



Clustering alternatives in preference-approvals via novel pseudometrics

Alessandro Albano¹  · José Luis García-Lapresta² · Antonella Plaia¹ · Mariangela Sciandra¹

Accepted: 4 August 2023
© The Author(s) 2023

Abstract

Preference-approval structures combine preference rankings and approval voting for declaring opinions over a set of alternatives. In this paper, we propose a new procedure for clustering alternatives in order to reduce the complexity of the preference-approval space and provide a more accessible interpretation of data. To that end, we present a new family of pseudometrics on the set of alternatives that take into account voters' preferences via preference-approvals. To obtain clusters, we use the Ranked k -medoids (RKM) partitioning algorithm, which takes as input the similarities between pairs of alternatives based on the proposed pseudometrics. Finally, using non-metric multidimensional scaling, clusters are represented in 2-dimensional space.

Keywords Preference-approvals · Pseudometric · Clustering · Non metric multidimensional scaling · Voting systems

✉ Alessandro Albano
alessandro.albano@unipa.it

José Luis García-Lapresta
lapresta@eco.uva.es

Antonella Plaia
antonella.plaia@unipa.it

Mariangela Sciandra
mariangela.sciandra@unipa.it

¹ Department of Economics, Business and Statistics, University of Palermo, Viale delle Scienze, Edificio 13, 90129 Palermo, Sicily, Italy

² IMUVA, PRESAD Research Group, Departamento de Economía Aplicada, Universidad de Valladolid, Valladolid, Spain

1 Introduction

Preference-approvals are preference structures for declaring opinions over a set of alternatives. They combine decision makers' preference orderings and classify the alternatives as either acceptable or unacceptable (see Brams 2008, Chapter 3, Brams and Sanver 2009 and Sanver 2010). Thus, in preference-approval structures, voters should declare which alternatives are acceptable and rank-order them. Additionally, voters may either rank-order unacceptable alternatives or avoid displaying their preferences about them, as in *fallback voting* (Brams and Sanver 2009), by showing indifference between these alternatives.

Preference-approval structures have been studied from various perspectives, with a significant focus on exploring their basic properties (Dong et al. 2021) and achieving consensus in group decision making (GDM) (Erdamar et al. 2014; Liang et al. 2018; Barokas and Sprumont 2022). In particular, Barokas (2022a) introduced a social choice rule known as *majority approval* and compared it to other social choice rules. Additionally, Barokas (2022b) developed an axiomatic approach to allocation rules that are mathematically equivalent to preference-approvals but different from the voting rules.

A study conducted by Kruger and Sanver (2021) investigated preference-approvals and identified that there could be issues with reconciling ranking information and approval information in the method. In fact, they demonstrated that aggregating preference-approvals by decomposing the rankings and approvals could be dictatorial, indicating that the preferences of a single individual or a small group of individuals may excessively influence the resulting decision. Furthermore, Liu et al. (2023) proposed a model for Multi-Criteria Group Decision Making problems using the Preference Approval Structure approach, considering the Partial Information of Linguistic Terms to increase consistency between the preference-approvals and multi-criteria assessments.

Nevertheless, little effort has been devoted to developing clustering algorithms that deal with preference-approvals. The clustering task deals with classifying objects in homogeneous clusters, such that objects in a cluster have a higher degree of similarity than they do with items from other clusters. (see Jain et al. 1999 and Everitt et al. 2011). To the best of our knowledge, the only proposal applying clustering algorithms to preference-approval structures is found in Albano et al. (2023). They introduced a family of distances between preference-approvals and used a simple hierarchical clustering algorithm to find homogeneous groups of individuals. However, the possibility of clustering alternatives in preference-approvals has not yet been addressed. The goal of this paper is to fill this gap by demonstrating that identifying homogeneous groups of alternatives can be beneficial in reducing the complexity of the preference-approval space and making the data easier to interpret. Indeed, developing a method for clustering alternatives based on preference-approvals has several practical implications that can benefit decision-makers in a variety of settings by enabling them to identify potential trade-offs and conflicts between different policy options. For instance, clustering alternatives could be helpful for identifying groups of politicians that

are most similar in terms of voters' preferences when evaluating candidates in an election. This data can help political campaigns determine which groups of voters to target with their message. Furthermore, clustering alternatives can be used to identify outliers, which are politicians who are not similar to others, and their presence can be interpreted as an indication of heterogeneous opinions among voters.

Another potential application of clustering algorithms for preference-approvals is in the context of online product recommendations. Online retailers often use recommendation algorithms to suggest products to their customers based on their previous purchases and browsing history. However, the complexity of the preference-approval space could be a limit for these algorithms since it can make it difficult to identify meaningful patterns and make accurate recommendations. Therefore, by applying clustering algorithms to the preference-approval data, retailers can more effectively group products based on their similarity, leading to more accurate and relevant customer recommendations.

Although the literature on clustering algorithms applied to preference orderings is rich, it is not straightforward to transfer it directly to the preference-approval framework because preference-approvals are more complex structures. Clustering approaches for preference rankings can be applied to both individuals and alternatives. Most commonly used methods for clustering individuals involve an algorithmic model, such as hierarchical clustering, or an approach that aims to optimize a badness-of-fit function, such as K-means, PCA, MDS, or fuzzy clustering. Further details on these methods can be found in Heiser and D'Ambrosio (2013, pp. 19–31).

Despite being less studied, the task of clustering alternatives rather than individuals in preference rankings is undoubtedly relevant. Marden (1996) defined a distance between two alternatives as the squared Euclidean distance of the ranks assigned to them. Thus, objects will be close if the voters give them similar ranks. Finally, they applied a simple hierarchical clustering to find meaningful groups. Sciandra et al. (2020) proposed a projection pursuit-based clustering method to simultaneously identify clusters of both individuals and alternatives in preference rankings.

Similarly to the task of clustering alternatives in preference rankings, González del Pozo et al. (2017) focused on clustering alternatives in ordered qualitative scales. They designed an agglomerative hierarchical clustering algorithm, relying on the concept of ordinal proximity measure, to cluster nine US presidential candidates. The degree of consensus is measured by the proximity of all pairs of individual appraisals over the evaluated alternatives.

In this work, we introduce a new family of pseudometrics on the set of alternatives taking into account voters' opinions on these alternatives through preference-approvals. To obtain clusters, we apply an order-invariant partitioning algorithm, known as Ranked k -medoids (RKM), see Zadegan et al. (2013), taking the similarities as input among pairs of alternatives based on the proposed pseudometrics. Finally, clusters are represented in 2-dimensional space using non-metric multidimensional scaling. This paper is an extended version of the paper presented at the 51st Scientific Meeting of the Italian Statistical Society in June, 2022 (Albano et al. 2022).

The paper is organized as follows. Section 2 is devoted to introducing basic notation and concepts we use throughout the article. Section 3 contains our proposal for clustering alternatives. Section 4 includes some case studies in order to emphasize the advantage of reducing the complexity of the preference-approval space. Finally, Sect. 5 concludes the paper with some remarks.

2 Preliminaries

Let $X = \{x_1, \dots, x_n\}$ represent a finite set of alternatives, with $n \geq 2$. A full and transitive binary relation on X is a *weak order* (or *complete preorder*). While, a *linear order* on X is an antisymmetric weak order on X .

The set of weak and linear orders on X is denoted by $W(X)$ and $L(X)$, respectively. Given $R \in W(X)$, we represent the asymmetric and symmetric components of R with $>$ and \sim , respectively: $x_i > x_j$ if not $x_j R x_i$, and $x_i \sim x_j$ if $x_i R x_j$ and $x_j R x_i$.

Given a set Y , with $\mathcal{P}(Y)$ we denote its power set, i.e., $I \in \mathcal{P}(Y) \Leftrightarrow I \subseteq Y$. In turn, with $\#Y$ we denote the cardinality of Y .

2.1 Preference-approvals

Consider a scenario in which a group of voters $V = \{v_1, \dots, v_m\}$, with $m \geq 2$ have to declare their preferences on a set of alternatives $X = \{x_1, \dots, x_n\}$, with $n \geq 2$.

By splitting X into A , the set of *acceptable* alternatives, and $U = X \setminus A$, the set of *unacceptable* alternatives, where A and U can both be the empty set, we assume that each voter uses a weak order to rank the options in X and additionally determines whether each option is acceptable or unacceptable.

We also make the following consistency assumption: given two alternatives x_i and x_j , if x_j is acceptable and x_i is ranked above x_j , then x_i should be acceptable as well.

Definition 1 A *preference-approval* on X is a pair $(R, A) \in W(X) \times \mathcal{P}(X)$ satisfying the following condition:

$$\forall x_i, x_j \in X [(x_i R x_j \text{ and } x_j \in A) \Rightarrow x_i \in A].$$

With $\mathcal{R}(X)$ we denote the set of preference-approvals on X .

A *profile* is a vector of preference-approvals $[(R_1, A_1), \dots, (R_m, A_m)] \in \mathcal{R}(X)^m$, where (R_k, A_k) is the preference-approval of the voter $v_k \in V$.

Remark 1 If $(R, A) \in \mathcal{R}(X)$, then the following conditions are satisfied:

1. $\forall x_i, x_j \in X [(x_i \in A \text{ and } x_j \in U) \Rightarrow x_i > x_j]$.
2. $\forall x_i, x_j \in X [(x_i R x_j \text{ and } x_i \in U) \Rightarrow x_j \in U]$.

Example 1 Consider the preference-approval $(R, A) \in \mathcal{R}(\{x_1, x_2, x_3, x_4\})$ represented by

Table 1 Number of approvals, linear orders, weak orders and preference-approvals

n	Approvals	Linear orders	Weak orders	Preference-approvals
2	4	2	3	8
3	8	6	13	44
4	16	24	75	308
5	32	120	541	2612
6	64	720	4683	25,988
7	128	5040	47,293	296,564
8	256	40,320	545,835	3,816,548
9	512	362,880	7,087,261	54,667,412
10	1024	3,628,800	102,247,563	862,440,068

$$x_1$$

$$x_2 \quad x_3$$

$$x_4$$

The alternatives above the line are acceptable, i.e., $A = \{x_1\}$, and those below the line are unacceptable, i.e., $U = \{x_2, x_3, x_4\}$. This means that alternatives in the upper rows are preferred to those in the lower rows, and alternatives in the same row are indifferent.

The number of approvals, linear orders, weak orders, and preference-approvals when the number of alternatives is $n = 2, 3, \dots, 10$ are listed in Table 1. The total number of approvals (subsets of X) and linear orders is widely known to be $2^n!$ and $n!$, respectively. While, according to Good (1975) and Bailey (1998), there are $n!(\log_2 e)^{n+1}/2$ weak orders. Finally, the last column of Table 1 shows the exact number of preference-approvals (these data come from Albano et al. 2023).

Table 1 provides a comprehensive overview of the number of possible preferences and rankings that can be generated for a given number of alternatives. It allows to gain a better understanding of the combinatorial explosion that occurs as the number of alternatives increases. Indeed, the complexity of the preference-approval space poses a significant challenge in developing algorithms or models for preference aggregation and prediction.

2.2 A pseudometric on preferences

Positions are easily assigned to alternatives in linear orders: given $R \in L(X)$, the *position* of each alternative $x_i \in X$ in R is defined through the mapping $P_R : X \rightarrow \{1, \dots, n\}$ that gives the first choice a score of 1, the second alternative a score of 2, and so on.

In weak orders, the positions of the alternatives can be assigned in a variety of ways. One of them, employed by García-Lapresta and Pérez-Román (2011) is based on Smith (1973), Black (1976), and Cook and Seiford (1982). Given $R \in W(X)$, the position of $x_i \in X$ in R is determined by the mapping $P_R : X \rightarrow [1, n]$ defined as

$$P_R(x_i) = n - \#\{x_k \in X \mid x_i > x_k\} - \frac{1}{2} \cdot \#\{x_k \in X \setminus \{x_i\} \mid x_i \sim x_k\}, \quad (1)$$

that is, the position of x_i in R is determined by subtracting the number of alternatives to which x_i is strictly preferred (i.e., they appear after x_i in R) from n , the total number of alternatives. This value is then adjusted by subtracting half the number of alternatives that are tied with x_i (i.e., are indifferent to x_i). The resulting positions can be used to compare the rankings of alternatives across different weak orders. From Eq. (1), we introduce a pseudometric on the set of alternatives that measures the difference between the positions of two alternatives in a weak order.

Proposition 1 *Given $R \in W(X)$, the mapping $d_p : X \times X \rightarrow \mathbb{R}$ defined as*

$$d_p(x_i, x_j) = |P_R(x_i) - P_R(x_j)| \quad (2)$$

is a pseudometric on X , i.e., it satisfies the following conditions for all $x_i, x_j, x_k \in X$:

1. $d_p(x_i, x_j) \geq 0$.
2. $d_p(x_i, x_i) = 0$.
3. $d_p(x_i, x_j) = d_p(x_j, x_i)$.
4. $d_p(x_i, x_j) \leq d_p(x_i, x_k) + d_p(x_k, x_j)$.

Additionally, it is satisfied $d_p(x_i, x_j) = 0 \Leftrightarrow x_i \sim x_j$, for all $x_i, x_j \in X$.

Obviously, if $R \in L(X)$, then d_p is a metric, i.e., $d_p(x_i, x_j) = 0 \Leftrightarrow x_i = x_j$, for all $x_i, x_j \in X$. Note that $d_p(x_i, x_j) \in \{0, 1, \dots, n-1\}$ for all $x_i, x_j \in X$.

2.3 A pseudometric on approvals

Given $A \subseteq X$, the *indicator function* (or *characteristic function*) of A , $I_A : X \rightarrow \{0, 1\}$, is defined as

$$I_A(x_i) = \begin{cases} 1, & \text{if } x_i \in A, \\ 0, & \text{if } x_i \in X \setminus A. \end{cases} \quad (3)$$

From Eq. (3), we now introduce a pseudometric on the set of alternatives that measures the difference between the membership of two alternatives in a set.

Proposition 2 *Given $A \subseteq X$, the mapping $d_A : X \times X \rightarrow \mathbb{R}$ defined as*

$$d_A(x_i, x_j) = |I_A(x_i) - I_A(x_j)| \tag{4}$$

is a pseudometric on X , i.e., it satisfies the following conditions for all $x_i, x_j, x_k \in X$:

1. $d_A(x_i, x_j) \geq 0$.
2. $d_A(x_i, x_i) = 0$.
3. $d_A(x_i, x_j) = d_A(x_j, x_i)$.
4. $d_A(x_i, x_j) \leq d_A(x_i, x_k) + d_A(x_k, x_j)$.

Additionally, it is satisfied $d_A(x_i, x_j) = 0 \Leftrightarrow [x_i, x_j \in A \text{ or } x_i, x_j \notin A]$, for all $x_i, x_j \in X$.

Note that $d_A(x_i, x_j) \in \{0, 1\}$ for all $x_i, x_j \in X$.

Remark 2 Every preference-approval $(R, A) \in \mathcal{R}(X)$ can be codified in terms of $P_R(x_i)$ (Eq. 1) and $I_A(x_i)$ (Eq. 3) as follows:

$$[P_R(x_1), P_R(x_2), \dots, P_R(x_n)] \ \& \ [I_A(x_1), I_A(x_2), \dots, I_A(x_n)]. \tag{5}$$

For instance, in Example 1, (R, A) is codified as $(1, 2.5, 2.5, 4) \ \& \ (1, 0, 0, 0)$.

3 The proposal

Given a profile $[(R_1, A_1), \dots, (R_m, A_m)] \in \mathcal{R}(X)^m$ and two alternatives $x_i, x_j \in X$, we now present two indices that quantify the distance between these items in terms of preference and approvals, respectively, for each voter $v_k \in V$. They are based on the pseudometrics introduced in Eqs. (2) and (4).

3.1 Preference discordances

The *preference-discordance* between x_i and x_j for the voter $v_k \in V$ is defined as

$$p_{ij}^k = \frac{1}{n-1} \cdot |P_{R_k}(x_i) - P_{R_k}(x_j)|. \tag{6}$$

Note that $p_{ij}^k \in [0, 1]$.

Remark 3 Note that if a voter expresses a linear order $R \in L(X)$, then: (i) there will not be any pair of different alternatives whose preference-discordance is 0 and (ii) there will be only one pair of alternatives whose preference-discordance is maximum, equal to 1:

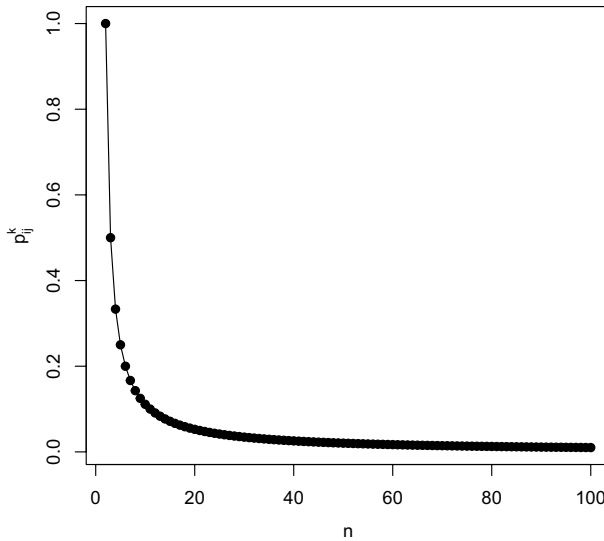


Fig. 1 Preference-discordance of two adjacent alternatives by n

$$R \in L(X) \Rightarrow \begin{cases} p_{ij}^k \neq 0 \text{ for all } x_i, x_j \in X, x_i \neq x_j, \\ \exists! x_i, x_j \in X \quad p_{ij}^k = 1. \end{cases}$$

On the contrary, if a voter expresses a weak order that is not a linear order $R' \in [W(X) \setminus L(X)]$, and indifference between different alternatives happens, then: i) there will exist at least a pair of different alternatives whose preference-discordance is 0; ii) no pair of alternatives produces a preference-discordance equal to 1:

$$R' \in [W(X) \setminus L(X)] \Rightarrow \begin{cases} \exists x_i, x_j \in X, x_i \neq x_j, \quad p_{ij}^k = 0, \\ p_{ij}^k \neq 1 \text{ for all } x_i, x_j \in X. \end{cases}$$

Remark 4 Note that, p_{ij}^k is decreasing as the total number of alternatives, n , increases. Figure 1 plots the preference-discordance $p_{ij}^k(x_i, x_k)$ as a function of n , where x_i, x_j are two adjacent alternatives for the k -th voter.

The total number of different alternatives, n , determines the expressivity of voters. Two alternatives $x_i, x_j \in X$ that are adjacent are considered more similar in a large order than in a small one. For example, consider the universe of weak orders for $n = 4$: x_i and x_j are adjacent in 40 out of the 75 possible scenarios, about 53%. On the contrary, when the number of alternatives doubles, $n = 8$, the number of weak orders in which x_i and x_j are adjacent drops to 170,440 out of 545,835, approximately 31%. As n increases, the percentage of scenarios in which x_i and x_j are adjacent decreases, and so does the average distance between them.

Finally, the *average preference-discordance*, \bar{p}_{ij} , summarizes the average dissimilarity between two alternatives according to the whole set of voters:

$$\bar{p}_{ij} = \frac{1}{m} \sum_{k=1}^m p_{ij}^k. \tag{7}$$

3.2 Approval discordances

The *approval-discordance* between x_i and x_j for the voter $v_k \in V$ is defined as

$$a_{ij}^k = |I_{A_k}(x_i) - I_{A_k}(x_j)|, \tag{8}$$

where $a_{ij}^k \in \{0, 1\}$.

Unlike p_{ij}^k , the approval-discordance is not influenced by the number of alternatives whose acceptability is established. Considering all possible approvals of n alternatives, the percentage of approval vectors in which x_i and x_j receive the same rating remains constant as n varies.

Finally, the *average approval-discordance*, \bar{a}_{ij} , summarizes the average dissimilarity between two alternatives according to the whole set of approvals:

$$\bar{a}_{ij} = \frac{1}{m} \sum_{k=1}^m a_{ij}^k. \tag{9}$$

3.3 Global discordances

In order to define an overall measure of discordance between each pair of alternatives, we consider the family of *weighted means*, $h : [0, 1] \times [0, 1] \rightarrow [0, 1]$, defined as

$$h(x, y) = \lambda \cdot x + (1 - \lambda) \cdot y, \tag{10}$$

where $\lambda \in [0, 1]$.

Taking into account the preference and approval discordances introduced in Eqs. (6), (7), (8) and (9), respectively, and the family of weighted means defined in Eq. (10), we now introduce a global measure of discordance between pairs of alternatives.

Definition 2 Given a profile $[(R_1, A_1), \dots, (R_m, A_m)] \in \mathcal{R}(X)^m$ and $\lambda \in [0, 1]$, the mapping $\delta_\lambda : X \times X \rightarrow [0, 1]$ is defined as

$$\delta_\lambda(x_i, x_j) = \frac{1}{m} \cdot \sum_{k=1}^m \left(\lambda \cdot p_{ij}^k + (1 - \lambda) \cdot a_{ij}^k \right) = \lambda \cdot \bar{p}_{ij} + (1 - \lambda) \cdot \bar{a}_{ij}. \tag{11}$$

Proposition 3 Given a profile $[(R_1, A_1), \dots, (R_m, A_m)] \in \mathcal{R}(X)^m$, the mapping δ_λ is a pseudometric on X for every $\lambda \in [0, 1]$. We say that δ_λ is the pseudometric associated with λ .

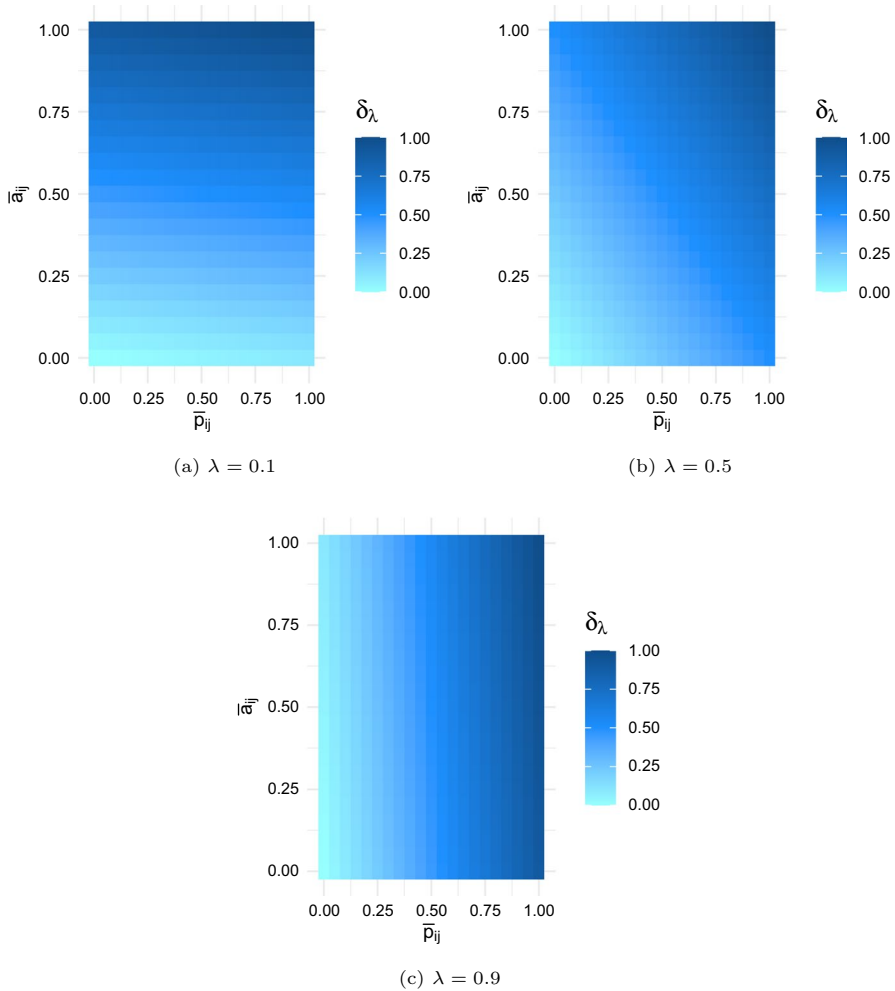


Fig. 2 Heatmaps δ_λ

Proof Taking into account Propositions 1 and 2, it is obvious that δ_λ satisfies the following conditions for all $x_i, x_j \in X$: $\delta_\lambda(x_i, x_j) \geq 0$, $\delta_\lambda(x_i, x_i) = 0$ and $\delta_\lambda(x_i, x_j) = \delta_\lambda(x_j, x_i)$. Finally, δ_λ satisfies the triangle inequality being a convex combination of pseudometrics. \square

Figure 2 shows how δ_λ varies as a function of \bar{p}_{ij} and \bar{a}_{ij} for $\lambda = 0.1, 0.5, 0.9$.

In Fig. 2a, λ is set to 0.5. Thus, \bar{p}_{ij} and \bar{a}_{ij} have the same weight in determining the final distance $\delta_\lambda(x_i, x_j)$. As a result, the corresponding heatmap is symmetrical with respect to the secondary diagonal, and δ_λ increases diagonally from bottom to top and from left to right.

On the contrary, $\lambda = 0.1$ Fig. 2a and $\lambda = 0.9$ (Fig. 2)c correspond to two unbalanced settings. Giving much more importance to approvals, $\lambda = 0.1$ Fig. 2a, causes the bottom area of the graph to contain lower distances, δ_λ grows much more noticeably vertically rather than horizontally. Finally, in Fig. 2c, δ_λ is dominated by the preference-discordance. The lesser distances are found on the left side of the graph, and δ_λ expands horizontally significantly more than vertically.

The choice of λ as a weighting parameter in metrics for preference-approvals has been the subject of debate in other scientific articles (Erdamar et al. 2014; Dong et al. 2021; Albano et al. 2023). There is not a λ value that is always the best choice for preference-approvals problems. Generally, as the relative importance of the two types of information is unknown, a recommended value of λ is 0.5, which assigns equal importance to both the ranking and approval components.

3.4 Clustering procedure and visualization

In this paper, we use the algorithm *Ranked k-medoids* (RKM) (see Zadegan et al. 2013) to find clusters but we highlight that our pseudometrics can be used jointly with any distance-based clustering algorithm.

The RKM technique employs a function that assigns a rank to alternatives based on how similar they are to each other, with the more similar alternatives receiving a lower rank. In other words, $\text{rank}(x_i, x_j) = l$ shows that x_j is the l -th similar alternative to x_i among n alternatives in the dataset. Sorting the similarity values between x_i and other items in the dataset allows one to determine the ranks of the remaining objects in relation to an item like x_i . A rank matrix is also expressed by the rank function $K = [k_{ij}]$, where $\text{rank}(x_i, x_j) = k_{ij}$ for all $x_i, x_j \in X$.

Note that, due to the fact that two items are rarely at the same rank as one another, K is not always symmetric. Thus, K is an $n \times n$ matrix that shows the *hostility* relationship among alternatives in the dataset.

The hostility value (hv) of a particular object, x_i , within a collection of alternatives, G , is introduced in order to identify the medoids. The hostility value, hv_i , of x_i within the set G is defined as:

$$hv_i = \sum_{x_j \in G} k_{ij}. \quad (12)$$

Starting from the similarities among pairs of objects based on $\delta_\lambda(x_i, x_j)$, the RKM algorithm firstly calculates K matrix and selects the medoids randomly. Then, for each medoid, select the group of the most similar objects to each medoid, using the sorted index matrix, and calculate the hostility values of every object in those groups using Eq. (12). Afterwards, select the object with the highest hostility value as the new medoid and move one of the medoids placed in the same group. Finally, iterate the process and assign each object to the most similar medoid.

This algorithm requires the number of clusters to be specified before. However, some methods, such as the Silhouette Coefficient, can be used to estimate the optimal number of clusters in our data.

The RKM method is particularly suitable in our case since it analyzes a ranking of dissimilarities, which makes the results order-invariant, meaning that data transformations that preserve the original order of the data have no impact on clusters.

In order to represent the resulting clusters in a 2-dimensional space, the multidimensional scaling (MDS) is employed. This class of methods attempts to express an observable proximity or distance matrix by a simple geometrical model or map so that the greater the perceived distance between two alternatives, the more apart the points representing them in the final geometrical model are.

Such models estimate q -dimensional coordinate values to represent n alternatives of a distance matrix. They optimize a chosen goodness of fit index, how closely the predicted distances approximate the observed ones. A number of optimization strategies, when combined with a variety of goodness of fit indices, result in a variety of MDS algorithms (Hothorn and Everitt 2006).

In this paper, given the nature of the objects, the Non-metric Multidimensional Scaling is employed. This method constructs fitted distances in the same rank order as the original distance, thus preserving the rank order of the proximities. Algorithms for accomplishing this are described in Kruskal (1964). The required coordinates for a given set of disparities are found by minimizing a function called *Stress* based on the squared differences between the observed proximities and the derived disparities. The process iterates until a suitably chosen convergence condition is satisfied.

4 Case studies

This section shows how the proposed metric can be used to perform cluster analysis on real data.

4.1 Eurobarometer dataset

Eurobarometer is a collection of cross-country public opinion surveys conducted on the authority of the European Commission and other European Union (EU) institutions since 1973. These surveys address a variety of topics pertaining to the EU and its member countries. Specifically, the data used in this paper come from survey question QA7 named “Public opinion in the European Union”.¹ Voters, divided up by country, were asked to indicate which of the values listed in Table 2 the EU meant the most to them.

As a result, data are stored in Table 11, which has 15 columns (each indicating an object of $X = \{x_1, \dots, x_{15}\}$) and 27 rows (one row for each EU member country). The table’s generic cell ij displays the total number of votes that the i -th country gives in favor of the j -th alternative.

The original table is transformed into a set of preference-approvals to perform the analysis. Following Albano et al. (2023), the alternatives are ranked in order of

¹ <https://europa.eu/eurobarometer/surveys/detail/2553>.

Table 2 Values in the EU

Alternatives	Names
x_1	Peace
x_2	Economic prosperity
x_3	Democracy
x_4	Social protection
x_5	Freedom to travel, study and work anywhere in the EU
x_6	Cultural diversity
x_7	Stronger say in the world
x_8	Euro
x_9	Unemployment
x_{10}	Bureaucracy
x_{11}	Waste of money
x_{12}	Loss of our cultural identity
x_{13}	More crime
x_{14}	Not enough control at external borders
x_{15}	Quality of life of future generations

Table 3 Votes in Italy

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
Italy	24	13	19	11	43	15	27	32	13	12	12	10	9	16	22

popularity, from the most to the least voted, and approvals are derived by approving the alternatives that obtained a higher number of votes than the national average. For instance, Table 3 displays the votes cast in Italy (Table 11 contains the votes cast in all countries).

Following Eq. (5) and considering that the average vote in Italy is 18.53, the votes in Italy are converted into a preference-approval as:

$$(4, 9.5, 6, 13, 1, 8, 3, 2, 9.5, 11.5, 11.5, 14, 15, 7, 5) \& (1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1)$$

that can be visualized as follows

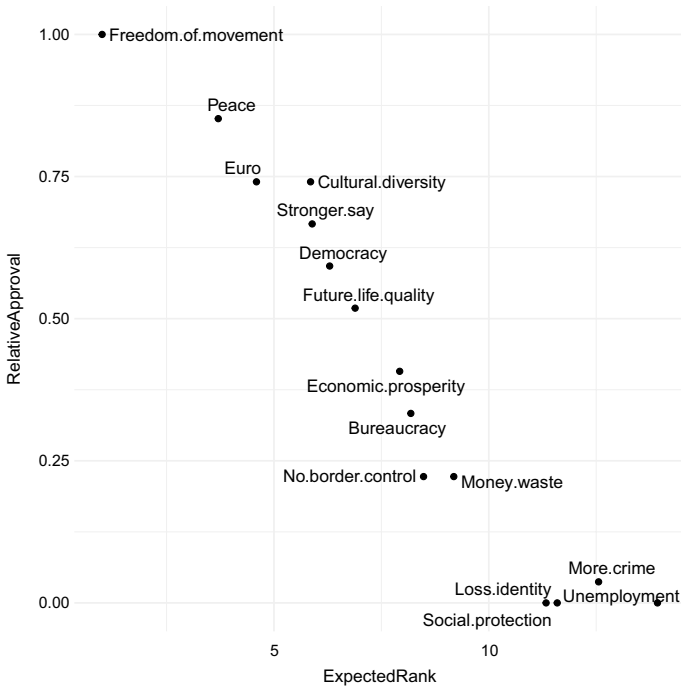


Fig. 3 Preference-approval plane, Eurobarometer

x_5
 x_8
 x_7
 x_1
 x_{15}
 x_3

 x_{14}
 x_6
 x_2 x_9
 x_{10} x_{11}
 x_4
 x_{12}
 x_{13}

In Fig. 3, the 15 alternatives are arranged on the preference-approval plane. The location of each alternative in this 2-dimensional space is identified by its ExpectedRank (i.e., the average rank over the whole set of voters) and by its RelativeApproval, i.e., the relative frequency of voters who considered it acceptable.

The preference-approval plane provides a summary of the evaluations of voters on average. In particular, it reveals that all voters consider “Freedom of movement” the

Table 4 ExpectedRank and RelativeApproval

Alternative	ExpectedRank	RelativeApproval
x_1	2.5	0.5
x_2	2	1
x_3	3	0
x_4	2.5	0.5

Table 5 Distances $\delta_{0.5}$

	x_1	x_2	x_3	x_4
x_1	0			
x_2	0.5	0		
x_3	0.5	0.67	0	
x_4	1	0.5	0.5	0

best alternative: it is unanimously approved and always placed first in the preference-approvals; its RelativeApproval and its ExpectedRank are equal to 1. The other alternatives tend to lie in a straight line with a negative angular coefficient. The further we move away from the point (1, 1), the worse the corresponding alternatives obtained average ratings.

Note that the preference-approval plane aids the interpretation of clusters once they have been estimated. However, it should not be considered a tool to identify clusters since the distance between points in the preference-approval plane does not necessarily reflect the pseudometric in Eq. (11). Alternatives having similar average ranking positions and approvals may show discordance among the voters.

Example 2 To further clarify this concept, let us consider $(R_1, A_1), (R_2, A_2) \in \mathcal{R}(\{x_1, x_2, x_3, x_4\})$ the following preference-approvals:

$$(R_1, A_1) \equiv \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \quad (R_2, A_2) \equiv \begin{array}{c} x_4 \\ x_2 \\ x_3 \\ x_1 \end{array}$$

For each alternative $x_i \in X$, the ExpectedRank, expected approval and $\delta_{0.5}$ distance matrix are reported in Tables 4 and 5.

Note that x_1 and x_4 have the same RelativeApproval and ExpectedRank, thus identical coordinates in the preference-approval plane, but show maximum discordance over the voters, i.e., $\delta_{0.5}(x_1, x_4) = 1$. In fact, they are placed at the opposite extremes in both preference-approvals. Therefore, the preference-approval

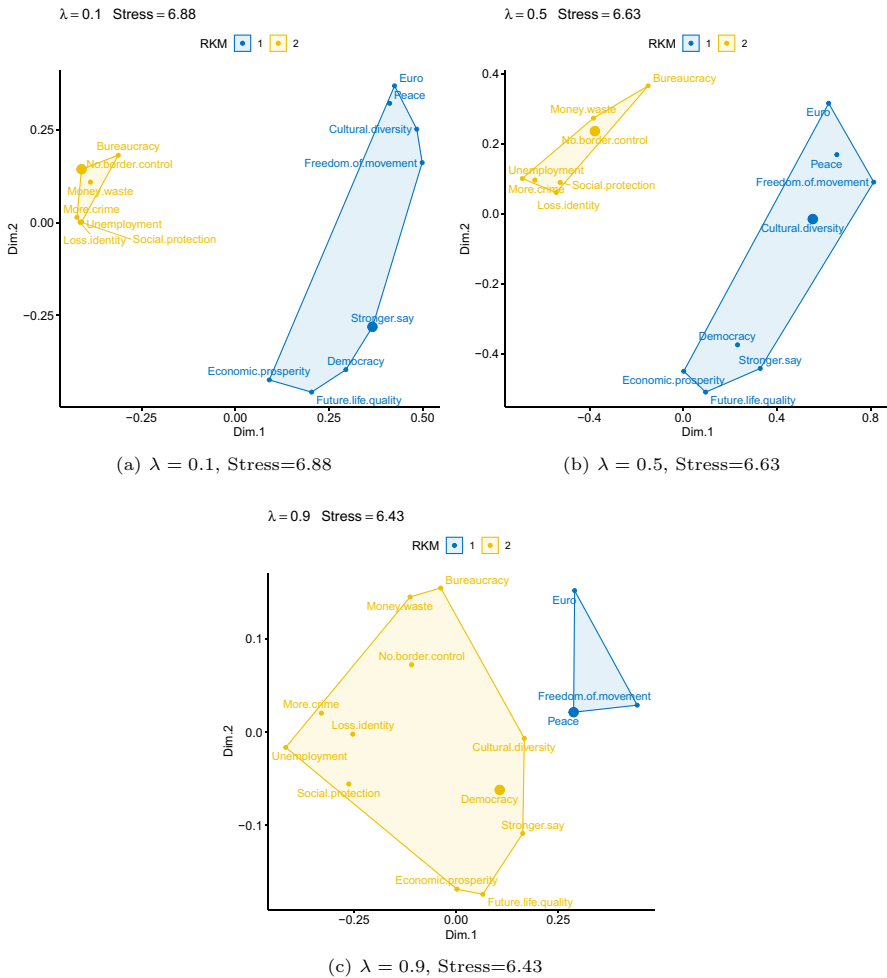


Fig. 4 Graphical representation of RKM clusters

plane is intended to be an interpretative tool to visualize average judgments and interpret clusters once they have been estimated. At the same time, it is not appropriate to identify clusters since it does not reflect similarities among elements.

Figure 4 shows the clusters estimated by the RKM algorithm, where the central medoid for each cluster is highlighted through the dimension of the point. We investigate the effect of the λ parameter on the output, by setting $\lambda = 0.1, 0.5, 0.9$. In this way, we are able to study three scenarios: $\lambda = 0.5$, which corresponds to giving the same importance to approvals and preferences, and $\lambda = 0.1, 0.9$, which corresponds to the opposite unbalanced situations.

We also show the Stress values in each scenario to assess the goodness of the graphical representation obtained with the MDS. Note that the position of the points in the new space found by MDS depends on the value of λ . If the parameter, λ varies, the graphical representation does as well.

In general, the Stress coefficient varies between 6.88 and 6.43, showing a good adaptation that tends to improve slightly as λ increases. The optimal number of clusters, chosen through the Silhouette criterion, is two independently from the λ value.

When $\lambda = 0.1, 0.5$, the clusters found are the same, but the degree of separation between them clearly changes. In fact, the two clusters exhibit a higher separation index² under $\lambda = 0.1$, i.e., assigning much more weight to the approvals than under $\lambda = 0.5$. Thus, a clear division is obtained between frequently accepted and not accepted alternatives. The two clusters become closer as λ reaches 0.5. In this example, there are clearly two different types of alternatives: those referring to negative aspects (“Bureaucracy”, “Unemployment”, “Money waste”, etc.) and those referring to positive aspects (“Freedom”, “Democracy”, etc.). For this reason, a voter with a bad opinion about the EU will prefer the former and vice versa. Indeed, the two clusters are robust and remain unchanged for small and moderate values of λ . In this sense, clusters also provide a measure of the consistency of voters’ judgment when alternatives can be divided into natural groups. In this instance, when $\lambda = 0.1, 0.5$, the identified clusters split options related to negative attributes from those related to positive qualities. If the clusters were a mixture of good and bad options, it would imply low consistency among the judges.

Note that the proximity between points in the two-dimensional space discovered by the MDS (Fig. 4a–c) reflects the similarities based on δ_λ , between the alternatives over the voters. Thus, the position of the elements in this new space addresses the cluster interpretation.

Indeed, although in the preference-approval plane (see Fig. 3), “Money waste” is closer to the alternatives belonging to Cluster 1, its position in the MDS space reveals that actually, it is part of Cluster 2.

Figure 4c displays clusters under $\lambda = 0.9$, i.e., unbalanced towards preferences. In this case, Cluster 1 isolates the three alternatives frequently placed in the first positions (see the preference-approval plane Fig. 3), namely: “Freedom”, “Peace” and “Euro”.

To better understand how these can be used, consider a policymaker seeking to design a campaign to improve the EU’s public image. By analyzing the clusters and the responses within each cluster, the policymaker can tailor the campaign message to better resonate with the target audience. For instance, the campaign could emphasize the positive aspects of the EU, such as “Freedom of movement”, “Peace” and “Democracy”, to appeal to those with a positive view of the EU. Conversely, the campaign could address negative aspects, such as “Bureaucracy”, “Money waste” and “Loss of our cultural identity” to appeal to those with a negative view of the EU. As a matter of fact, the clusters can inform policymakers and political parties about the values and concerns that are most important to voters in different countries. This allows a geographical type analysis to describe EU preferences.

² Based on the distances for every point to the closest point not in the same cluster.

Table 6 Average ranking position and approval rating for items in Cluster 1 and Cluster 2 for each country

Country	Cluster 1		Cluster 2	
	Average rankings	Average approvals	Average rankings	Average approvals
Belgium	5.00	0.62	11.43	0.14
Bulgaria	5.00	0.50	10.86	0.00
Czech Rep	5.88	0.75	10.14	0.29
Denmark	5.12	0.75	11.00	0.14
Germany	5.38	0.75	11.00	0.14
Estonia	5.62	0.62	9.86	0.14
Ireland	4.38	1.00	11.86	0.00
Greece	6.38	0.62	9.71	0.14
Spain	4.62	0.75	11.57	0.00
France	5.88	0.50	10.29	0.29
Croatia	4.25	1.00	12.00	0.00
Italy	4.75	0.75	11.43	0.00
Cyprus	6.38	0.50	9.43	0.29
Latvia	5.38	0.75	11.00	0.00
Lithuania	5.00	0.62	10.71	0.00
Luxembourg	5.50	0.62	10.57	0.00
Hungary	4.75	0.75	11.29	0.14
Malta	4.50	0.88	11.86	0.00
Netherlands	4.75	0.88	11.57	0.14
Austria	6.75	0.38	8.86	0.43
Poland	4.88	0.75	11.00	0.00
Portugal	4.38	0.88	11.43	0.00
Romania	4.38	0.62	11.14	0.00
Slovenia	4.38	0.88	11.71	0.00
Slovakia	7.12	0.50	8.43	0.43
Finland	5.88	0.50	10.00	0.14
Sweden	6.00	0.50	10.14	0.29

Table 6 displays the average ranking positions and the average approval ratings given by each country to the alternatives in the two clusters (identified using $\lambda = 0.5$). As an example, let's consider Belgium's rankings and approval ratings for Cluster 1 and Cluster 2. In Cluster 1, the alternatives $\{x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_{15}\}$ received the following ranking positions (3, 8, 7, 1, 6, 4, 2, 9). The approval ratings for these alternatives are: (1, 0, 0, 1, 1, 1, 1, 0). Therefore, the average ranking for Cluster 1 alternatives is 5 and the average approval rating is 0.62. On the other hand, for Cluster 2, the alternatives $\{x_4, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}$ received the following ranking positions (12, 15, 11, 5, 13, 14, 10). The approval ratings for

Table 7 Lines of action

Alternatives	Names
x_1	Searching for life and planets that could support life
x_2	Searching for raw materials and natural resources that could be used on Earth
x_3	Conducting basic scientific research to increase knowledge and understanding of space
x_4	Developing technologies that could be adapted for uses other than space exploration
x_5	Monitoring asteroids and other objects that could potentially hit the Earth
x_6	Monitoring key parts of the Earth's climate system
x_7	Sending human astronauts to explore the Moon
x_8	Sending human astronauts to explore Mars
x_9	Conducting scientific research on how space travel affects human health

these alternatives are (0, 0, 0, 1, 0, 0, 0). The average ranking for Cluster 2 is 11.43, while the average approval rating is only 0.14. This indicates that Belgium prefers the alternatives in Cluster 1.

Table 6 shows that Cluster 1 exhibited overall better ranking positions and higher approval ratings compared to Cluster 2, indicating more positive evaluations of its alternatives. However, it should be noted that some countries, such as Austria and Slovakia, assign better ranking positions and higher approval than the other countries to Cluster 2 items, which could indicate that negative aspects of the EU, such as “No border control” and “Money waste”, may be of particular concern to people in these countries. Thus, they may require more targeted and nuanced messaging that addresses specific concerns or criticisms that they have about the EU. Understanding these specific concerns can help policymakers craft messages that resonate with these groups and ultimately improve their overall perception of the EU. In this sense, these two countries can be considered outliers compared to the rest of the EU countries.

On the other hand, countries such as Croatia, Ireland, Portugal, and Slovenia assign particularly high approval ratings and good ranking positions to alternatives in Cluster 1. This could indicate that positive aspects of the EU, such as “Freedom of movement” and “Peace”, may resonate exceptionally well with the people of these countries.

Furthermore, the clusters can identify potential areas of disagreement or conflict among EU member countries. Policymakers can address these differences and find common ground by understanding the values and concerns that are most important to voters in different countries. For example, countries with significantly different ratings between clusters, such as Austria, which gave the highest rating in Cluster 2, and Slovakia, which gave the highest rating in Cluster 1, may have opposing ideologies and concerns that could lead to conflicts. In contrast, countries with comparable ratings across clusters, like Finland and Sweden, might have more common values and issues, making cooperation easier.

Table 8 Linguistic terms

	Linguistic term
l_1	Top priority
l_2	Important but lower priority
l_3	Not too important
l_4	Should not be done
l_5	No answer

Table 9 Pew Research Center example

Respondent	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
v_{10}	l_4	l_2	l_1	l_1	l_4	l_3	l_4	l_4	l_2

Overall, the clusters identified through the Eurobarometer dataset have practical applications for policymakers, political parties, and anyone interested in understanding public opinion within the EU.

4.2 Pew Research Center dataset

The Pew Research Center is a research institute that specializes in data-driven social science research, including public opinion surveys, demographic studies, content analysis, and more.

In this analysis, the survey “American Trends Panel Wave 33”³ is considered. Data in this report is drawn from the panel wave conducted from March 27 to April 9, 2018, to collect the opinions of United States citizens regarding the space agency NASA.

In this analysis, we focus specifically on a query in which a total of 2541 respondents were asked to assess how much priority NASA should give to a list of nine lines of action, listed in Table 7. Individuals employed the linguistic terms from the qualitative scale in Table 8 to accomplish this.

In order to remove neutral answers, the respondents giving at least a “No answer” responses were excluded, i.e., about 3% of the total sample size. Furthermore, for each respondent, alternatives were arranged into a preference-approval. The two linguistic terms l_1 and l_2 were used to indicate an acceptable alternative. An example is provided in Table 9.

The respondent v_{10} preference-approval (see Eq. 5) is

$$(7.5, 3.5, 1.5, 1.5, 7.5, 5, 7.5, 7.5, 3.5) \& (0, 1, 1, 1, 0, 0, 0, 0, 1)$$

that can be visualized as follows

³ <https://www.pewresearch.org/science/dataset/american-trends-panel-wave-33/>.

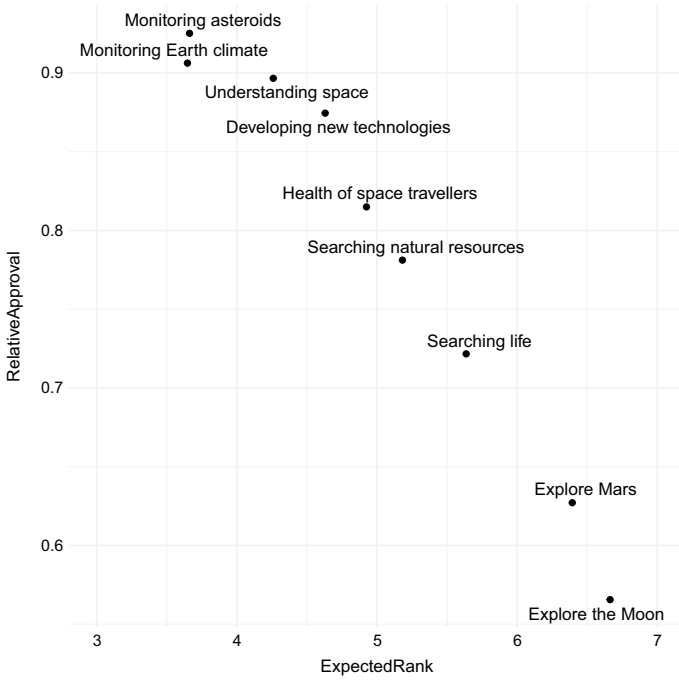


Fig. 5 Preference-approval plane, Pew Research Center

$$\begin{array}{cc}
 x_3 & x_4 \\
 x_2 & x_9 \\
 \hline
 & x_6 \\
 x_1 & x_5 & x_7 & x_8
 \end{array}$$

Here, approvals are generated directly by the voters through linguistic terms so that each voter can define all alternatives as acceptable or vice versa.

Figure 5 shows the nine alternatives on the preference-approval plane.

In this example, the RelativeApproval of each item ranges between 55 and 95%, meaning that each alternative has been considered as acceptable by more than half of the individuals. Therefore, although the alternatives may be regarded as acceptable by voters on average, less urgent alternatives, such as exploration of other planets and satellites (Moon and Mars) and more urgent alternatives, such as earth monitoring (Climate and Asteroids), can be identified.

The clusters estimated by the RKM algorithm are shown in Fig. 6. As in the previous example, three different values of $\lambda = 0.1, 0.5, 0.9$ are used. For each scenario, the clusters, medoids and Stress values reached by the MDS are illustrated for graphical representation.

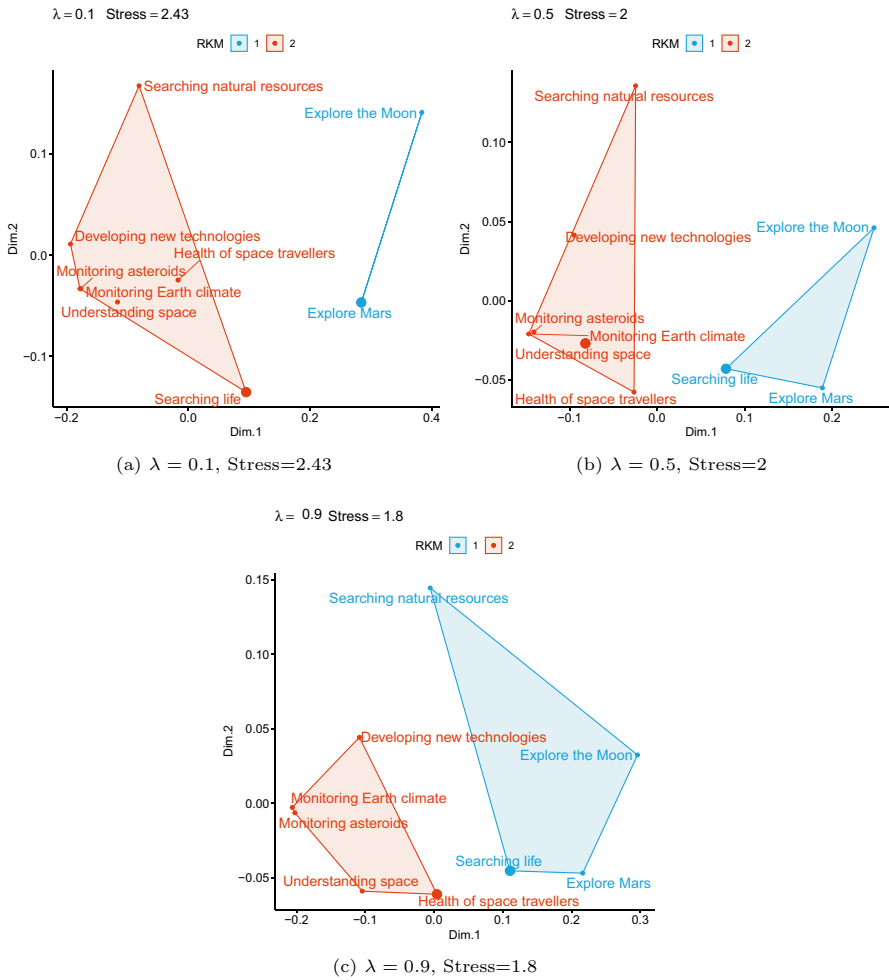


Fig. 6 Graphical representation of RKM clusters

Table 10 Comparison of two Clusters Based on ExpectedRank, RelativeApproval, and average distance

	Average ExpectedRank	Average Relative-Approval	Average within distance
Cluster 1	4.46	0.86	0.19
Cluster 2	6.08	0.66	0.20

In general, the stress coefficient varies between 2.43 and 1.8, showing an excellent adaptation that tends to improve as λ increases. The optimal number of clusters, chosen through the Silhouette criterion, turns out to be two independently from the λ value.

The effect of the parameter λ on cluster building is visible. Setting $\lambda = 0.1$ results in Cluster 1 including only the two alternatives related to space exploration (Moon and Mars), which have the lowest RelativeApproval (see Fig. 5). Voters strongly tend to attribute the same approvals to these two alternatives.

Increasing the value of $\lambda = 0.5$ causes Cluster 1 to enlarge by including the alternative “Searching life”. Finally, giving much more weight to the rankings, i.e. $\lambda = 0.9$, results in Cluster 1 also including the alternative Searching natural resources. In this way, Cluster 1 contains the four alternatives that are most frequently placed in the last positions in voters’ preference-approvals.

To use the clusters obtained in the Pew Research Center dataset in practice, one could identify which lines of action have similar approval profiles and use this information to inform policymakers. For instance, in this case study, we can state that the fact that the three alternatives x_1 (Search for life and planets that could support life) x_2 (Explore the Moon) and x_3 (Explore Mars) belong to the same cluster will suggest that policies focused on one of these areas will also receive the consent of voters who have a preference for the alternatives linked to it. Therefore, solutions shared by the three lines could significantly reduce the economic resources otherwise necessary for interventions on the individual dimensions. Moreover, by clustering the lines of action based on preference-approval profiles, policymakers can gain insights into the relationships between different policy options and the values and priorities of the electorate. These insights can inform the development of policies that align more closely with public sentiment and are, therefore, more likely to be successful. Table 10 shows the average ExpectedRank, RelativeApproval and within distance of the alternatives in the two clusters, identified using $\lambda = 0.5$.

The table shows that Cluster 1 alternatives have a higher average RelativeApproval (0.86) and a better average ExpectedRank (4.46) than Cluster 2. This means that, on average, individuals were more supportive of Cluster 1’s than Cluster 2’s activities. Furthermore, the average within-cluster distance is the same in both groups even though only three action lines are in the second cluster $\{x_1, x_7, x_8\}$ and six action lines $\{x_2, x_3, x_4, x_5, x_6, x_9\}$ belong to the first cluster. From a practical point of view, this means that investments in terms of time, money, and human resources aimed at one or more lines of action in Cluster 1 are likely succeed in attracting the support of all those who have expressed a preference for one or more of these actions, appealing a wider range of citizens as a result. Therefore, the action lines in Cluster 1 shall have a higher priority than those in Cluster 2.

It is worth noting that, in contrast to the first case study, the preference-approvals in this study were obtained using a different method. Instead of ranking the alternatives in order of popularity and approving those that received a higher number of votes than the national average, individuals employed a qualitative scale consisting of five linguistic terms to indicate their level of approval for each alternative. Consequently, due to the qualitative nature of this scale, there were significantly more approved alternatives in this study than in the first case study.

As a result, the ranking of alternatives played a more critical role in evaluating the distance between alternatives, and the impact of λ on the clustering process is more pronounced.

5 Concluding remarks

Preference-approvals structures are gaining increasing attention in social choice as they allow decision-makers to describe their preferences using more flexible and intuitive ordinal information. In this paper, we propose a new method for clustering alternatives in preference-approvals. First, we introduce a family of pseudometrics, δ_λ , able to quantify the distance between alternatives based on two main components: the *preference-discordance* p_{ij} and the *approval-discordance* a_{ij} , and on the λ parameter, which regulates the weight to give to each component.

To obtain clusters, we apply the Ranked k -medoids partitioning algorithm, taking as input the similarities among pairs of alternatives based on the proposed pseudometrics. Finally, clusters are represented in 2-dimensional space using Non-Metric Multidimensional Scaling.

Through two applications to real data, we demonstrate how our algorithm allows dividing a heterogeneous population of alternatives into homogeneous groups, reducing the complexity of the preference-approval space and providing a more accessible interpretation of data. We also show the effect of the λ parameter on cluster identification and visualization.

Future research should consider using the proposed clustering method to collapse categories in the context of multiple-choice models. Moreover, it will be important that future research investigate a method to identify simultaneous clusters of both individuals and alternatives in the preference-approval framework, extracting helpful information in a low-dimensional subspace. In the future, we will certainly consider any relevant alternatives that may arise in the context of preference-approval clustering and include a comparison with them.

Appendix

See Table 11.

Table 11 Votes in the EU

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
Belgium	32	22	23	14	52	24	27	42	7	18	25	12	11	19	20
Bulgaria	17	15	17	10	57	20	18	11	7	14	14	17	9	12	24
Czech Republic	33	30	31	8	66	18	32	17	2	37	28	18	7	20	30
Denmark	50	31	35	15	56	27	32	15	4	32	13	11	10	20	22
Germany	51	25	37	12	61	36	31	46	7	33	29	11	20	26	21
Estonia	30	19	28	13	76	29	18	50	2	34	19	19	4	19	22
Ireland	30	28	22	15	57	25	26	35	4	12	5	7	5	8	25
Greece	45	14	25	18	66	35	45	48	30	17	20	28	23	36	19
Spain	16	22	18	14	47	24	20	27	6	16	11	3	3	9	15
France	37	10	19	12	46	31	20	37	9	18	30	14	8	24	15
Croatia	26	27	23	20	55	24	26	23	4	11	12	18	9	13	39
Italy	24	13	19	11	43	15	27	32	13	12	12	10	9	16	22
Cyprus	38	20	26	24	67	36	26	45	31	21	22	24	35	36	25
Latvia	24	22	16	15	60	20	10	31	6	19	17	14	4	12	25
Lithuania	33	17	20	18	75	32	25	18	3	18	20	13	4	17	31
Luxembourg	46	23	34	17	61	36	22	51	6	26	27	12	18	26	22
Hungary	21	18	23	13	48	24	20	15	6	15	8	10	10	20	26
Malta	22	29	27	21	56	24	34	31	3	16	10	16	5	17	32
Netherlands	53	41	28	10	67	30	37	47	3	31	17	11	9	17	33
Austria	39	25	30	25	58	26	32	50	24	34	37	25	31	38	26
Poland	26	19	29	10	47	16	23	10	5	12	12	11	7	11	29
Portugal	19	27	24	18	57	27	32	44	6	6	7	7	6	16	22
Romania	23	19	21	14	45	16	16	23	9	15	14	18	15	16	21
Slovenia	36	25	27	16	54	27	20	47	5	18	15	13	15	16	24
Slovakia	28	20	16	10	64	21	28	44	12	28	33	20	22	38	19
Finland	32	18	24	6	67	27	21	54	3	35	24	11	11	22	18
Sweden	46	19	34	12	67	25	44	14	4	40	29	10	17	23	19

Acknowledgements The authors would like to express their gratitude to two anonymous reviewers for their valuable feedback and suggestions, and also to the Spanish Agencia Estatal de Investigación (Project PID2021-122506NB-I00) and the University of Palermo (Projects: FFR_D16_PLAIA and FFR_D16_SCIANDRA) for their financial support.

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

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

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

- Albano A, Sciandra M, Plaia A (2022) Towards the definition of distance measures in the preference-approval structures. *SIS 2022 book of short papers*
- Albano A, García-Lapresta JL, Plaia A, Sciandra M (2023) A family of distances between preference-approvals. *Ann Oper Res* 323:1–29. <https://doi.org/10.1007/s10479-022-05008-4>
- Bailey RW (1998) The number of weak orderings of a finite set. *Soc Choice Welf* 15(4):559–562
- Barokas G (2022a) Majority-approval social choice. *J Math Psychol* 109:102694
- Barokas G (2022b) Revealed desirability: a novel instrument for social welfare. *Theory Decis* 93(4):649–661
- Barokas G, Sprumont Y (2022) The broken Borda rule and other refinements of approval ranking. *Soc Choice Welf* 58(1):187–199
- Black D (1976) Partial justification of the Borda count. *Public Choice* 28(1):1–15. <https://doi.org/10.1007/BF01718454>
- Brams SJ (2008) Mathematics and democracy: designing better voting and fair-division procedures. *Math Comput Model* 48(9):1666–1670. <https://doi.org/10.1016/j.mcm.2008.05.013>
- Brams SJ, Sanver MR (2009) *Voting systems that combine approval and preference*. Springer, Berlin, pp 215–237
- Cook WD, Seiford LM (1982) On the Borda–Kendall consensus method for priority ranking problems. *Manag Sci* 28(6):621–637
- Dong Y, Li Y, He Y, Chen X (2021) Preference-approval structures in group decision making: axiomatic distance and aggregation. *Decis Anal* 18(4):273–295
- Erdamar B, García-Lapresta JL, Pérez-Román D, Sanver MR (2014) Measuring consensus in a preference-approval context. *Inf Fusion* 17:14–21
- Everitt BS, Landau S, Leese M, Stahl D (2011) *Cluster analysis*, 5th edn. Wiley, Hoboken
- García-Lapresta JL, Pérez-Román D (2011) Measuring consensus in weak orders. In: Herrera-Viedma E, García-Lapresta JL, Kacprzyk J, Fedrizzi M, Nurmi H, Zadrozny S (eds) *Consensual processes*. Springer, Berlin, pp 213–234. https://doi.org/10.1007/978-3-642-20533-0_13
- González del Pozo R, García-Lapresta JL, Pérez-Román D (2017) Clustering US 2016 presidential candidates through linguistic appraisals. In: Kacprzyk J, Szmidt E, Zadrozny S, Atanassov K, Krawczak M (eds) *Advances in fuzzy logic and technology 2017*. Springer, Cham, pp 143–153
- Good IJ (1975) The number of orderings of n candidates when ties are permitted. *Fibonacci Q* 13:11–18
- Heiser WJ, D'Ambrosio A (2013) Clustering and prediction of rankings within a Kemeny distance framework. In: Lausen B, Van den Poel D, Ultsch A (eds) *Algorithms from and for nature and life*. Springer, Cham, pp 19–31
- Hothorn T, Everitt BS (2006) *A handbook of statistical analyses using R*. Chapman and Hall/CRC, Boca Raton
- Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM Comput Surv (CSUR)* 31(3):264–323
- Kruger J, Sanver MR (2021) An Arrowian impossibility in combining ranking and evaluation. *Soc Choice Welf* 57:535–555
- Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29(2):115–129
- Liang H, Xiong W, Dong Y (2018) A prospect theory-based method for fusing the individual preference-approval structures in group decision making. *Comput Ind Eng* 117:237–248. <https://doi.org/10.1016/j.cie.2018.01.001>
- Liu H, Xu Z, Jiang L, Zhu J (2023) Multi-criteria group decision making with preference approval structures: a personalized individual semantics approach. *Inf Fusion* 96:80–91. <https://doi.org/10.1016/j.inffus.2023.03.009>

Marden JI (1996) Analyzing and modeling rank data. CRC Press, Boca Raton

Sanver MR (2010) Approval as an intrinsic part of preference. In: Laslier JF, Sanver MR (eds) Handbook on approval voting. Studies in choice and welfare. Springer, Berlin, pp 469–481. <https://doi.org/10.1007/978-3-642-02839-7>

Sciandra M, d'Ambrosio A, Plaia A (2020) Projection clustering unfolding: a new algorithm for clustering individuals or items in a preference matrix. *Data Anal Appl 3 Comput Classif Financ Stat Stochas Methods* 5:215–230

Smith JH (1973) Aggregation of preferences with variable electorate. *Econometrica* 41(6):1027–1041

Zadegan SMR, Mirzaie M, Sadoughi F (2013) Ranked k-medoids: a fast and accurate rank-based partitioning algorithm for clustering large datasets. *Knowl-Based Syst* 39:133–143

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