1 - \exp \left[-\beta_2 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} \right\} - 1 \tag{20}$$

By considering (18):

$$\exp \left(\frac{\lambda}{p} - 1 \right) = \frac{1}{\sum_{c=1}^C \left[\frac{1}{\left\{ 1 - \exp \left[-\beta_2 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\}} \right]} \tag{21}$$

and by replacing Equation (21) in (20), we have:

$$u_{lc} = \frac{1}{\sum_{c'=1}^C \left[\frac{\exp \left(\frac{1}{p} \left\{ 1 - \exp \left[-\beta_2 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_c(i,j)| \right)^2 \right] \right\} \right)}{\exp \left(\frac{1}{p} \left\{ 1 - \exp \left[-\beta_2 \left(\frac{1}{2} \sum_{i,j=1}^k |\pi_l(i,j) - \tilde{\pi}_{c'}(i,j)| \right)^2 \right] \right\} \right)} \right]} \tag{22}$$

□

Appendix C

Figures C.1, C.2, C.3, C.4, C.5, C.6, C.7, C.8.

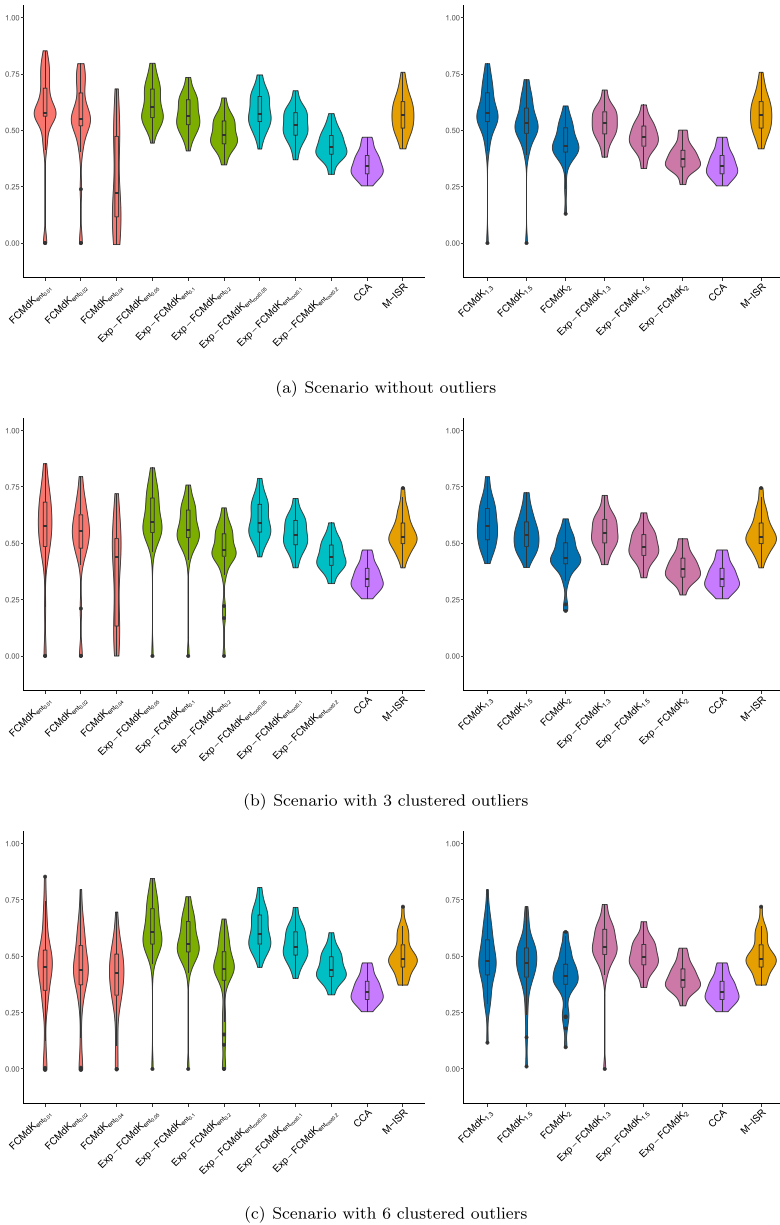
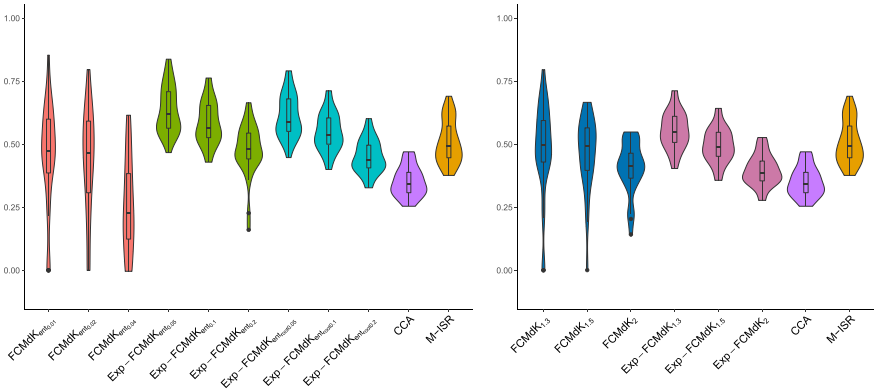
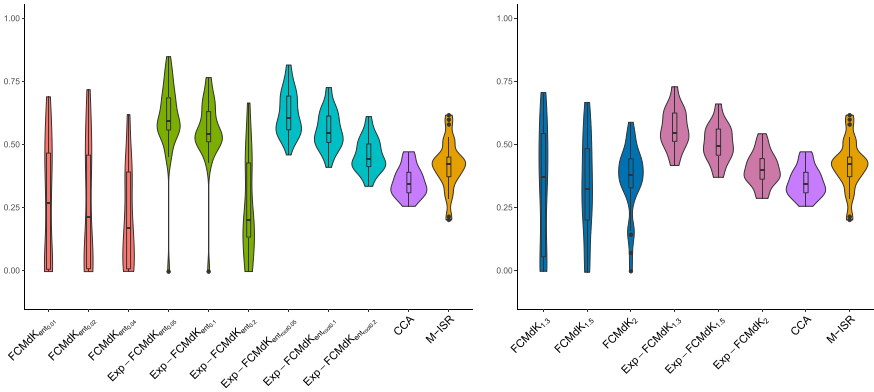


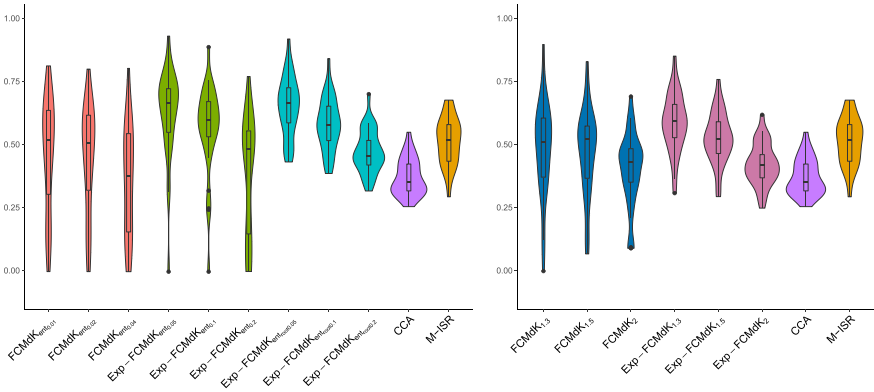
Fig. C.1 ISR generative model with two groups: boxplots of ACI indexes for scenarios (I) 2 natural groups and no outliers, (II) 2 natural groups and 3 clustered outliers and (III) 2 natural groups and 6 clustered outliers. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{ent,root} and FCMdK_{ent} methods



(a) Scenario with 6 radial outliers

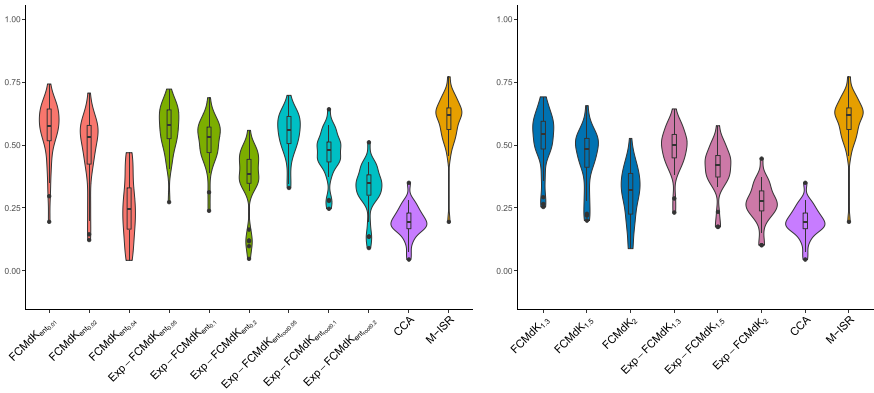


(b) Scenario with 12 radial outliers

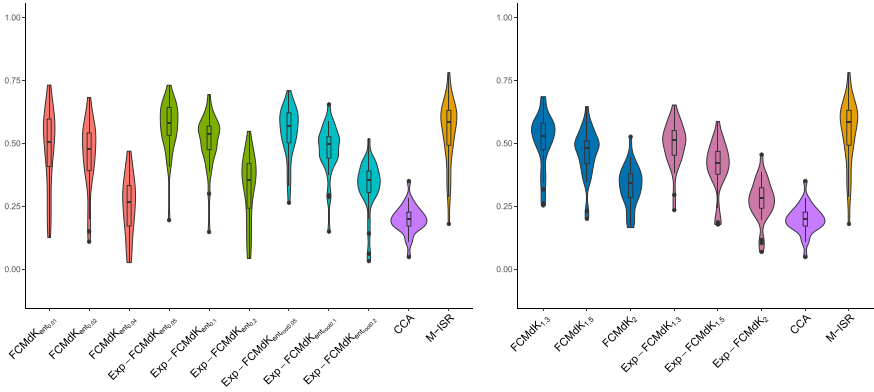


(c) Scenario with 20 noisy data

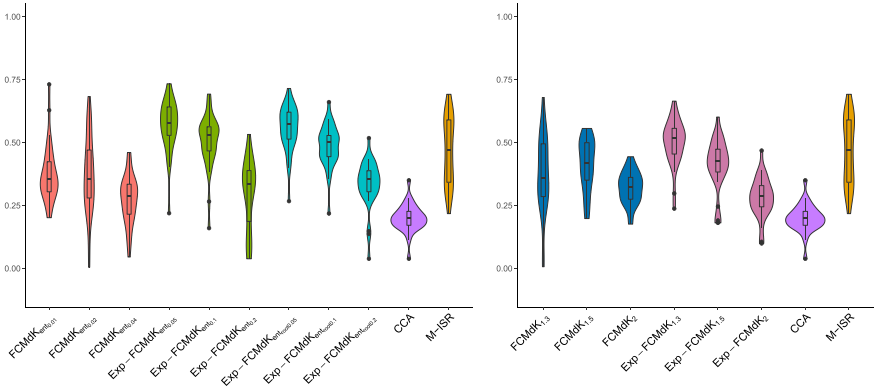
Fig. C.2 ISR generative model with two groups: boxplots of ACI indexes for scenarios (IV) 2 natural groups and 6 radial outliers, (V) 2 natural groups and 12 radial outliers and (VI) 2 natural groups and 20 noisy data. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{ent,root} and FCMdK_{ent} methods



(a) Scenario without outliers

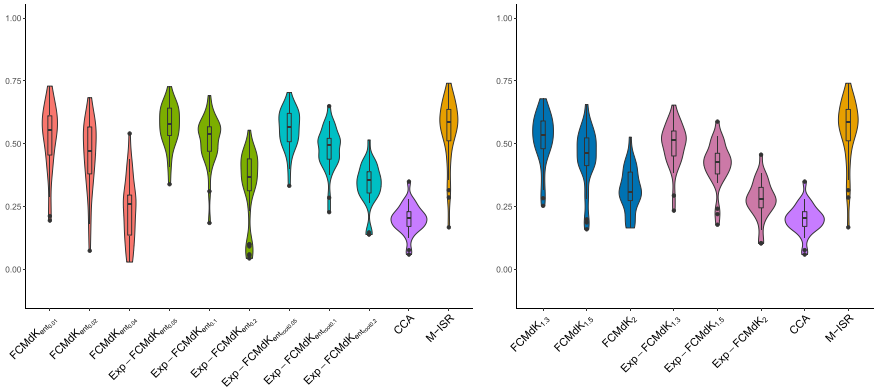


(b) Scenario with 3 clustered outliers

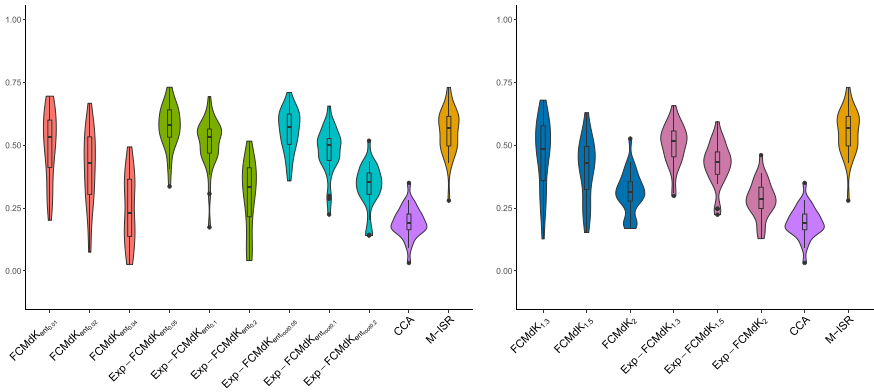


(c) Scenario with 6 clustered outliers

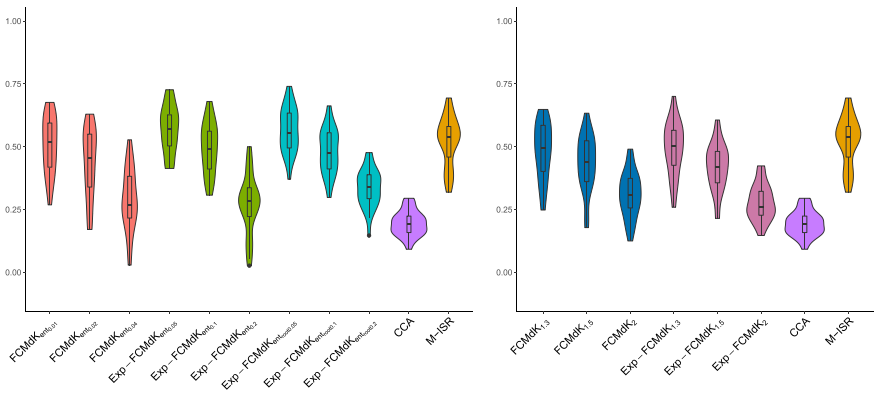
Fig. C.3 ISR generative model with three groups: boxplots of ACI indexes for scenarios (I) 3 natural groups and no outliers, (II) 3 natural groups and 3 clustered outliers and (III) 3 natural groups and 6 clustered outliers. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{ent,root} and FCMdK_{ent} methods



(a) Scenario with 6 radial outliers

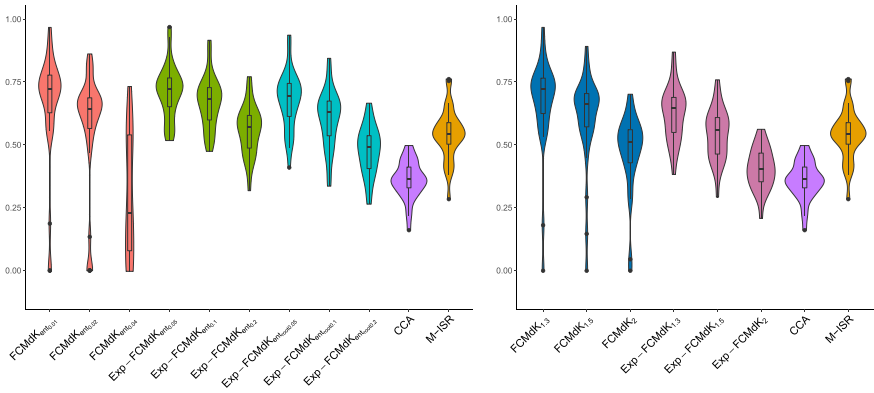


(b) Scenario with 12 radial outliers

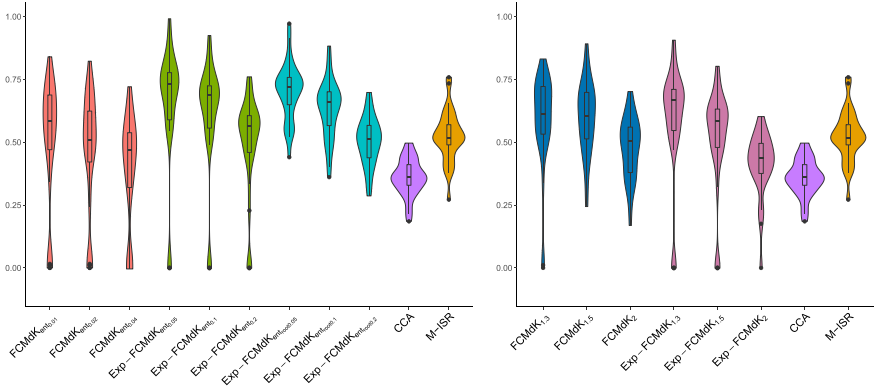


(c) Scenario with 20 noisy data

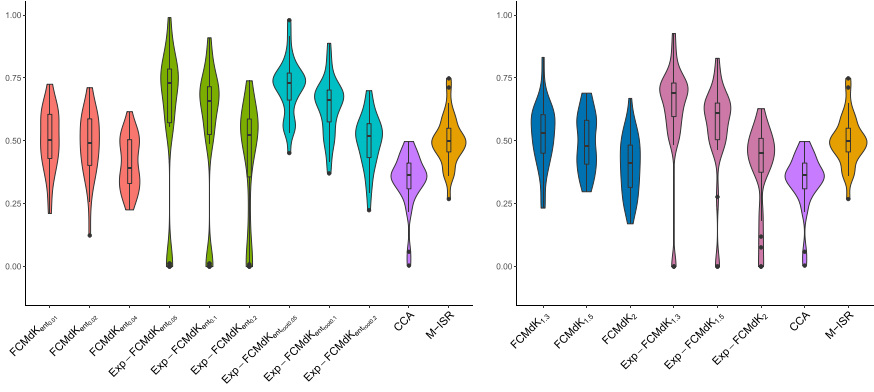
Fig. C.4 ISR generative model with three groups: boxplots of ACI indexes for scenarios (IV) 3 natural groups and 6 radial outliers, (V) 3 natural groups and 12 radial outliers and (VI) 3 natural groups and 20 noisy data. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{ent,root} and FCMdK_{ent} methods



(a) Scenario without outliers

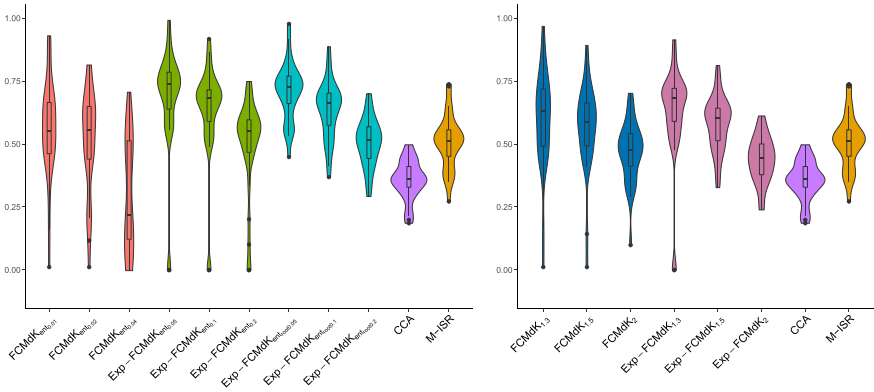


(b) Scenario with 3 clustered outliers

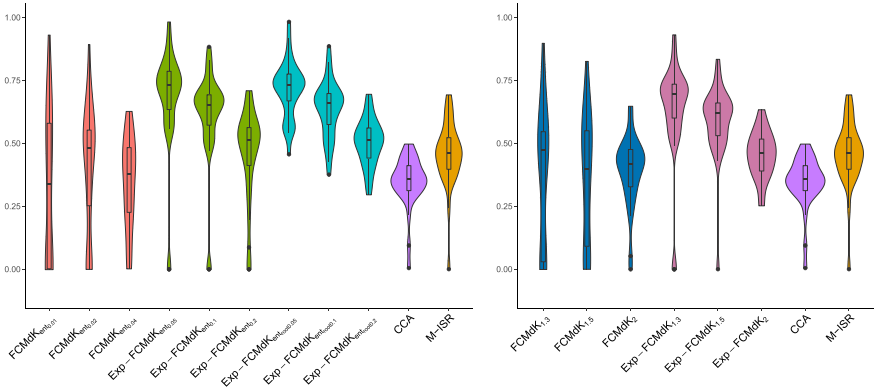


(c) Scenario with 6 clustered outliers

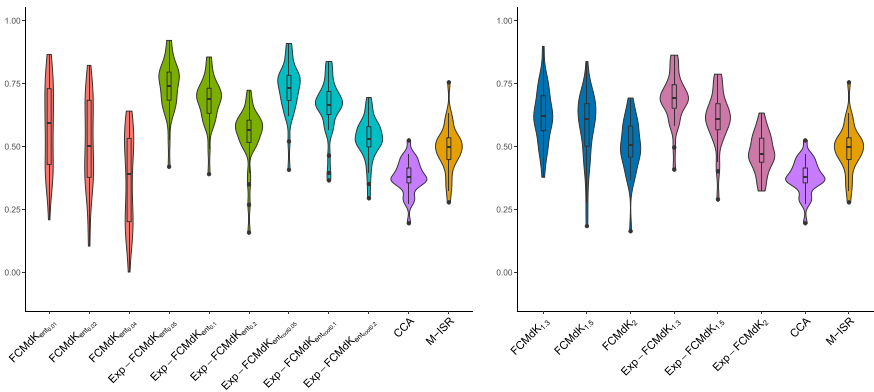
Fig. C.5 Mallows generative model with two groups: boxplots of ACI indexes for scenarios (I) 2 natural groups and no outliers, (II) 2 natural groups and 3 clustered outliers and (III) 2 natural groups and 6 clustered outliers. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{ent,root} and FCMdK_{ent} methods



(a) Scenario with 6 radial outliers

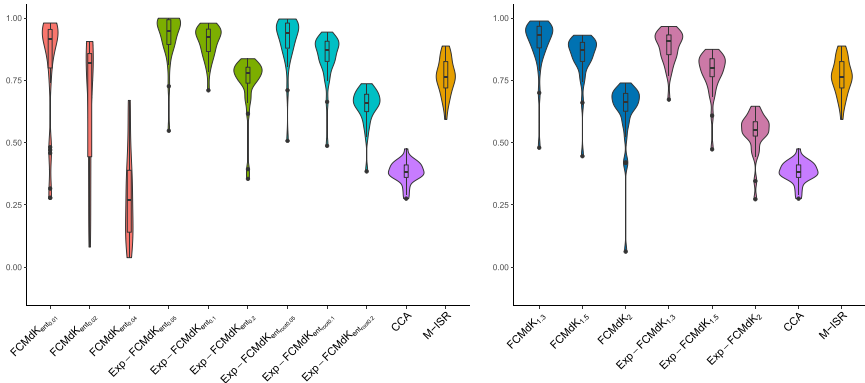


(b) Scenario with 12 radial outliers

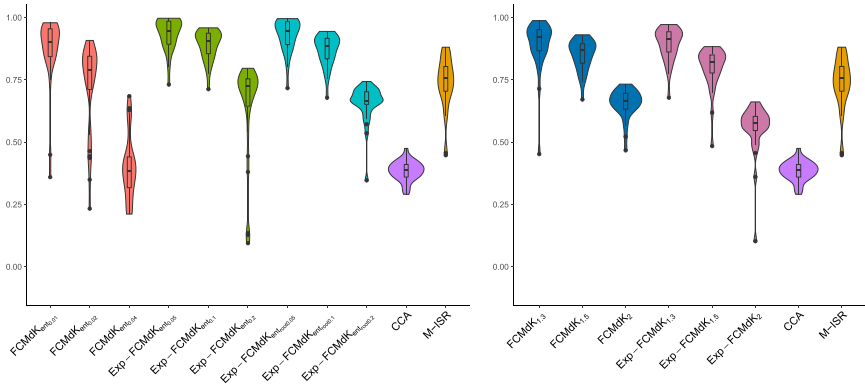


(c) Scenario with 20 noisy data

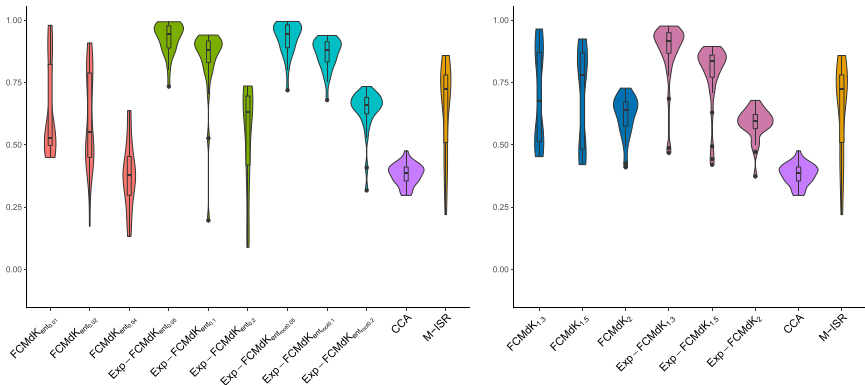
Fig. C.6 Mallows generative model with two groups; boxplots of ACI indexes for scenarios (IV) 2 natural groups and 6 radial outliers, (V) 2 natural groups and 12 radial outliers and (VI) 2 natural groups and 20 noisy data. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{ent,root} and FCMdK_{ent} methods



(a) Scenario without outliers

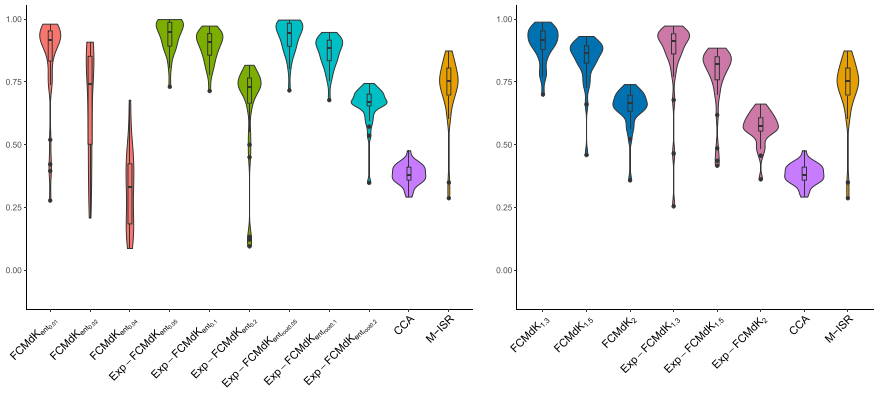


(b) Scenario with 3 clustered outliers

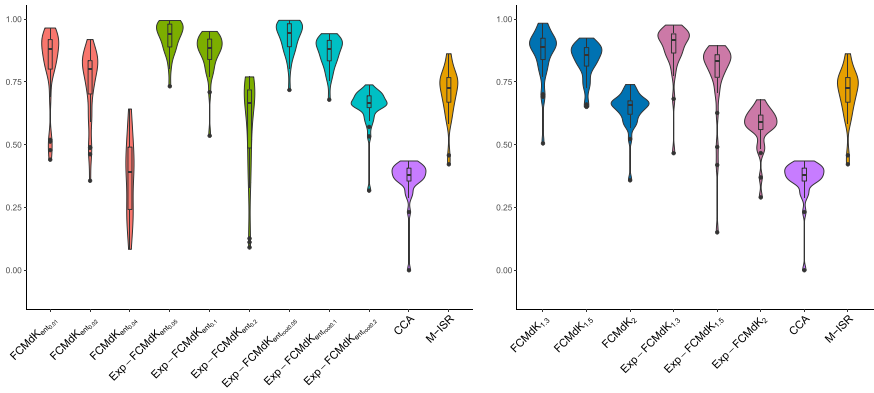


(c) Scenario with 6 clustered outliers

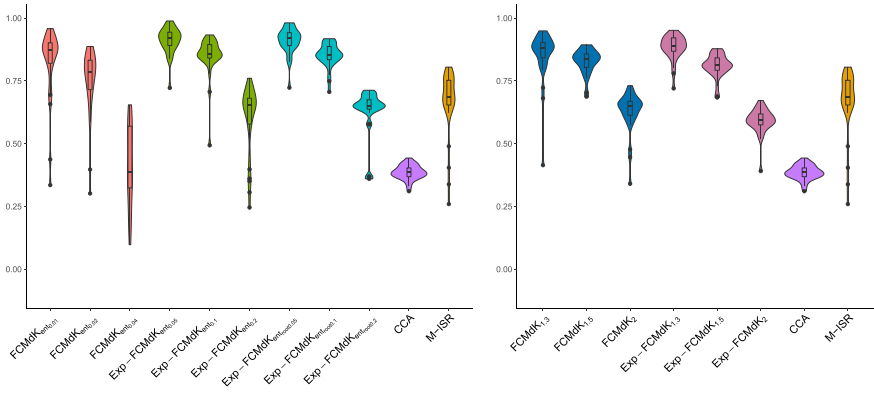
Fig. C.7 Mallows generative model with three groups: boxplots of ACI indexes for scenarios (I) 3 natural groups and no outliers, (II) 3 natural groups and 3 clustered outliers and (III) 3 natural groups and 6 clustered outliers. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{entroot} and FCMdK_{ent} methods



(a) Scenario with 6 radial outliers



(b) Scenario with 12 radial outliers



(c) Scenario with 20 noisy data

Fig. C.8 Mallows generative model with three groups: boxplots of ACI indexes for scenarios (IV) 3 natural groups and 6 radial outliers, (V) 3 natural groups and 12 radial outliers and (VI) 3 natural groups and 20 noisy data. The subscript in the label provides the value of the corresponding m parameter for the FCMdK and Exp-FCMdK methods and of the p parameter for the Exp-FCMdK_{ent}, Exp-FCMdK_{ent,root} and FCMdK_{ent} methods

Appendix D

Tables D.4, D.5.

Table D.4 *Gaming Platforms* dataset: membership degrees matrix provided by the applied clustering methods

Row	X-box	Play-Station	PSPort-able	Game-Cube	Game-Boy	Per-sonal Com-puter	Exp-FCMdK _{ent} ($p=0.05$)			Exp-FCMdK ($m=1.3$)			M-ISR			Exp-FCMdK _{ent,ent} ($p=0.10$)			CCA		
							Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
							1	2	1	3	5	6	4	0.00	1.00	0.00	0.00	1.00	0.10	0.90	0.00
2	4	2	3	5	6	1	1.00	0.00	0.92	0.08	0.44	0.56	0.44	0.15	0.70	0.15	0.09	0.55	0.36		
3	3	1	2	4	6	5	0.00	1.00	0.03	0.97	0.06	0.94	0.06	0.02	0.04	0.94	0.20	0.27	0.53		
4	3	5	6	1	4	2	1.00	0.00	0.78	0.22	0.17	0.83	0.17	0.79	0.12	0.09	0.41	0.31	0.28		
5	1	2	3	4	6	5	0.00	1.00	0.16	0.84	0.02	0.98	0.02	0.08	0.03	0.89	0.25	0.25	0.50		
6	3	2	5	4	6	1	1.00	0.00	0.92	0.08	0.84	0.16	0.16	0.82	0.15	0.03	0.11	0.54	0.36		
7	3	2	4	6	5	1	1.00	0.00	0.92	0.08	0.99	0.01	0.99	0.00	1.00	0.00	0.00	1.00	0.00		
8	2	3	4	1	5	6	0.02	0.98	0.40	0.60	0.01	0.99	0.01	0.41	0.17	0.41	0.67	0.13	0.19		
9	2	3	1	5	6	4	0.00	1.00	0.24	0.76	0.00	1.00	0.00	0.12	0.06	0.82	0.27	0.27	0.47		
10	2	1	4	5	6	3	0.01	0.99	0.00	1.00	0.08	0.92	0.08	0.00	0.00	1.00	0.00	0.00	1.00		
11	2	1	4	5	6	3	0.01	0.99	0.00	1.00	0.07	0.93	0.07	0.00	0.00	1.00	0.00	0.00	1.00		
12	3	2	4	6	5	1	1.00	0.00	0.92	0.08	0.99	0.01	0.99	0.00	1.00	0.00	0.00	1.00	0.00		
13	3	4	5	2	1	6	0.36	0.64	0.48	0.52	0.02	0.98	0.02	0.34	0.32	0.34	0.59	0.20	0.21		
14	3	2	4	6	5	1	1.00	0.00	0.92	0.08	0.99	0.01	0.99	0.00	1.00	0.00	0.00	1.00	0.00		
15	1	2	3	5	6	4	0.01	0.99	0.08	0.92	0.01	0.99	0.01	0.05	0.02	0.93	0.17	0.24	0.59		
16	3	2	1	5	6	4	0.00	1.00	0.03	0.97	0.02	0.98	0.02	0.02	0.04	0.94	0.20	0.27	0.53		
17	5	1	3	6	4	2	0.50	0.50	0.24	0.76	0.26	0.74	0.26	0.03	0.77	0.20	0.12	0.50	0.38		
18	5	3	2	1	4	6	0.10	0.90	0.41	0.59	0.01	0.99	0.01	0.31	0.29	0.40	0.56	0.19	0.25		

Table D.4 (continued)

Row	X-box	Play-Station	PSP/Portable	Game-Cube	Game-Boy	Personal Computer	Exp-FCMdK _{ent} ($p=0.05$)			Exp-FCMdK ($m=1.3$)			M-ISR			Exp-FCMdK _{ent,not} ($p=0.10$)			CCA							
							Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
							1	2	1	2	1	2	1	2	1	2	1	2	3	1	2	3	1	2	3	
19	3	2	4	5	6	1	0.99	0.01	0.99	0.01	0.99	0.01	0.83	0.17	0.13	0.74	0.13	0.06	0.63	0.31						
20	2	4	5	3	6	1	1.00	0.00	0.99	0.01	0.99	0.01	0.52	0.48	0.99	0.01	0.00	0.22	0.43	0.35						
21	2	5	4	6	3	1	1.00	0.00	0.93	0.07	0.43	0.57	0.43	0.57	0.24	0.71	0.06	0.20	0.51	0.29						
22	3	4	5	6	2	1	1.00	0.00	0.93	0.07	0.21	0.79	0.21	0.79	0.24	0.71	0.06	0.17	0.52	0.30						
23	1	2	3	6	4	5	0.00	1.00	0.24	0.76	0.08	0.92	0.08	0.92	0.10	0.23	0.68	0.24	0.34	0.42						
24	2	1	3	6	5	4	0.00	1.00	0.01	0.99	0.63	0.37	0.63	0.37	0.01	0.17	0.82	0.12	0.35	0.53						
25	1	4	3	6	5	2	1.00	0.00	0.84	0.16	0.30	0.70	0.30	0.70	0.17	0.66	0.17	0.16	0.48	0.36						
26	1	2	6	4	5	3	0.50	0.50	0.16	0.84	0.00	1.00	0.00	1.00	0.49	0.02	0.49	0.20	0.27	0.53						
27	1	2	3	5	6	4	0.01	0.99	0.08	0.92	0.01	0.99	0.01	0.99	0.05	0.02	0.93	0.17	0.24	0.59						
28	3	2	5	4	6	1	1.00	0.00	0.92	0.08	0.85	0.15	0.85	0.15	0.82	0.15	0.03	0.11	0.54	0.36						
29	1	4	5	3	6	2	1.00	0.00	0.84	0.16	0.09	0.91	0.09	0.91	0.93	0.02	0.05	0.24	0.34	0.42						
30	1	3	2	5	6	4	0.00	1.00	0.16	0.84	0.01	0.99	0.01	0.99	0.08	0.03	0.89	0.25	0.25	0.50						
31	4	2	3	6	5	1	1.00	0.00	0.84	0.16	0.95	0.05	0.95	0.05	0.01	0.98	0.01	0.06	0.75	0.19						
32	3	5	4	1	2	6	0.36	0.64	0.48	0.52	0.00	1.00	0.00	1.00	0.34	0.31	0.34	0.75	0.11	0.14						
33	4	5	6	3	2	1	0.97	0.03	0.70	0.30	0.21	0.79	0.21	0.79	0.56	0.31	0.13	0.31	0.40	0.28						
34	3	2	4	6	5	1	1.00	0.00	0.92	0.08	0.99	0.01	0.99	0.01	0.00	1.00	0.00	0.00	1.00	0.00						
35	4	6	1	2	3	5	0.50	0.50	0.53	0.47	0.01	0.99	0.01	0.99	0.34	0.32	0.34	0.59	0.20	0.21						
36	4	5	2	3	6	1	1.00	0.00	0.93	0.07	0.08	0.92	0.08	0.92	0.60	0.26	0.14	0.29	0.38	0.33						
37	6	2	1	4	5	3	0.01	0.99	0.36	0.64	0.01	0.99	0.01	0.99	0.19	0.29	0.53	0.26	0.34	0.40						
38	3	2	6	5	4	1	1.00	0.00	0.76	0.24	0.96	0.04	0.96	0.04	0.17	0.81	0.02	0.11	0.64	0.26						
39	1	2	4	6	5	3	0.50	0.50	0.08	0.92	0.09	0.91	0.09	0.91	0.04	0.17	0.79	0.12	0.35	0.53						
40	2	1	4	5	6	3	0.01	0.99	0.00	1.00	0.07	0.93	0.07	0.93	0.00	0.00	1.00	0.00	0.00	1.00						

Table D.4 (continued)

Row	X-box	Play-Station	PSP/Portable	Game-Cube	Game-Boy	Personal Computer	Exp-FCMdK _{ent} ($p=0.05$)			Exp-FCMdK ($m=1.3$)			M-ISR			Exp-FCMdK _{ent,not} ($p=0.10$)			CCA							
							Cluster		Cluster		Cluster		Cluster		Cluster		Cluster		Cluster		Cluster		Cluster		Cluster	
							1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	3	
41	2	1	4	3	5	6	0.00	1.00	0.07	0.93	0.17	0.83	0.12	0.06	0.82	0.34	0.24	0.42								
42	4	5	1	2	3	6	0.27	0.73	0.47	0.53	0.00	1.00	0.32	0.31	0.37	0.65	0.15	0.20								
43	2	1	6	4	5	3	0.00	1.00	0.01	0.99	0.03	0.97	0.05	0.02	0.93	0.14	0.25	0.62								
44	5	6	4	3	1	2	0.67	0.33	0.56	0.44	0.01	0.99	0.39	0.33	0.28	0.41	0.33	0.25								
45	1	3	4	6	5	2	1.00	0.00	0.92	0.08	0.55	0.45	0.15	0.70	0.15	0.11	0.54	0.36								
46	3	6	4	1	2	5	0.50	0.50	0.52	0.48	0.01	0.99	0.37	0.31	0.32	0.68	0.16	0.17								
47	2	3	4	6	5	1	1.00	0.00	1.00	0.00	0.83	0.17	0.15	0.84	0.01	0.07	0.75	0.19								
48	1	4	3	5	6	2	1.00	0.00	0.92	0.08	0.02	0.98	0.44	0.12	0.44	0.18	0.35	0.47								
49	5	1	3	2	4	6	0.00	1.00	0.30	0.70	0.01	0.99	0.26	0.22	0.52	0.41	0.24	0.35								
50	3	6	4	2	5	1	1.00	0.00	0.86	0.14	0.13	0.87	0.89	0.07	0.05	0.43	0.31	0.27								
51	4	5	2	1	3	6	0.36	0.64	0.48	0.52	0.00	1.00	0.34	0.31	0.34	0.75	0.11	0.14								
52	2	3	5	6	4	1	1.00	0.00	0.99	0.01	0.51	0.49	0.17	0.81	0.02	0.11	0.64	0.26								
53	1	2	5	6	4	3	0.50	0.50	0.16	0.84	0.02	0.98	0.07	0.19	0.74	0.15	0.37	0.49								
54	2	3	4	6	5	1	1.00	0.00	1.00	0.00	0.83	0.17	0.15	0.84	0.01	0.07	0.75	0.19								
55	3	5	2	6	4	1	1.00	0.00	0.93	0.07	0.27	0.73	0.24	0.71	0.06	0.20	0.51	0.29								
56	2	1	5	6	4	3	0.00	1.00	0.01	0.99	0.26	0.74	0.01	0.17	0.82	0.10	0.36	0.54								
57	3	2	4	6	5	1	1.00	0.00	0.92	0.08	0.99	0.01	0.00	1.00	0.00	1.00	0.00	0.00								
58	2	3	4	6	5	1	1.00	0.00	1.00	0.00	0.81	0.19	0.15	0.84	0.01	0.07	0.75	0.19								
59	2	4	3	6	5	1	1.00	0.00	0.99	0.01	0.84	0.16	0.17	0.81	0.02	0.12	0.62	0.25								
60	2	4	3	6	5	1	1.00	0.00	0.99	0.01	0.83	0.17	0.17	0.81	0.02	0.12	0.62	0.25								
61	3	4	1	5	2	6	0.16	0.84	0.45	0.55	0.01	0.99	0.29	0.32	0.38	0.44	0.26	0.29								
62	5	4	6	2	3	1	0.93	0.07	0.60	0.40	0.26	0.74	0.53	0.29	0.19	0.29	0.38	0.33								

Table D.4 (continued)

Row	X-box	Play-Station	PSP/Portable	Game-Cube	Game-Boy	Personal Computer	Exp-FCMdK _{ent} ($p=0.05$)			Exp-FCMdK _{ent} ($m=1.3$)			M-ISR			Exp-FCMdK _{ent,not} ($p=0.10$)			CCA						
							Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
							1	2	1	2	1	2	1	2	1	2	1	2	1	2	3	1	2	3	1
63	1	6	3	4	5	2	1.00	0.00	0.76	0.24	0.01	0.99	0.41	0.18	0.41	0.31	0.31	0.38							
64	2	3	6	4	5	1	1.00	0.00	0.99	0.01	0.07	0.93	0.99	0.01	0.00	0.18	0.45	0.36							
65	3	2	4	5	6	1	0.99	0.01	0.99	0.01	0.82	0.18	0.13	0.74	0.13	0.06	0.63	0.31							
66	1	4	3	2	5	6	0.02	0.98	0.40	0.60	0.02	0.98	0.41	0.17	0.41	0.67	0.13	0.19							
67	5	6	2	1	3	4	0.64	0.36	0.56	0.44	0.04	0.96	0.41	0.30	0.29	0.61	0.20	0.19							
68	3	1	4	6	5	2	0.50	0.50	0.08	0.92	0.83	0.17	0.01	0.84	0.15	0.06	0.63	0.31							
69	1	2	4	6	5	3	0.50	0.50	0.08	0.92	0.07	0.93	0.04	0.17	0.79	0.12	0.35	0.53							
70	2	3	4	5	6	1	1.00	0.00	1.00	0.00	0.19	0.81	0.82	0.15	0.03	0.11	0.54	0.36							
71	1	3	6	2	4	5	0.50	0.50	0.36	0.64	0.02	0.98	0.44	0.12	0.44	0.42	0.23	0.35							
72	3	1	6	5	4	2	0.50	0.50	0.24	0.76	0.56	0.44	0.17	0.66	0.17	0.13	0.49	0.37							
73	1	2	5	3	4	6	0.01	0.99	0.36	0.64	0.00	1.00	0.44	0.12	0.44	0.42	0.23	0.35							
74	3	1	5	4	6	2	0.50	0.50	0.08	0.92	0.24	0.76	0.45	0.10	0.45	0.12	0.35	0.53							
75	2	1	4	3	5	6	0.00	1.00	0.07	0.93	0.17	0.83	0.12	0.06	0.82	0.34	0.24	0.42							
76	2	1	4	5	6	3	0.01	0.99	0.00	1.00	0.07	0.93	0.00	0.00	1.00	0.00	1.00	1.00							
77	2	3	5	4	6	1	1.00	0.00	1.00	0.00	0.24	0.76	1.00	0.00	0.00	0.16	0.48	0.36							
78	2	3	6	4	5	1	1.00	0.00	0.99	0.01	0.07	0.93	0.99	0.01	0.00	0.18	0.45	0.36							
79	2	1	3	6	5	4	0.00	1.00	0.01	0.99	0.66	0.34	0.01	0.17	0.82	0.12	0.35	0.53							
80	5	2	3	6	4	1	1.00	0.00	0.76	0.24	0.85	0.15	0.02	0.96	0.02	0.09	0.65	0.26							
81	6	3	5	2	4	1	0.93	0.07	0.60	0.40	0.11	0.89	0.53	0.29	0.19	0.25	0.40	0.35							
82	5	6	2	3	4	1	1.00	0.00	0.78	0.22	0.01	0.99	0.46	0.30	0.23	0.35	0.35	0.31							
83	4	3	1	6	5	2	0.50	0.50	0.69	0.31	0.96	0.04	0.04	0.93	0.04	0.15	0.56	0.28							
84	4	5	6	2	3	1	1.00	0.00	0.78	0.22	0.02	0.98	0.79	0.12	0.09	0.35	0.35	0.31							

Table D.4 (continued)

Row	X-box	Play-Station	PSPort-able	Game-Cube	Game-Boy	Per-sonal Com-puter	Exp-FCMdK _{ent} ($p=0.05$)		Exp-FCMdK ($m=1.3$)		M-ISR		Exp-FCMdK _{ent,msr} ($p=0.10$)			CCA		
							Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
85	3	2	5	4	6	1	1.00	0.00	0.92	0.08	0.86	0.14	0.82	0.15	0.03	0.11	0.54	0.36
86	2	3	5	4	6	1	1.00	0.00	1.00	0.00	0.21	0.79	1.00	0.00	0.00	0.16	0.48	0.36
87	1	3	5	6	2	4	0.50	0.50	0.31	0.69	0.03	0.97	0.14	0.26	0.60	0.25	0.34	0.41
88	3	2	4	6	5	1	1.00	0.00	0.92	0.08	0.99	0.01	0.00	1.00	0.00	0.00	1.00	0.00
89	5	2	6	4	3	1	0.98	0.02	0.64	0.36	0.87	0.13	0.24	0.71	0.06	0.17	0.52	0.30
90	5	3	2	4	6	1	1.00	0.00	0.76	0.24	0.28	0.72	0.20	0.60	0.20	0.18	0.45	0.36
91	4	1	3	5	6	2	0.50	0.50	0.08	0.92	0.05	0.95	0.04	0.17	0.79	0.10	0.36	0.54

Table D.5 University Rankings dataset: membership degrees matrix provided by the applied clustering methods

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{ent,post} $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
1	1	5	1	2	4	3	0.60	0.40	0.55	0.45
2	1	4	3	5	2	5	0.01	0.99	0.24	0.76
3	2	4	6	5	1	3	0.33	0.67	0.44	0.56
4	1	4	3	3	2	2	0.38	0.62	0.48	0.52
5	1	3	4	2	6	5	0.97	0.03	0.73	0.27
6	2	4	3	5	1	3	0.14	0.86	0.37	0.63
7	1	3	1	3	3	2	0.38	0.62	0.48	0.52
8	2	1	5	3	6	4	0.99	0.01	0.68	0.32
9	2	1	6	3	4	5	0.94	0.06	0.60	0.40
10	1	4	3	5	2	6	0.03	0.97	0.29	0.71
11	1	2	4	6	3	5	0.00	1.00	0.09	0.91
12	1	3	2	2	1	4	0.16	0.84	0.43	0.57
13	1	4	4	2	4	3	0.95	0.05	0.70	0.30
14	1	3	1	1	2	4	0.65	0.35	0.56	0.44
15	5	2	3	4	1	6	0.18	0.82	0.38	0.62
16	1	1	3	1	2	4	0.79	0.21	0.62	0.38
17	1	2	1	5	3	4	0.04	0.96	0.32	0.68
18	4	5	1	2	3	6	0.50	0.50	0.52	0.48
19	2	2	1	4	3	3	0.16	0.84	0.39	0.61
20	2	4	1	2	3	4	0.50	0.50	0.52	0.48
21	1	6	4	3	5	2	0.74	0.26	0.61	0.39
22	1	1	2	4	3	3	0.06	0.94	0.33	0.67
23	1	5	2	3	1	4	0.14	0.86	0.40	0.60
24	2	1	4	3	5	6	0.94	0.06	0.60	0.40
25	5	1	3	4	2	6	0.10	0.90	0.36	0.64

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{ent_{max}} $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
26	1	3	3	3	2	3	0.10	0.90	0.41	0.59
27	2	3	5	1	2	4	0.81	0.19	0.61	0.39
28	3	1	1	2	2	4	0.38	0.62	0.45	0.55
29	1	2	6	5	3	4	0.19	0.81	0.38	0.62
30	1	5	3	2	4	6	0.84	0.16	0.62	0.38
31	3	2	4	3	4	1	0.82	0.18	0.59	0.41
32	1	2	5	3	6	4	1.00	0.00	0.73	0.27
33	1	3	3	1	3	2	0.92	0.08	0.68	0.32
34	1	4	5	2	5	3	0.98	0.02	0.74	0.26
35	1	2	5	3	4	3	0.95	0.05	0.62	0.38
36	1	4	4	3	5	2	0.86	0.14	0.64	0.36
37	2	1	1	4	3	5	0.13	0.87	0.38	0.62
38	2	1	3	2	3	2	0.95	0.05	0.62	0.38
39	2	1	4	4	3	1	0.33	0.67	0.44	0.56
40	2	1	5	3	4	6	0.81	0.19	0.54	0.46
41	4	2	4	1	4	3	0.81	0.19	0.65	0.35
42	2	1	2	3	2	4	0.08	0.92	0.35	0.65
43	2	6	1	5	3	4	0.26	0.74	0.42	0.58
44	1	3	2	4	4	4	0.21	0.79	0.40	0.60
45	1	2	2	4	3	4	0.01	0.99	0.23	0.77
46	1	1	2	4	3	2	0.23	0.77	0.41	0.59
47	1	2	3	2	4	2	0.96	0.04	0.71	0.29
48	2	3	4	1	4	4	0.99	0.01	0.80	0.20
49	1	2	4	5	3	5	0.00	1.00	0.00	1.00
50	1	5	2	3	4	5	0.50	0.50	0.52	0.48

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{opt} , $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
51	2	4	3	2	1	5	0.35	0.65	0.48	0.52
52	4	3	2	5	1	5	0.11	0.89	0.34	0.66
53	1	2	4	3	4	4	1.00	0.00	0.89	0.11
54	1	2	2	3	3	4	0.19	0.81	0.38	0.62
55	3	2	1	2	1	3	0.40	0.60	0.46	0.54
56	2	4	3	5	1	5	0.02	0.98	0.29	0.71
57	1	3	2	5	6	4	0.50	0.50	0.47	0.53
58	2	1	4	5	3	5	0.00	1.00	0.15	0.85
59	2	1	5	3	4	6	0.81	0.19	0.54	0.46
60	2	4	3	1	2	4	0.62	0.38	0.56	0.44
61	4	5	5	3	1	2	0.46	0.54	0.49	0.51
62	3	4	4	2	1	3	0.50	0.50	0.52	0.48
63	1	1	1	3	2	2	0.10	0.90	0.37	0.63
64	1	4	4	2	3	3	0.90	0.10	0.65	0.35
65	1	3	3	3	2	3	0.10	0.90	0.41	0.59
66	3	5	4	1	3	2	0.62	0.38	0.59	0.41
67	2	1	1	1	2	3	0.62	0.38	0.56	0.44
68	1	2	2	3	2	2	0.23	0.77	0.41	0.59
69	1	3	4	5	2	6	0.01	0.99	0.23	0.77
70	1	3	5	2	6	4	0.99	0.01	0.76	0.24
71	1	1	2	1	2	2	1.00	0.00	0.85	0.15
72	2	1	3	4	2	3	0.03	0.97	0.29	0.71
73	1	4	2	6	3	5	0.05	0.95	0.33	0.67
74	1	2	4	6	3	5	0.00	1.00	0.09	0.91
75	3	2	2	4	2	1	0.46	0.54	0.46	0.54

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{ent,mod} $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
76	1	2	3	4	1	5	0.02	0.98	0.27	0.73
77	1	3	4	2	4	2	0.95	0.05	0.68	0.32
78	3	2	4	1	5	5	0.93	0.07	0.67	0.33
79	2	1	3	3	3	3	0.81	0.19	0.54	0.46
80	2	6	4	1	3	5	0.62	0.38	0.56	0.44
81	1	3	2	6	4	5	0.03	0.97	0.29	0.71
82	2	4	1	3	4	4	0.65	0.35	0.56	0.44
83	4	3	1	4	2	4	0.31	0.69	0.40	0.60
84	1	4	4	3	2	3	0.33	0.67	0.48	0.52
85	1	3	5	5	2	4	0.08	0.92	0.35	0.65
86	3	4	4	1	4	2	0.81	0.19	0.65	0.35
87	1	1	3	1	2	3	0.94	0.06	0.69	0.31
88	2	3	4	1	4	4	0.99	0.01	0.80	0.20
89	2	5	5	1	3	4	0.73	0.27	0.59	0.41
90	1	2	5	4	6	3	0.97	0.03	0.65	0.35
91	3	1	2	5	6	4	0.50	0.50	0.48	0.52
92	3	2	4	4	1	4	0.13	0.87	0.38	0.62
93	2	1	4	2	3	5	0.79	0.21	0.53	0.47
94	1	2	3	3	3	3	0.84	0.16	0.56	0.44
95	1	4	3	5	2	6	0.03	0.97	0.29	0.71
96	1	3	5	2	4	5	0.95	0.05	0.73	0.27
97	1	5	5	3	2	4	0.50	0.50	0.52	0.48
98	1	2	3	3	3	4	0.50	0.50	0.44	0.56
99	2	4	4	3	1	4	0.50	0.50	0.52	0.48
100	1	3	4	2	5	5	0.98	0.02	0.77	0.23

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{ent_{max}} $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
101	1	4	6	3	2	5	0.50	0.50	0.53	0.47
102	1	3	3	3	3	2	0.67	0.33	0.57	0.43
103	4	3	2	6	1	5	0.15	0.85	0.35	0.65
104	1	3	2	5	4	5	0.02	0.98	0.27	0.73
105	2	1	1	4	3	3	0.16	0.84	0.39	0.61
106	2	1	6	4	5	3	0.90	0.10	0.58	0.42
107	1	3	2	3	4	3	0.50	0.50	0.53	0.47
108	1	2	2	2	2	2	0.77	0.23	0.60	0.40
109	2	1	3	2	4	4	0.97	0.03	0.65	0.35
110	4	2	5	1	5	3	0.89	0.11	0.67	0.33
111	2	5	4	6	3	1	0.30	0.70	0.42	0.58
112	4	2	6	1	3	5	0.77	0.23	0.60	0.40
113	3	4	5	1	2	5	0.81	0.19	0.61	0.39
114	1	4	4	4	3	2	0.35	0.65	0.48	0.52
115	2	1	2	3	2	4	0.08	0.92	0.35	0.65
116	1	1	4	2	5	3	0.99	0.01	0.71	0.29
117	2	1	4	5	3	5	0.00	1.00	0.15	0.85
118	5	1	2	3	4	5	0.50	0.50	0.48	0.52
119	2	3	6	1	5	4	0.95	0.05	0.68	0.32
120	1	1	1	2	1	2	0.04	0.96	0.32	0.68
121	3	2	3	5	1	4	0.05	0.95	0.32	0.68
122	1	4	2	3	3	4	0.33	0.67	0.48	0.52
123	3	5	2	4	1	6	0.18	0.82	0.41	0.59
124	2	3	5	1	5	4	0.98	0.02	0.74	0.26
125	1	1	1	1	1	1	0.65	0.35	0.56	0.44

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{opt} , $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
126	2	1	6	5	4	3	0.50	0.50	0.47	0.53
127	3	1	4	5	2	5	0.01	0.99	0.24	0.76
128	1	2	3	4	5	6	0.50	0.50	0.45	0.55
129	1	2	4	6	3	5	0.00	1.00	0.09	0.91
130	1	2	4	5	3	6	0.01	0.99	0.11	0.89
131	2	5	1	3	4	6	0.50	0.50	0.52	0.48
132	1	1	2	3	3	3	0.19	0.81	0.38	0.62
133	3	1	4	4	2	4	0.08	0.92	0.35	0.65
134	1	3	6	2	5	4	0.97	0.03	0.73	0.27
135	1	2	3	3	3	3	0.84	0.16	0.56	0.44
136	1	4	3	4	2	2	0.23	0.77	0.44	0.56
137	1	2	4	5	3	6	0.01	0.99	0.11	0.89
138	2	1	3	4	5	6	0.50	0.50	0.47	0.53
139	2	3	1	4	4	4	0.25	0.75	0.42	0.58
140	1	2	3	3	3	4	0.50	0.50	0.44	0.56
141	3	1	5	6	2	4	0.05	0.95	0.33	0.67
142	1	3	4	2	6	5	0.97	0.03	0.73	0.27
143	1	3	5	2	4	5	0.95	0.05	0.73	0.27
144	2	1	2	4	1	3	0.03	0.97	0.30	0.70
145	2	1	4	6	5	3	0.28	0.72	0.43	0.57
146	3	1	2	4	2	3	0.10	0.90	0.36	0.64
147	1	2	3	4	4	4	0.17	0.83	0.36	0.64
148	1	4	6	2	5	3	0.95	0.05	0.68	0.32
149	1	2	3	6	4	5	0.01	0.99	0.23	0.77
150	1	2	3	5	4	6	0.04	0.96	0.27	0.73

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{ent_{max}} $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
151	1	2	4	4	4	3	0.79	0.21	0.53	0.47
152	1	2	4	3	5	6	0.96	0.04	0.64	0.36
153	2	1	4	5	3	4	0.02	0.98	0.27	0.73
154	1	2	4	5	4	3	0.21	0.79	0.40	0.60
155	2	3	3	1	3	4	0.87	0.13	0.63	0.37
156	3	2	2	5	1	4	0.11	0.89	0.34	0.66
157	1	2	4	3	4	4	1.00	0.00	0.89	0.11
158	3	1	4	2	4	4	0.99	0.01	0.71	0.29
159	1	1	3	2	3	2	0.98	0.02	0.67	0.33
160	2	4	1	3	5	6	0.50	0.50	0.52	0.48
161	1	2	3	4	2	4	0.00	1.00	0.09	0.91
162	1	2	3	3	4	4	0.96	0.04	0.64	0.36
163	3	2	6	5	1	4	0.33	0.67	0.44	0.56
164	1	6	5	4	3	2	0.50	0.50	0.52	0.48
165	3	4	5	1	2	6	0.67	0.33	0.57	0.43
166	3	2	5	4	1	6	0.28	0.72	0.43	0.57
167	4	3	6	2	5	1	0.63	0.37	0.58	0.42
168	1	5	4	2	6	3	0.92	0.08	0.68	0.32
169	2	2	1	2	3	4	0.50	0.50	0.52	0.48
170	2	4	1	6	3	5	0.10	0.90	0.36	0.64
171	1	1	2	1	2	2	1.00	0.00	0.85	0.15
172	2	1	5	4	3	6	0.19	0.81	0.38	0.62
173	2	1	5	3	4	6	0.81	0.19	0.54	0.46
174	3	1	2	4	5	6	0.50	0.50	0.47	0.53
175	1	6	5	3	4	2	0.70	0.30	0.59	0.41

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{opt} , $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
176	3	2	3	4	1	3	0.14	0.86	0.37	0.63
177	4	3	2	5	1	6	0.18	0.82	0.38	0.62
178	2	6	5	1	4	3	0.74	0.26	0.61	0.39
179	2	1	5	3	6	4	0.99	0.01	0.68	0.32
180	2	1	4	4	3	4	0.05	0.95	0.31	0.69
181	3	2	3	1	3	1	0.66	0.34	0.60	0.40
182	5	2	1	2	4	3	0.54	0.46	0.54	0.46
183	3	1	4	5	6	2	0.67	0.33	0.52	0.48
184	3	5	6	2	1	4	0.50	0.50	0.52	0.48
185	4	1	3	2	5	6	0.84	0.16	0.57	0.43
186	1	4	2	5	6	3	0.50	0.50	0.48	0.52
187	3	3	4	2	1	2	0.45	0.55	0.49	0.51
188	2	4	4	1	3	5	0.70	0.30	0.57	0.43
189	4	5	5	1	2	3	0.60	0.40	0.57	0.43
190	2	5	3	6	1	4	0.08	0.92	0.33	0.67
191	2	5	6	1	4	3	0.82	0.18	0.63	0.37
192	1	4	4	2	4	3	0.95	0.05	0.70	0.30
193	1	2	3	6	5	4	0.06	0.94	0.33	0.67
194	2	1	3	5	3	4	0.02	0.98	0.27	0.73
195	4	6	5	1	2	3	0.58	0.42	0.57	0.43
196	3	1	4	5	2	6	0.03	0.97	0.29	0.71
197	4	3	1	5	2	6	0.30	0.70	0.42	0.58
198	2	2	2	1	2	3	0.77	0.23	0.60	0.40
199	3	6	1	4	5	2	0.50	0.50	0.51	0.49
200	1	6	5	2	4	3	0.82	0.18	0.63	0.37

Table D.5 (continued)

Row	London	Paris	Milan	St. Gallen	Barcelona	Stockholm	Exp-FCMdK _{ent,ross} $p = 0.10$		CCA	
							Cluster 1	Cluster 2	Cluster 1	Cluster 2
201	2	1	2	5	3	4	0.04	0.96	0.32	0.68
202	1	5	6	2	3	4	0.84	0.16	0.62	0.38
203	1	4	5	2	1	3	0.50	0.50	0.52	0.48
204	3	1	5	4	2	5	0.50	0.50	0.47	0.53
205	2	2	1	4	2	3	0.14	0.86	0.37	0.63
206	1	2	4	3	6	5	0.99	0.01	0.69	0.31
207	4	3	1	6	2	5	0.26	0.74	0.39	0.61
208	1	3	2	2	2	1	0.58	0.42	0.55	0.45
209	4	1	3	5	2	5	0.02	0.98	0.29	0.71
210	1	2	4	3	5	6	0.96	0.04	0.64	0.36
211	2	3	4	6	1	5	0.01	0.99	0.26	0.74
212	4	3	2	3	1	4	0.34	0.66	0.43	0.57

Acknowledgements The authors thank the Editor and the referees for their useful comments and suggestions which helped to improve the quality and presentation of this manuscript.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Data Availability The datasets generated during the simulation study are available from the corresponding author on reasonable request.

Declarations

This research does not contain any studies with human participations or animals performed by any of the authors.

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bellman, R., Kalaba, R., & Zadeh, L. (1966). Abstraction and pattern classification. *Journal of Mathematical Analysis and Applications*, 13(1), 1–7.
- Bezdek, J. C. (1974). Numerical taxonomy with fuzzy sets. *Journal of Math Biology*, 1, 57–71. <https://doi.org/10.1007/BF02339490>
- Bezdek, J.C. (1981). Objective function clustering. In *Pattern recognition with fuzzy objective function algorithms* (pp. 43–93). Springer.
- Biernacki, C., & Jacques, J. (2013). A generative model for rank data based on insertion sort algorithm. *Computational Statistics & Data Analysis*, 58, 162–176.
- Bonanomi, A., Nai Ruscone, M., & Osmetti, S. (2017). Defining subjects distance in hierarchical cluster analysis by copula approach. *Quality & Quantity*, 51(2), 859–872.
- Bonanomi, A., Nai Ruscone, M., & Osmetti, S. (2019). Dissimilarity measure for ranking data via mixture of copulae. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(5), 412–425.
- Bradley, R. A. (1984). Paired comparisons: Some basic procedures and examples. *Handbook of Statistics*, 4, 299–326.
- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: i. the method of paired comparisons. *Biometrika*, 39(3/4), 324–345.
- Brentari, E., Dancelli, L., & Manisera, M. (2016). Clustering ranking data in market segmentation: a case study on the Italian Mcdonald's customers' preferences. *Journal of Applied Statistics*, 43(11), 1959–1976.
- Campello, J. R., & Hruschka, E. R. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157(21), 2858–2875.
- Church, R. (1978). Contrasts between facility location approaches and non-hierarchical cluster analysis. In *ORSA/TIMS Joint National Meeting*.
- Coppi, R., & D'Urso, P. (2006). Fuzzy unsupervised classification of multivariate time trajectories with the shannon entropy regularization. *Computational Statistics & Data Analysis*, 50(6), 1452–1477.
- Critchlow, D. E., Fligner, M. A., & Verducci, J. S. (1991). Probability models on rankings. *Journal of Mathematical Psychology*, 35(3), 294–318.
- D'Ambrosio, A. (2021). ConsRankClass, classification and clustering of preference rankings. R package version 101. <https://CRAN.R-project.org/package=ConsRankClass>
- D'Ambrosio, A., & Heiser, W. J. (2019). A distribution-free soft-clustering method for preference rankings. *Behaviormetrika*, 46(2), 333–351.

- D'Ambrosio, A., Amodio, S., Iorio, C., Pandolfo, G., & Siciliano, R. (2021). Adjusted concordance index: an extension of the adjusted rand index to fuzzy partitions. *Journal of Classification*, 38(1), 112–128.
- Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3), 32–57.
- D'Urso, P. In C. Hennig, M. Meila, F. Murtagh, & R. Rocci (Eds.) (2015). *Fuzzy clustering*. Chap: Chapman and Hall/CRC.
- D'Urso, P., & De Giovanni, L. (2014). Robust clustering of imprecise data. *Chemometrics and Intelligent Laboratory Systems*, 136, 58–80.
- D'Urso, P., De Giovanni, L., & Massari, R. (2018). Robust fuzzy clustering of multivariate time trajectories. *International Journal of Approximate Reasoning*, 99, 12–38.
- Emond, E. J., & Mason, D. W. (2002). A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis*, 11(1), 17–28.
- Everitt, S. B. S., & Landau, L. M. (2001). *Cluster Analysis*. London: Arnold Press.
- Fligner, M. A., & Verducci, J. S. (1988). Multistage ranking models. *Journal of the American Statistical Association*, 83(403), 892–901.
- Fok, D., Paap, R., & Van Dijk, B. (2012). A rank-ordered logit model with unobserved heterogeneity in ranking capabilities. *Journal of Applied Econometrics*, 27(5), 831–846.
- Franczak, B. C., Castura, J. C., Browne, R. P., Findlay, C. J., & McNicholas, P. D. (2016). Handling missing data in consumer hedonic tests arising from direct scaling. *Journal of Sensory Studies*, 31(6), 514–523. <https://doi.org/10.1111/joss.12241>
- García-Escudero, L. A., & Gordaliza, A. (2005). A proposal for robust curve clustering. *Journal of Classification*, 22(2), 185–201.
- García-Escudero, L. A., Gordaliza, A., & Matrán, C. (2003). Trimming tools in exploratory data analysis. *Journal of Computational and Graphical Statistics*, 12(2), 434–449.
- García-Escudero, L. A., Gordaliza, A., Matrán, C., & Mayo-Iscar, A. (2008). A general trimming approach to robust cluster analysis. *The Annals of Statistics*, 36(3), 1324–1345.
- Heiser, W.J., & D'Ambrosio, A. (2013). Clustering and prediction of rankings within a kemeny distance framework. In B. Lausen, D. Van den Poel, & A. Ultsch (Eds.) *Algorithms from and for Nature and Life* (pp. 19–31). Springer International Publishing.
- Hullermeier, E., Rifqi, M., Henzgen, S., & Senge, R. (2011). Comparing fuzzy partitions: a generalization of the rand index and related measures. *IEEE Transactions on Fuzzy Systems*, 20(3), 546–556.
- Hwang, H., DeSarbo, W. S., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, 72(2), 181–198.
- Irurozki, E., Calvo, B., & Lozano, J. (2016). Permallows: an r package for mallows and generalized mallows models. *Journal of Statistical Software, Articles*, 71(12), 1–30. <https://doi.org/10.18637/jss.v071.i12>.
- Jacques, J., & Biernacki, C. (2014). Model-based clustering for multivariate partial ranking data. *Journal of Statistical Planning and Inference*, 149, 201–217.
- Jacques, J., Grimonprez, Q., & Biernacki, C. (2014). Rankcluster: an r package for clustering multivariate partial rankings. *The R Journal*, 6(1), 101–110. <https://doi.org/10.32614/RJ-2014-010>
- Kaufman, L., & Rousseeuw, P. (1987). Clustering by means of medoids. In *Statistical data analysis based on the L1-norm and related methods* (pp. 405–416). North-Holland, Amsterdam: Y. Dodge.
- Kemeny, J.G., & Snell, L. (1962). *Mathematical models in the social sciences*. Ginn and Company.
- Kendall, M. (1948). *Rank correlation methods*. London: Charles Griffin and Company Limited.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81–93.
- Krishnapuram, R., Joshi, A., & Yi, L. (1999). A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering. In *International fuzzy systems conference (FUZZIEEE99)*, (Vol. 3 pp. 1281–1286). Seoul: IEEE.
- Krishnapuram, R., Joshi, A., Nasraoui, O., & Yi, L. (2001). Low-complexity fuzzy relational clustering algorithms for web mining. *IEEE Transactions on Fuzzy Systems*, 9(4), 595–607.
- Lee, P. H., & Yu, P. L. (2012). Mixtures of weighted distance-based models for ranking data with applications in political studies. *Computational Statistics & Data Analysis*, 56(8), 2486–2500.
- Li, P., Mazumdar, A., & Milenkovic, O. (2017). Efficient rank aggregation via lehmer codes. In *Artificial Intelligence and Statistics, PMLR* (pp. 450–459).
- Li, R., & Mukaidono, M. (1995). A maximum entropy approach to fuzzy clustering. In *Proceedings of the Fourth IEEE Conference on Fuzzy Systems (FUZZ-IEEE/IFES '95)* (pp. 2227–2232).
- Li, R. P., & Mukaidono, M. (1999). Gaussian clustering method based on maximum-fuzzy-entropy interpretation. *Fuzzy Sets and Systems*, 102(2), 253–258.
- Luce, R.D. (1959). *Individual choice behavior: a theoretical analysis*. John Wiley & Sons Inc.
- Mallows, C. L. (1957). Non-null ranking models. *Biometrika*, 44, 114–130.

- Marden, J.I. (1995). Analyzing and modeling rank data, Monographs on Statistics and Applied Probability, vol 64. Chapman and Hall/CRC.
- McBratney, A., & Moore, A. (1985). Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 35(1–4), 165–185.
- Meila, M., & Chen, H. (2010). Dirichlet process mixtures of generalized mallows models. In P. Grunwald P. Spirtes (Eds.) *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*.
- Miyamoto, S., & Mukaidono, M. (1997). Fuzzy c-Means as a regularization and maximum entropy approach. In *Proc. of 7th international fuzzy systems association world congress (IFSA'97), II* (pp. 86–92).
- Mollica, C., & Tardella, L. (2017). Bayesian plackett-luce mixture models for partially ranked data. *Psychometrika*, 82(2), 442–458. <https://doi.org/10.1007/s11336-016-9530-0>
- Müllensiefen, D., Hennig, C., & Howells, H. (2018). Using clustering of rankings to explain brand preferences with personality and socio-demographic variables. *Journal of Applied Statistics*, 45(6), 1009–1029.
- Mulvey, J. M., & Crowder, H. P. (1979). Cluster analysis: an application of lagrangian relaxation. *Management Science*, 25(4), 329–340.
- Murphy, T. B., & Martin, D. (2003). Mixtures of distance-based models for ranking data. *Computational statistics & data analysis*, 41(3–4), 645–655.
- Plackett, R. L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(2), 193–202.
- Rao, M. (1971). Cluster analysis and mathematical programming. *Journal of the American statistical association*, 66(335), 622–626.
- Ruspini, E. (1969). A new approach to clustering. *Information and Control*, 16, 22–32.
- Ruspini, E. (1970). Numerical methods of fuzzy clustering. *Information Science*, 2, 319–350.
- Ruspini, E. (1973). New experimental results in fuzzy clustering. *Information Science*, 6, 273–284.
- Smith, B. B. (1950). Discussion of professor ross's paper. *Journal of the Royal Statistical Society B*, 12(1), 41–59.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273.
- Vinod, H. D. (1969). Integer programming and the theory of grouping. *Journal of the American Statistical association*, 64(326), 506–519.
- Wedel, M., & Kamakura, W. (1998). *Market segmentation: Conceptual and methodological foundations*. Boston: SKluwer Academic Press.
- Wu, K., & Yang, M. (2002). Alternative c-Means clustering algorithms. *Pattern recognition*, 35(10), 2267–2278.
- Zadeh, L. A. (1965). Information and control. *Fuzzy Sets*, 8(3), 338–353.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.