Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v14n1p117

Element weighted Kemeny distance for ranking data

By Albano, Plaia

Published: 20 May 2021

Electronic Journal of Applied Statistical Analysis
Vol. 14, Issue 01, May 2021, 117-145
DOI: 10.1285/i20705948v14n1p117

Element weighted Kemeny distance for ranking data

Alessandro Albano*and Antonella Plaia

Department of Economics, Business and Statistics - University of Palermo, Via delle Scienze, 90128. Palermo. Italy

Published: 20 May 2021

Preference data are a particular type of ranking data that arise when several individuals express their preferences over a finite set of items. Within this framework, the main issue concerns the aggregation of the preferences to identify a compromise or a "consensus", defined as the closest ranking (i.e. with the minimum distance or maximum correlation) to the whole set of preferences. Many approaches have been proposed, but they are not sensitive to the importance of items: i.e. changing the rank of a highly-relevant element should result in a higher penalty than changing the rank of a negligible one. The goal of this paper is to investigate the consensus between rankings taking into account the importance of items (element weights). For this purpose, we present: i) an element weighted rank correlation coefficient as an extension of the Emond and Mason's one, and ii) an element weighted rank distance as an extension of the Kemeny distance. The one-to-one correspondence between the weighted distance and the rank correlation coefficient is analytically proved. Moreover, a procedure to obtain the consensus ranking among several individuals is described and its performance is studied both by simulation and by the application to real datasets.

keywords: Weighted rank correlation coefficient, weighted Kemeny distance, element weights, consensus ranking.

1 Introduction

Ranking is one of the most effective cognitive processes used by people to handle many aspects of their lives. It is also a simple and efficient data collection technique to understand individuals' perception and preferences for some items. When some subjects

*Corresponding author: alessandro.albano@unipa.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

are asked to indicate their preferences over a set of alternatives, ranking data are called preference data. Therefore, preference data arise when a group of n individuals (e.g. judges, experts, voters, raters) express their preferences for a finite set of items (m different alternatives of objects, e.g., movies, activities, wines). Preference data can be expressed in two forms: by ordering the items (when alternatives are placed in order from best to worst), or rankings (when alternatives are fixed in any pre-specified order and preferences are expressed by using integers to indicate the rank of each alternative). If the m items, labelled $\{1, ...m\}$, are ranked in m distinguishable ranks, a complete (full) ranking or linear ordering is achieved (Cook, 2006): this ranking π is a mapping function from the set of items $\{1, ..., m\}$ to the set of ranks $\{1, ..., m\}$, endowed with the natural ordering of integers; $\pi = (\pi(1), \pi(2), ..., \pi(m))$ where $\pi(i)$ is the rank given by a judge to item i. For example, given 5 items, say $\{i_1, i_2, i_3, i_4, i_5\}$, the ordering $O = (i_2 \succ i_3 \succ i_4 \succ i_1 \succ i_5)$ corresponds to the ranking $\pi(O) = (4, 1, 2, 3, 5)$. Ranking π is, in this case, one of the 5! (or $m!$ with m items) possible permutations of 5 elements. When some items receive the same preference, then a tied ranking or a weak ordering is obtained. For example, given 5 items, say $\{i_1, i_2, i_3, i_4, i_5\}$, the ordering $O_2 = (i_2 \succ i_1 \sim i_3 \succ i_4 \succ i_5)$, where the judge likes i_1 and i_3 equally well (i.e. the items are tied), corresponds to the ranking $\pi(O_2) = (2, 1, 2, 4, 5)$. Finally, in real situations, sometimes not all items are ranked: we observe partial rankings when judges are asked to rank only a subset of the whole set of items (for example only $m - 1$ items), and incomplete rankings when judges can freely choose to rank only some items. Since every ordering can be transformed into a ranking and vice-versa, in this paper the two words are used interchangeably.

Because of their data reduction properties and ease of acquisition and representation, rankings have gained significant attention in the past few years. Within this framework, the interest lies in evaluating the distance and the correlation between two rankings. The most famous correlation measures between rankings include Kendall's and Spearman's rank correlation coefficient. As regards the distances, in 1962 Kemeny and Snell introduced a metric defined on linear and weak orders, known as Kemeny distance (or metric), later generalized to the framework of partial orders by Cook et al. in 1986, which satisfies the constraints of a distance measure suitable for rankings. Among the several axiomatic approaches proposed in the literature, here we consider the Kemeny's axiomatic framework (Kemeny and Snell, 1962) and, since we consider the possibility of ties, we assume that the geometrical space of preference rankings is the generalized permutation polytope (Heiser and D'Ambrosio 2013, D'Ambrosio et al. 2017), for which the natural distance measure is the Kemeny distance. Cook (2006) highlights the difficulties to treat the Kemeny metric, an issue already underlined by Emond and Mason (2002) and connected to the mathematical formulation using absolute values (see Eq.(6)). For this reason, the latter introduced a new correlation coefficient, strictly related to the Kemeny distance, and proposed the use of this coefficient as a basis for deriving a consensus among a set of rankings. A correlation coefficient takes values between -1 and +1, i.e. rankings in full agreement are assigned a correlation of +1, those in full disagreement are assigned a correlation of -1, and all others lie in between. A distance d between two rankings, instead, is a non-negative value, ranging in $[0, Dmax]$, where 0 is the distance

between a ranking and itself, while $Dmax$ varies among distances. This makes the correlation coefficient much more intuitive as a measure of agreement between rankings.
In general, distances between rankings consider all item equally important, and they are not sensitive toward where the disagreement occurs. Kumar and Vassilvitskii (2010) introduced two essential aspects for many applications involving distances between rankings: positional weights and element weights. Positional weights allow to take into account the particular position of disagreement between two rankings when computing their distance/similarity, i.e., for example, the researcher may want to consider swapping elements near the head of a ranking more critical than swapping elements in the tail of the ranking.
Conversely, element weights refer to the role played by the objects that judges are ranking: in certain situations swapping some particular objects should be less penalizing than swapping some others. For example, let us consider a survey in which a group of people is asked to rank ten social networks. In this case, it would be reasonable to assign weights proportional the social network's stock market value (e.g. Facebook would receive the highest weight), so that a disagreement between two popular platforms receive a larger penalization than an inversion between less famous ones. In other words, if two judges agree in assigning the position of the most important alternatives, they will be highly positively correlated. Another example comes from the voting theory: when ranking politicians, the weights allow taking into account that some candidates are similar (belonging to the same party or political coalition) and that transposing similar candidates induces a smaller cost than transposing dissimilar candidates. Here two judges that commit many inversions between politicians coming from different parties will be negatively correlated.
A critical issue involving rankings concerns the aggregation of the preferences in order to identify a compromise or a "consensus" (Kemeny and Snell 1962, Fligner and Verducci 1990). The same problem is known in the literature as the social choice problem, the rank aggregation problem, the median ranking problem, the central ranking detection, or the Kemeny problem, depending on the reference framework (D'Ambrosio et al., 2019). Different approaches have been proposed in the literature to cope with this problem, but probably the most popular is the one related to distances/correlations (Kemeny and Snell 1962,Cook et al. 1986, Fagot 1994, D'Ambrosio and Heiser 2016). As a matter of fact, in order to obtain homogeneous groups of subjects with similar preferences, it is natural to measure the spread between rankings through dissimilarity or distance measures. In this sense, a consensus is defined as the closest ranking (i.e. with the minimum distance) to the whole set of preferences. Another possible way for measuring (dis)-agreement between rankings in a consensus problem is in terms of a correlation coefficient. As already said, a correlation coefficient lies between -1 and +1, while the distance between two rankings ranges in $[0, Dmax]$, where $Dmax$ changes according to the distance. Considering the (finite) set S of all weak orderings of m objects, any rank correlation coefficient on S is also a distance metric on S, and vice versa. A distance metric d can be transformed into a correlation coefficient c (and vice-versa) using the linear transformation $c = 1 - \frac{2d}{Dmax}$.
While a position weighted correlation coefficient τ_x^w, for both linear and weak ordering,

has been proposed by Plaia et al. (2020), here we aim at introducing an element weighted correlation coefficient called $\tau_{x,e}$ as an extension of τ_x provided by Emond and Mason (2002), and a new weighted distance called $d_{K,e}$ as an extension of Kemeny distance. We will prove that the proposed correlation coefficient reduces to Emond and Mason's τ_x when equal weights are set, and that it is related to the proposed distance through the linear transformation $\tau_{x,e} = 1 - \frac{2d_{K,e}}{Dmax}$.
Moreover, the proposed weighted correlation coefficient will be used to deal with a consensus ranking problem, i.e. to find the ranking which best represents the rankings/preferences expressed by a group of judges.
The paper is organized as follows. The next section describes the most used distance measures for ranking data. Section 3 deals with the introduction of element weights in the distance. In Section 4, some intuitive methods to assign weights to elements are discussed. The algorithm for finding the consensus ranking is described in Section 5. In Section 6, the algorithm is applied to simulated and real data. Finally, the concluding remarks are presented in Section 7.

2 Distances and correlation for rankings

Several distance measures have been proposed for ranking data. Given a set X, a distance is a function $d : X \times X \to \mathbb{R}$ where, for all π and $\pi^* \in X$, holds:

1. reflexivity $d(\pi, \pi) = 0$;
2. positivity $d(\pi, \pi*) \geq 0$
3. symmetry $d(\pi, \pi*) = d(\pi*, \pi)$.

A distance measure is said to be a metric when it satisfies the triangle inequality:

4. triangle inequality $d(\pi, \pi*) \leq d(\pi, z) + d(z, \pi*)$, $\forall z \in X$.

Finally, d is said to be a pseudometric if it does not satisfy the identity of indiscernibles:

5. identity of indiscernibles $d(\pi, \pi*) = 0$ if and only if $\pi = \pi*$.

2.1 Kendall's correlation coefficient τ_b

Kendall's correlation coefficient is probably the best-known measure for ranking data (Kendall, 1948). It can be calculated by creating a score matrix of a ranking. A rank vector π with m objects can be transformed into a symmetric $m \times m$ score matrix, whose elements a_{ij} are defined by:

$$a_{ij} = \begin{cases} 1 & \text{if } i \text{ is preferred to } j \\ 0 & \text{if } i = j \text{ or } i \text{ is tied with } j \\ -1 & \text{if } j \text{ is preferred to } i \end{cases} \quad (1)$$

Kendall's correlation coefficient τ_b between two rankings, π with score matrix a_{ij} and π^* with score matrix b_{ij} is defined as:

$$\tau_b(\pi, \pi^*) = \frac{\sum_{i=1}^{m} \sum_{i=1}^{m} a_{ij} b_{ij}}{\sqrt{\sum_{i=1}^{m} \sum_{i=1}^{m} a_{ij}^2 \sum_{i=1}^{m} \sum_{i=1}^{m} b_{ij}^2}}. \tag{2}$$

When two rankings are the reversal of each other τ_b becomes -1. When comparing linear orderings, the denominator always works out to the constant $m(m-1)$. Conversely, when comparing weak orderings the denominator will compute to a lesser value, reduced according to the total number of ties declared in each ranking. Emond and Mason (2002) pointed out that an all-ties ranking results in a zero-filled score matrix and can never be estimated as a solution, because of the zeros in the numerator divided zeros in the denominator results in an unknown number. Kendall's correlation coefficient is a measure of similarity and can be transformed into a dissimilarity or distance measure via the linear transformation $d_{\tau_b} = 1 - \tau_b$, where d_{τ_b} is Kendall's distance.

2.2 Emond and Mason's correlation coefficient τ_x

When dealing with tied rankings, Emond and Mason (2002) showed that Kendall's distance (d_{τ_b}) violates the triangle inequality. To solve this difficulty, they redesigned the elements in Kendall's τ_b score matrix in Eq.(1) and renamed it to τ_x. The elements in the new score matrix a'_{ij} for rank vector π are now defined by:

$$a'_{ij} = \begin{cases} 1 & \text{if } i \text{ is preferred or tied with } j \\ 0 & \text{if } i = j \\ -1 & \text{if } j \text{ is preferred to } i. \end{cases} \tag{3}$$

The extended correlation coefficient is defined as:

$$\tau_x(\pi, \pi^*) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} a'_{ij} b'_{ij}}{m(m-1)}. \tag{4}$$

When ties are not allowed τ_x reduces to τ_b, the former differs from the latter in giving a score of 1 to ties instead of 0; this allows to solve the known Kendall's problems with weak orderings.

2.3 Spearman's distance d_s

The Spearman's distance is calculated by taking the square root of the well known Spearman's ρ. The distance between two rank vectors π and π^* is defined by:

$$d_s(\pi, \pi^*) = \sqrt{\sum_{i=1}^{m} (\pi(i) - \pi^*(i))^2}. \tag{5}$$

When a ranking contains tied objects, these objects must be given the average of the corresponding rank values. A problem identified by Emond and Mason (2000) is that Spearman's d_s suffers from what is known as the sensitivity to irrelevant alternatives (an irrelevant alternative is one that is asymmetrically dominated, this means that the object is less preferred in every ranking to another object but not by every other object (Emond and Mason 2000)). In other words, adding extra irrelevant objects to the ranking exercise could change the maximum agreement solution. This technical flaw arises because Spearman's d_s treats the ranks as numerical values instead of categorical ordered values. Because of this sensitivity to irrelevant alternatives, Spearman's d_s is not suitable as a rank correlation coefficient in the weighted rankings problem.

2.4 Kemeny distance d_K

Kemeny (1959) introduced several constraints that a suitable distance measure for rankings should satisfy:

1. reflexivity, positivity, symmetry and the triangular inequality;
2. the measure of distance should not be affected by a relabeling of the set of objects to be ranked;
3. if two rankings are in complete agreement at the beginning and at the end of the list and differ only in the middle, than the distance does not change after deleting both the first and the last objects to be ranked;
4. the minimum positive distance is one,

and introduced a distance, d_K, that satisfies all these constraints.
The Kemeny distance d_K between two rankings of size m, π with score matrix a_{ij} and π^* with score matrix b_{ij} (a_{ij} and b_{ij} defined as in Eq.(1)) is a city block distance defined as:

$$d_K(\pi, \pi^*) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left| a_{ij} - b_{ij} \right| . \tag{6}$$

The Kemeny distance takes the shortest path between two rankings. The factor a half takes into account that the two triangular matrices that are created by the sum of absolute differences of the score matrices are identical. The maximum distance from a complete ranking to its reversal is $m(m-1)$ while the maximum distance of a ranking containing t ties is given by: $m(m-1) - 2t$.
Considering the usual relation between a distance d and its corresponding correlation coefficient $\tau = 1 - \frac{2d}{Dmax}$, where $Dmax$ is the maximum distance, d_K is in a one-to-one correspondence to the rank correlation coefficient τ_x proposed by Emond and Mason (2002). Distances and correlation are two possible measure used to cope with the consensus ranking problem: given n full or weak rankings of m items, what best represents the consensus opinion? The consensus is the ranking that shows maximum

correlation, or equivalently, minimum distance with the whole set on n rankings.
Given the drawbacks affecting d_s and d_{τ_b}, in this paper we follow the approach based on the Kemeny distance d_K and its corresponding correlation coefficient τ_x.

3 Item weighted distances and correlation coefficient

Distances between rankings treat all items equally, and they are not sensitive to the point of disagreement. Kumar and Vassilvitskii (2010) introduced two issues that are essential for many applications involving distances between rankings, namely, positional weights and element weights. In brief, i) the importance given to swapping elements near the head of a ranking could be higher than the importance attributed to elements belonging to the tail of the list, and ii) swapping important items should receive a larger penalization than swapping negligible ones. The issue of positional weights has been explored by relevant researches (García-Lapresta and Pérez-Román 2010; Can 2014; Plaia et al. 2018, 2019). In this paper, we deal with case ii) and propose the weighted version of the Kemeny metric and the correlation coefficient introduced by Emond and Mason (2002).
The weighting vector $w = (w_1, w_2, ..., w_m)$ with $w_i \geq 0$ is used to take into account the importance of the items where w_i is the importance given to the i^{th}-item in a ranking.

3.1 Introducing element weights in the Kemeny distance

There are many ways to introduce weights in a distance, each of them corresponds to a different penalization of each inversion between two generic items in two rankings. For example, one can decide that an inversion of elements i and j should have a penalty proportional to the arithmetic average of their weights, say $\frac{w_i+w_j}{2}$. The corresponding weighted version of the Kemeny distance, in this case, will be:

$$_a d_{K,e}(\pi, \pi^*) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{w_i + w_j}{2} \left| a_{ij} - b_{ij} \right| . \qquad (7)$$

It can be easily demonstrated that the maximum value of Eq.(7) is equal to $(m - 1) \sum_{i=1}^{m} w_i$. An alternative could be the product of the weights $w_i w_j$, the corresponding weighted Kemeny distance will be defined as:

$$_p d_{K,e}(\pi, \pi^*) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j \left| a_{ij} - b_{ij} \right| , \qquad (8)$$

the maximum value of Eq.(8) being equal to $\sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j$. It can be proved that, regardless of the choice of weighting procedure, the mathematical properties of the Kemeny distance are preserved. Therefore, these methods should be compared in the light of their impact on the resulting weighted Kemeny distance.
Let's consider, for example, three different rankings of three items, say i_1, i_2 and i_3, $R_1 = (1, 2, 3)$, $R_2 = (2, 1, 3)$, $R_3 = (2, 3, 1)$ and a weighting vector $w = (10, 10, 1)$ (Table

1).

Table 1: Weighting vector and data matrix

w		
i_1	i_2	i_3
10	10	1

	Elements		
	i_1	i_2	i_3
R_1	1	2	3
R_2	2	1	3
R_3	2	3	1

Let us compute the weighted Kemeny distances between R_1 and the other rankings R_2, R_3 using the two penalization method discussed before.

Table 2: Weighted Kemeny distances

Items	${}_a d_{K,e}$	${}_p d_{K,e}$
R_1 vs R_2	20	200
R_1 vs R_3	22	40

According to ${}_a d_{K,e}$, the distance R_1 vs R_3 (22) is slightly higher than R_1 vs R_2 (20) while ${}_p d_{K,e}$ claims the contrary, stating that R_1 vs R_3 (40) is far lower than R_1 vs R_2 (200). R_1 assigns the first position to item i_1, the second one to item i_2 and finally item i_3 is ranked third. With R_1 used as reference, R_2 switches the ranks of i_1 and i_2 but keeps i_3 in the last position. R_3 changes the rank of every item moving i_3 to the first position, i_1 to the second one, and finally i_2 to the third one.

Apparently, R_3 changes more frequently the position of items, but it keeps unchanged the ordering of i_1 and i_2. That is to say, either R_3 and R_1 prefer i_1 to i_2, while R_2 doesn't. Since i_1 and i_2 are the most important elements according to the weighting vector w, their inversion should be over penalized. This logically implies that R_3 resembles R_1 more than R_2 does.

Why does the arithmetic average weighted Kemeny distance ${}_a d_{K,e}$ fail to spot the inversion between the most important items i_1 and i_2 committed by R_3?

It is a matter of relative weight, this measure is computed as the ratio of the ij inversion's weight over the total sum of weights, and it represents how much each inversion influences the resulting $d_{K,e}$.

When using the arithmetic average, the relative weight of each inversion between two generic elements i and j is defined as follow:

$$r_{ij}^{a} = \begin{cases} \frac{w_i + w_j}{(m-1)\sum_{i=1}^{m} w_i}, & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{9}$$

while in the case of the product, the relative weight of each inversion is:

$$r_{ij}^{p} = \begin{cases} \frac{2 w_i w_j}{\sum_{i=1}^{m}\sum_{j=1}^{m}(w_i w_j)}, & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases} \tag{10}$$

In both cases the relative weights must sum up to 1; $\sum_{i<j}^{m} r_{ij} = 1$. Let's compute the relative weights with the data of Tab.(1):

Table 3: Relative weights r_{ij}^{a} of each inversion with arithmetic average

	i_1	i_2	i_3
i_1	0	-	-
i_2	**0.476**	0	-
i_3	0.262	0.262	0

Table 4: Relative weights r_{ij}^{p} of each inversion with product

	i_1	i_2	i_3
i_1	0	-	-
i_2	**0.834**	0	-
i_3	0.083	0.083	0

The inversion between i_1 and i_2, when using the arithmetic average, will "cost" approximately the 48% of the maximum obtainable Kemeny distance (Tab.3). In contrast, when using the product (Tab.4), the same inversion has a more considerable influence, equal to 83%. In this example, the product of weights turns out to be an appropriate method, while the arithmetic average produces inconsistent results.

In broader terms, the product aggregation ${}_{p}d_{K,e}$ concentrates the mass of weights on the inversions of the most important items, while the arithmetic average ${}_{a}d_{K,e}$ distributes it more evenly.

The critical point for the researcher is to think about the relative weight of each inversion when assigning the individual weights. From now on, for the purpose of this paper, the product of strictly positive weights ($w_i > 0$) will be used as penalization, keeping in mind that the relative weights are what really matter.

3.2 A new weighted rank correlation coefficient

Combining the weighted Kemeny distance proposed, ${}_pd_{K,e}$, and the extension of τ_x provided by Emond and Mason (2002) we propose a new weighted rank correlation coefficient between two rankings π and π^*:

$$\tau_{x,e}(\pi,\pi^*) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j a_{ij} b_{ij}}{\max[d_{K,e}]} \tag{11}$$

where the denominator represents the maximum value of the weighted Kemeny distance $\max[d_{K,e}] = \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j$.

3.2.1 Correspondence between distance and correlation

Following the relation $\tau = 1 - \frac{2d}{Dmax}$, we prove the following equation:

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j a'_{ij} b'_{ij}}{\max[d_{K,e}]} = 1 - \frac{2d_{K,e}}{\max[d_{K,e}]}. \tag{12}$$

Proof:

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j a'_{ij} b'_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j} = 1 - \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j \left|a_{ij} - b_{ij}\right|}{\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j a'_{ij} b'_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j - \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j \left|a_{ij} - b_{ij}\right|$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j a'_{ij} b'_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j (1 - \left|a_{ij} - b_{ij}\right|)$$

the left side and the right side of the equation are equal, the proof is due to Emond and Mason (2002). To prove this equality we will show that over any pair of objects i and j the two summations correspond, i.e. that

$$\cancel{w_i w_j}(1 - \left|a_{ij} - b_{ij}\right|) + \cancel{w_i w_j}(1 - |a_{ji} - b_{ji}|) = \cancel{w_i w_j}(a'_{ij} b'_{ij}) + \cancel{w_i w_j}(a'_{ji} b'_{ji}) \tag{13}$$

There are nine possible combinations of preferences for objects i and j between rankings A and B, but only four distinct cases must be considered. The other five are equivalent to one of these four through a simple relabelling of the rankings and/or the objects.

Case 1: A prefers object i over j, as does B.
The values are: $a_{ij} = 1$, $a_{ji} = -1$, $b_{ij} = 1$, $b_{ji} = -1$. These yield the left side:

$1-|1-1|+1-|(-1)-(-1)|=2$, the right side values are identical in this case: $a'_{ij}=1$, $a'_{ji}=-1$, $b'_{ij}=1$, $b'_{ji}=-1$ yield the same total: $(1)(1)+(-1)(-1)=2$.

Case 2: A prefers object i over j, while B ranks them as tied.
The values are: $a_{ij}=1$, $a_{ji}=-1$, $b_{ij}=0$, $b_{ji}=-0$. These yield the left side: $1-|1-0|+1-|(-1)-0)|=0$. The right side values are: $a'_{ij}=1$, $a'_{ji}=-1$, $b'_{ij}=1$, $b'_{ji}=-1$ yield the same total: $(1)(1)+(-1)(1)=0$.

Case 3: A prefers object i over j, while B prefers j over i.
The values are: $a_{ij}=1$, $a_{ji}=-1$, $b_{ij}=-1$, $b_{ji}=1$. These yield the left side: $1-|1-(-1)|+1-|-1-1|=-2$. The right side values are: $a'_{ij}=1$, $a'_{ji}=-1$, $b'_{ij}=-1$, $b'_{ji}=1$ yield the same total: $(1)(-1)+(-1)(1)=-2$.

Case 4: Both A and B rank the objects as tied.
The values are: $a_{ij}=0$, $a_{ji}=0$, $b_{ij}=0$, $b_{ji}=0$. These yield the left side: $1-|0-0|+1-|0-0|=2$. The right side values are: $a'_{ij}=1$, $a'_{ji}=1$, $b'_{ij}=1$, $b'_{ji}=1$ yield the same total: $(1)(1)+(1)(1)=2$.
The two methods give identical results in all four distinct cases, completing the proof.

3.2.2 Minimum and maximum values of $\tau_{x,e}$

From the previous demonstrations, $\tau_{x,e}$ assumes its maximum value, equal to 1, if and only if for all i and j only *Case 1* or *Case 4* are observed. Therefore, contrary to what happens with Kendall's τ_b, $\tau_{x,e}$ assumes the maximum value even when a generic all tied ranking is compared with itself. Analogously, $\tau_{x,e}$ can be minimum and equal to -1, if and only for all i and j only *Case 3* occurs.

3.2.3 Correspondence between weighted and unweighted measures

For equal weights assigned to the items, $w_i=C$ with $i=1,2,...,m$ the weighted distance is proportional to the classic Kemeny distance.

$$d_{K,e}=C^2 d_K \tag{14}$$

the proof is straightforward:

$$d_K=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left|a_{ij}-b_{ij}\right| \qquad d_{K,e}=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}w_i w_j\left|a_{ij}-b_{ij}\right|$$

if $w_i=C$ for each $i=1,...,m \Rightarrow w_i w_j=C^2$ and $d_{K,e}=\frac{C^2}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left|a_{ij}-b_{ij}\right|$.

Corollary. Since $\tau_x\equiv d_K$ and $\tau_{x,e}\equiv d_{K,e}$, the weighted rank correlation coefficient is equivalent to the rank correlation coefficient defined by Emond and Mason when equal importance is given to items:
$\tau_{x,e}=\tau_x$ with $w_i=C$ for $i=1,2,...,m$.

3.3 The case of 0-weight items

Sometimes the data matrix $\mathbf{X}$, i.e. the $n \times m$ matrix $\mathbf{X}$, whose l^{th} row represents the ranking associated to the l^{th} judge (defined as in Tab.1), contains some negligible items representing just noise. One may want to compute the weighted Kemeny distance between two or more rankings overlooking the set irrelevant items: to do this those elements will be assigned weight equal to 0. To deal with the 0-weight situation, the data matrix $\mathbf{X}$ should be modified in order to lead back to the well known case $w_i > 0$. Let's define two rankings and one weighting vector: $R_1 = (1, 2, 3, 4, 5)$, $R_2 = (4, 1, 2, 5, 3)$ and $w = (0, 1, 1, 1, 0)$. The weighting vector states that elements i_1 and i_5 shouldn't

Table 5: Weighting vector and original data matrix

w				
i_1	i_2	i_3	i_4	i_5
0	1	1	1	0

	Elements				
	i_1	i_2	i_3	i_4	i_5
R_1	1	2	3	4	5
R_2	4	1	2	5	3

influence the distance between R_1 and R_2. We proceed to remove these two items defining:

- two new rankings R'_1 and R'_2 that keep just the elements with a non-zero weight and re-assign the positions: $R'_1 = (1, 2, 3)$, $R'_2 = (1, 2, 3)$;
- a new weighting vector w' with all non-zero entries $w' = (1, 1, 1)$.

Table 6: Weighting vector and modified data matrix

w'		
i_2	i_3	i_4
1	1	1

	Elements		
	i_2	i_3	i_4
R'_1	1	2	3
R'_2	1	2	3

The new rankings R'_1 and R'_2 concern only three items i_2, i_3 and i_4 (the ones with a non-zero weight). It should be notice that element i_2 is ranked 2^{nd} by R_1, while in the new ranking R'_1 i_2 is ranked 1^{st} since i_1 is removed, a similar situation is met for the other elements i_3 and i_4.
Therefore, the distance $d_{K,e}$ between R_1 and R_2 with weighting vector $w = (0, 1, 1, 1, 0)$

reduces to the distance $d_{K,e}$ between R'_1 and R'_2 with weighting vector $w' = (1, 1, 1)$, and it's equal to 0.
This transformation allows to move from the tricky case of $w_i \geq 0$ to the straightforward case of $w_i > 0$.

4 How to choose the weights

The choice of the weights is crucial, because it strongly influences the Kemeny distance and the corresponding correlation coefficient between two rankings. As pointed out in Subsection 3.1, the weights determine the relative penalization of each inversion. Therefore, the researcher can use the weighting procedure to express his apriori knowledge on the candidates. In general, there is not a unique optimal solution to cope with this problem. Many times is up to the researcher to assign the weights subjectively, while in other situations some unequivocal parameters allow distinguishing the important elements from the irrelevant ones. In this section, two intuitive methods to assign weights are shown.

4.1 Frequency-based weights

This method uses a deterministic procedure to assign individual weights. Suppose that the $n \times m$ data matrix contains n incomplete rankings of m elements, in this case not all the items are ranked by all the judges. Assuming that choosing to rank an item is a proxy of the greater importance that a judge gives to that item (with respect to the items not ranked), the weights w_i can be defined as

$$w_i = 100\frac{T_i}{n} \quad \text{for } i = 1, ..., m, \tag{15}$$

where T_i stands for "number of judges that assigns a non-zero rank to the i^{th}-element", the weights w_i are rounded down so that $w_i \in \mathbb{N}$.
In other words, the frequency-based method assigns higher weights to items that are included several times in partial rankings of the data matrix. When using this method, in order to observe $\tau_{x,e} > \tau_x$ a particular situation must occur: the ordering of the generic elements i and j (e.g. $i \succ j$), who have the highest inclusion probabilities, must be respected by the vast majority of the incomplete rankings. This usually happens, when there is a high agreement between judges assigning the first and the last positions, but there is uncertainty on the middle positions.
Let us consider an example from the Ballon d'Or award voting to clarify the notion of frequency-based weights.
The Ballon d'Or is an annual football award presented by France Football[1] that is generally regarded as the most prestigious individual award for football players. The winner of the FIFA Ballon d'Or is annually chosen, in a system based on positional voting, by international journalists, the coaches, and the FIFA national teams' captains.

[1]https://www.francefootball.fr/

Voters are provided with a shortlist of 23 players from which they could select the three players they deemed to have performed the best in the previous calendar year. That is, each judge returns a partial ranking expressing only the top 3 positions.
Our example will focus on the Ballon d'Or award vote that took place in 2018. The total number of judges was 503, while the players who received at least a vote were: Ronaldo, De Bruyne, Griezmann, Hazard, Kane, Mbappé, Messi, Modric, Salah, Varane. The judges' preferences are reported in table 7.

Table 7: Ordering data matrix

	Players		
	1	2	3
R_1	De Bruyne	Ronaldo	Modric
R_2	Ronaldo	Modric	De Bruyne
R_3	Modric	Ronaldo	De Bruyne
...	...	...	...
R_{503}	Modric	Mbappé	Griezmann

The weights of each footballer, computed according to Eq.(15), are reported in Table 8. Table 9 compares the item-weighted and unweighted Kemeny distances computed

Table 8: Weighting vector

Weights									
Ronaldo	De Bruyne	Griezmann	Hazard	Kane	Mbappé	Messi	Modric	Salah	Varane
56	14	27	20	3	40	24	79	26	11

between the first judges.

Table 9: Item weighted Kemeny distances $d_{K,e}$

Items	d_K	$d_{K,e}$
R_1 vs R_2	4	5894
R_2 vs R_3	2	10962

The introduction of frequency-based weights allows distinguishing high-relevant footballers from negligible ones. In particular, according to the weighted distance, $d_{K,e}$, the judge rankings R_2 and R_3 appear to be more different than the couple R_1-R_2, since they disagree on the ordering of the most important footballers (with the highest weights), i.e. Ronaldo and Modric. On the contrary, the unweighted distance d_K, which does not consider the importance of items, regards the couple R_2-R_3 more similar than the couple R_1-R_2.

4.2 Item similarities

The item similarity criterion follows the idea that swapping two similar elements should be less penalized than swapping two dissimilar ones. A way to assign weights in this situation is to define a symmetric penalization matrix P which reflects the item similarities. The P matrix establishes the penalty ($p_{ij} = p_{ji}$) for each inversion of two generic items. The proposed weighted Kemeny distance between two rankings π and π^* when using the item similarities method is:

$$d_{K,e}(\pi, \pi^*) = \sum_{i<j}^{m} p_{ij} \left| a_{ij} - b_{ij} \right| \tag{16}$$

where p_{ij} is the generic element of the penalization matrix P. The relative weight of each generic inversion can still be computed:

$$r_{ij} = \begin{cases} \frac{p_{ij}}{\sum_{i<j}^{m} p_{ij}}, & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{17}$$

with $\sum_{i<j}^{m} r_{ij} = 1$. To illustrate the notion of similarities, let us consider an example from voting theory.
When dealing with rankings of politicians, it should be taken into account that candidates are gathered into political parties, aimed at the pursuit of common objectives such as adherence to a specific ideological area. Swapping candidates from the same political party should have a smaller impact on the results of an election than swapping candidates from different parties. Let R_1, R_2, R_3 be three rankings of politicians.

Table 10: Data matrix

	Politicians		
	Clinton	Obama	Bush
R_1	1	2	3
R_2	2	1	3
R_3	1	3	2

The rankings R_2 and R_3 differ from R_1 only in one adjacent transposition. In the first case, the swap involves members of the same political party, while in the second case, the transposed candidates belong to two different parties. Hence it is reasonable to assume that the first distance should be smaller than the second one.
In this example, we decided to penalize swapping politicians of the same party with weight equal to 1 while swapping politician of different parties with weight equal to 10, the penalization matrix:

Table 11: Penalization matrix

	Clinton	Obama	Bush
Clinton	0	-	-
Obama	1	0	-
Bush	**10**	**10**	0

Thus, the relative weights are:

Table 12: Relative weights r_{ij} of each inversion with item similarities

	Clinton	Obama	Bush
Clinton	0	-	-
Obama	0.04	0	-
Bush	**0.48**	**0.48**	0

The resulting Kemeny distances are reported in table 13.

Table 13: Unweighted and weighted Kemeny distances

Items	d_K	$d_{K,e}$
R_1 vs R_2	2	2
R_1 vs R_3	2	20

The introduction of weights allows to account for similarities of politicians, in fact according to the weighted distance $d_{K,e}$, R_2 resembles R_1 more than R_3 does.

In conclusion, the item similarity method is handy when dealing with multi-level data. In this case, the data matrix contains rankings of politicians (level 1) who belong to political parties (level 2). Although in this case we compute distances according to formula (16), the item similarity method can be seen as a special case of distance (8) with $p_{ij} = w_i w_j$.

5 Reaching the consensus ranking

There are many approaches for searching for a ranking representative of a group of judges, such as Heuristic methods: Borda (1781) and Dwork et al. (2001), DeConde et al. (2006) or model-based methods Thurstone (1927). Here we follow the approach based on a measure of distance (Cook, 2006), the median ranking approach. The proposed weighted correlation coefficient $\tau_{x,e}$ can be used to deal with a consensus ranking problem. Given a $n \times m$ matrix $\mathbf{X}$, whose l^{th} row represents the ranking associated to the l^{th} judge, the purpose is to identify the median ranking $\hat{S}$ within the universe of the permutations (with ties) of m elements that best represents the average consensus of the subjects involved (i.e. the matrix $\mathbf{X}$). Considering that there is a one-to-one correspondence between a rank correlation coefficient and a distance, the solution ranking is reached by minimizing the average distance or, similarly, maximizing the average rank correlation:

$$\sum_{l=i}^{n} d_{K,e}(x^{(l)}, S) = \min \tag{18}$$

$$\sum_{l=i}^{n} \tau_{x,e}(x^{(l)}, S) = \max \tag{19}$$

Emond and Mason (2002) proposed the BB algorithm to deal with the consensus ranking problem. Recently, Amodio et al. (2016) and D'Ambrosio et al. (2015) proposed two accurate algorithms, they called QUICK and FAST, for identifying the median ranking when dealing with weak and partial rankings, in the framework of the Kemeny approach. The procedure proposed here is based on their approach, but τ_x is replaced with $\tau_{x,e}$. Indicating as s_{ij} and $x_{ij}^{(l)}$ the scoring matrices for S and the l^{th}row of $\mathbf{X}$, $l = 1, .., n$, the problem is:

$$\max \sum_{l=1}^{n} \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j s_{ij} x_{ij}^{(l)}}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j} = \max \sum_{i=1}^{m} \sum_{j=1}^{m} s_{ij} c_{ij}^{ew} \tag{20}$$

where $c_{ij}^{ew} = \sum_{l=1}^{n} w_i w_j x_{ij}^{(l)}$. The score matrix $CI^{ew} = [c_{ij}^{ew}]$ is a modified version of the *Combined Input Matrix* (CI) proposed by Emond and Mason. It is the result of a summation of each input ranking multiplied by the weight. Defined in this way, it summarizes the information about the input rankings and the weights in a single matrix. Emond and Mason conceived a branch-and-bound algorithm to maximize the numerator of Eq (20) (since the denominator is a fixed quantity depending on the number of items and their weights), by defining an upper limit on the value of that dot product. This limit is given by the sum of the absolute values of the elements of CI^{ew}:

$$V = \sum_{i=1}^{m} \sum_{j=1}^{m} \left| c_{ij}^{ew} \right| . \tag{21}$$

Let $Q = \mathbf{1}$ be a vector of ones of size m. Let c_{ij}^{ew} be the $m \times m$ element weighted combined input matrix. By taking into account all the combinations of m objects, each pair of items is evaluated once by considering the two associated cells in CI^{ew}. A moderately accurate first candidate to be the median ranking can be computed as follow:

- If sign $c_{ij} = 1$ and sign $c_{ji} = -1$ then $Q_i = Q_i + 1$;
- If sign $c_{ij} = -1$ and sign $c_{ji} = 1$ then $Q_j = Q_j + 1$;
- If sign $c_{ij} = 1$ and sign $c_{ji} = 1$ then $Q_i = Q_i + 1, Q_j = Q_j + 1$

In this way, we obtain the updated rank vector Q containing the number of times each object is preferred to the others in the pairwise comparisons. This vector is the starting point for the algorithm. The detailed algorithm employing the defined quantities can be found in Amodio et al. (2016) and D'Ambrosio et al. (2015).
Data analysis is performed using our code written in `R` language (available upon request). The proposed BB algorithm has been implemented in `R` environment by suitably modifying the corresponding functions of the ConsRank package (D'Ambrosio et al., 2016).

6 Experimental evaluation

This section aims to show the impact of the element weighting procedure on the consensus ranking. As soon as the weighted version of the QUICK algorithm finds the consensus ranking, a numerical measure of agreement is provided: the weighted correlation coefficient $\tau_{x,e}$. In a consensus problem the value of the corresponding $\tau_{x,e}$ is crucial, because it represents the overall agreement between the estimated consensus $\hat{S}$ and the input rankings $\mathbf{X}$. That is to say, if the consensus ranking's $\tau_{x,e}$ is close to 0 then it's uncorrelated with the input rankings, therefore there is not a real optimal solution. The interest lies in pointing out how the consensus ranking and the corresponding $\tau_{x,e}$ vary according to the weighting vector w employed.
In order to study the performance of the $\tau_{x,e}$ we will consider two simulation studies and two real datasets.

6.1 Simulation under model I

In the first simulation study (Model I) ranking data were generated according to a vector of random variables with 5 independent components $X = (X_1, X_2, X_3, X_4, X_5)^T$, each one following a Gaussian distribution $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. The vector of expected values

is $\mu = (\mu_1 = 0.8, \mu_2 = 1.2, \mu_3 = 1.6, \mu_4 = 1.6, \mu_5 = 1.7)$, and the vector of standard deviations is $\sigma = (\sigma_1 = 0.4, \sigma_2 = 0.3, \sigma_3 = 0.6, \sigma_4 = 0.6, \sigma_5 = 0.4)$.
Each judge observes one realization of the random vector X; $x = (x_1, x_2, x_3, x_4, x_5)^T$ and produces his ranking by assigning the first position, i.e. rank 1, to the item that has the lowest value of x and so on. For example, the k^{th} judge observes the k^{th} realization of X, say $x_k = (1.021, 1.521, 1.474, 2.16, 1.857)$ and assigns the following ranking vector $\pi(x_k) = (1, 3, 2, 5, 4)$.
Since $\mu_1 < \mu_2 < \mu_3 = \mu_4 < \mu_5$, item number 1 will be reasonably placed most of the times in first position while item number 5 in the last one, furthermore having ties is improbable.
The item weighting vectors employed are $w_1 = (1, 1, 1, 1, 1)$, $w_2 = (10, 1, 1, 1, 10)$ and $w_3 = (1, 1, 10, 10, 1)$. Let's remind that w_1 will produce an unweighted version of consensus since it assigns the same weight to each item (see Eq.(14)). In contrast, w_2 assigns higher weights to the external items, and finally w_3 assigns higher weights to the internal items.
We generated 1000 samples of size 100, i.e. $X_{100\times5}$, according to Model I. For each sample, the consensus ranking and the corresponding $\tau_{x,e}$ are estimated according to each weighting vector.
In Tables 14, 15, 16 the relative weights of each generic inversion depending on the weighting vector are reported.

Table 14: Relative weights r_{ij} of each generic inversion when using w_1

	Item1	Item2	Item3	Item4	Item5
Item1	0	-	-	-	-
Item2	0.1	0	-	-	-
Item3	0.1	0.1	0	-	-
Item4	0.1	0.1	0.1	0	-
Item5	0.1	0.1	0.1	0.1	0

Table 15: Relative weights r_{ij} of each generic inversion when using w_2

	Item1	Item2	Item3	Item4	Item5
Item1	0	-	-	-	-
Item2	0.061	0	-	-	-
Item3	0.061	0.001	0	-	-
Item4	0.061	0.001	0.001	0	-
Item5	**0.613**	0.061	0.061	0.061	0

Table 16: Relative weights r_{ij} of each generic inversion when using w_3

	Item1	Item2	Item3	Item4	Item5
Item1	0	-	-	-	-
Item2	0.001	0	-	-	-
Item3	0.061	0.061	0	-	-
Item4	0.061	0.061	**0.613**	0	-
Item5	0.001	0.001	0.061	0.061	0

When equal weights are set (Table 14), each inversion has the same relative weight determining the $d_{K,e}$ and $\tau_{x,e}$. That is to say, the mass of weights is evenly distributed. On the contrary, vectors w_2 and w_3 manly emphasize the inversion of the two most important items attributing the 61.3% of the total weight.
Table 17 counts how many times the i^{th} candidate is chosen to be the consensus ranking by the QUICK algorithm when using the j^{th} weighting vector. Six candidates have been chosen at least once as consensus. It can be noticed that the weighted QUICK, regardless of the weighting vector employed, picks as the optimal solution predominantly the candidates $(1, 2, 4, 3, 5)$ and $(1, 2, 3, 4, 5)$ coherently with the generating model parameters. As one may notice, the algorithm finds more than one optimal solution approximately in 10% of the simulations (total ≈ 1100).
Which are the main differences between the three weighting schemes?
Figure 1 compares the conditional distributions of $\tau_{x,e}$ for the three different weighting vector.

Table 17: Distribution of consensus ranking vs weighting vector

Consensus ranking	w_1	w_2	w_3	Total
1 2 3 4 5	456	449	458	1363
1 2 3 5 4	112	121	109	342
1 2 4 3 5	413	405	413	1231
1 2 4 5 3	16	13	17	46
1 2 5 3 4	119	127	114	360
1 2 5 4 3	26	23	27	76
Total	1142	1138	1138	3418

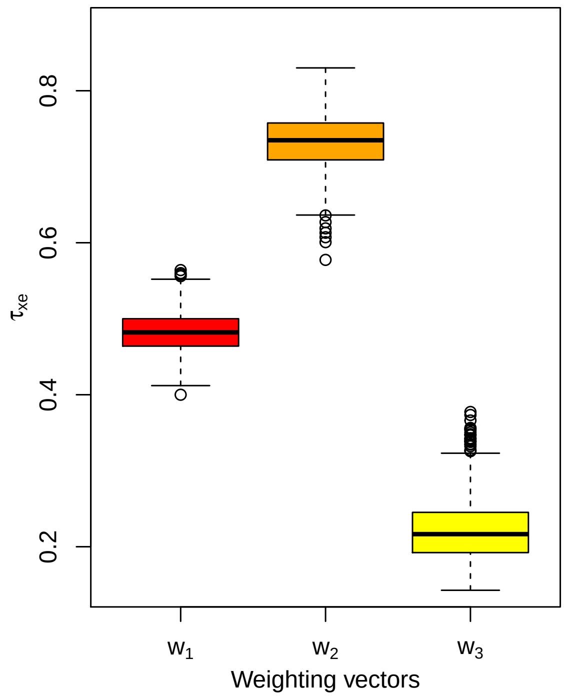

Figure 1: Distribution of $\tau_{x,e}$ vs weighting vectors

The conditional distributions of $\tau_{x,e}$ depending on the weighing vectors are very different.

In particular, when using $w_2 = (10, 1, 1, 1, 10)$ the corresponding $\tau_{x,e}$ takes high values varying from 0.57 to 0.83 with median and mean approximately equal to 0.73. This happens because the vast majority of the judges prefers item number 1 to item number 5. Thus, there is a strong concordance between them assigning the ranking of the items with the highest weight. In fact, as pointed out in Table 15, the inversion of item number 1 with item number 5 has the largest relative weight equal to 61.3%. This implies that, if most of the judges do not commit the over-penalized inversion they will exhibit a firm agreement and this fact will be disclosed by the $\tau_{x,e}$ of the consensus ranking, that is to say that the optimal solution is a proper synthesis of the input rankings.
On the contrary conditioning to $w_3 = (1, 1, 10, 10, 1)$, the corresponding $\tau_{x,e}$ takes small values ranging from 0.14 to 0.37, the median is equal to 0.21 and mean equal to 0.22. Again this is a strong evidence of the impact of weights. The weighting vector w_3 brings out the strong disagreement that exists between the judges in the determination of the rank of item number 3 and item number 4. In this case, just over half of the judges prefer item number 3 to item number 4. Therefore, the consensus ranking found is not a proper synthesis of the input rankings.
Such results are due to either the weighting vectors and the weighting aggregation procedure (i.e. product aggregation) manly emphasizing the inversion of the most important items. If one wants to distribute more evenly the mass of weights, then he should decrease the individual weights, or use another type of weighting scheme (e.g. arithmetic mean or geometric mean).

6.2 Simulation under model II

The second simulation (Model II) is run in order to include ties in the model matrix. Data were generated according to a vector of random variables with 5 independent components $Y = (Y_1, Y_2, Y_3, Y_4, Y_5)^T$, each one following a Gaussian distribution $Y_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. The vector of expected values is $\mu = (\mu_1 = 0.8, \mu_2 = 1.2, \mu_3 = 1.6, \mu_4 = 1.6, \mu_5 = 1.7)$, and the vector of standard deviations is $\sigma' = (\sigma_1 = 0.4, \sigma_2 = 0.3, \sigma_3' = 0.005, \sigma_4' = 0.005, \sigma_5 = 0.4)$. Each judge observes one realization of the random vector Y rounded to the second decimal place $y = (y_1, y_2, y_3, y_4, y_5)^T$ and produces his ranking. The item weighting vectors employed are again $w_1 = (1, 1, 1, 1, 1)$, $w_2 = (10, 1, 1, 1, 10)$ and $w_3 = (1, 1, 10, 10, 1)$. We generated 1000 samples of size 100, i.e. $Y_{100\times5}$ according to Model II. For each sample, the weighted QUICK algorithm estimates the consensus ranking and the corresponding $\tau_{x,e}$ according to the weighting vectors. Due to either the rounding of digital places and the choice of small standard deviation of item number 3 and item number 4, many judges will produce ties in their rankings. The results of the simulation are shown in Table 18 and Figure 2.

Table 18: Distribution of consensus ranking vs weighting vector

Consensus ranking	w_1	w_2	w_3	Total
1 2 3 3 3	4	4	4	12
1 2 3 3 4	979	979	980	2938
1 2 4 4 3	17	17	17	51
Total	1000	1000	1001	3001

Table 18 shows that the choice of the consensus ranking is unequivocal. Over 97% of the time QUICK selects the candidate (1, 2, 3, 3, 4) as optimal solution consistently with the data generator model. This is evidence of the goodness of the algorithm performance.

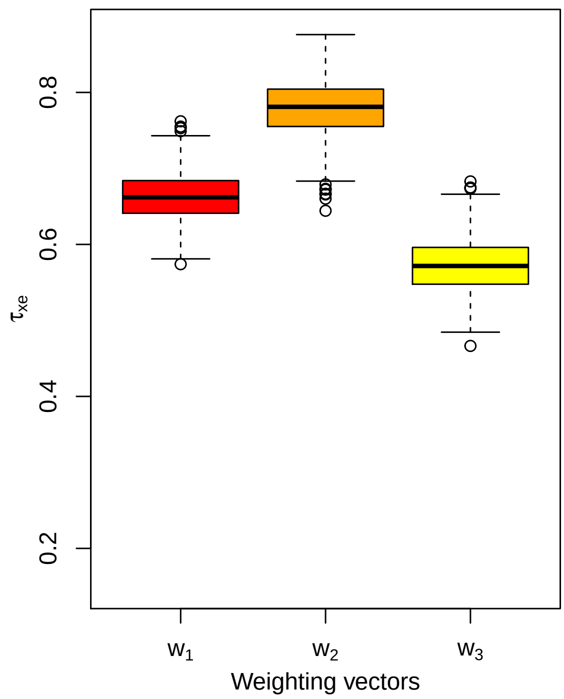

Figure 2: Distribution of $\tau_{x,e}$ vs weighting vectors

Once again, the highest agreement between the judges and the consensus ranking is reached using the weighting vector w_2. In fact, the corresponding $\tau_{x,e}$ varies from 0.64 to 0.87 with mean and median equal to 0.78. The lowest agreement is reached with w_3, when the corresponding $\tau_{x,e}$ takes values between 0.47 and 0.68 with median and mean equal to 0.57. The three conditional distributions turn out to be much more similar than they were in the first simulation. This is due to two factors, firstly the standard deviations of item number 3 and item number 4 ($\sigma_3' = \sigma_4' = 0.005$) are much lower than in the first simulation ($\sigma_3 = \sigma_4 = 0.6$). Therefore item number 3 and item number 4 cause less noise and, consequently, their rankings are indeed defined. This is visible in the Table 18 where there is only one real candidate to be the consensus. In other words, there is less uncertainty about the internal items. Secondly, ties are allowed. In this example, item number 3 and item number 4 are equally likable; therefore, the average agreement among judges will be higher if allowed to express a tie. This is particularly evident in the case of w_3, in the first simulation the similarity between item number 3 and item number 4 caused strong disagreement between the judges, while in the second simulation the two factors menage to mediate.

6.3 ISTAT dataset

ISTAT[2] dataset concerns the sample survey "Aspetti della vita quotidiana" (aspects of daily life); it provides basic information on the daily lives of individuals and families. Since 2005 it has been conducted annually in February. The information gathered makes it possible to learn about citizens' habits and the problems they face every day. Thematic areas on different social aspects follow each other in the questionnaires, allowing to understand how individuals live and how satisfied they are with their conditions, their economic situation, the area in which they live, the functioning of services, etc. The data matrix dimension is 22×10; the rows are the 20 regions of Italy and the autonomous provinces of Trento and Bolzano, the columns stand for the problems related to the city such as: parking difficulties (A), inefficiency of public transport (B), traffic (C), poor street lighting (D), poor road conditions (E), dirty roads (F), air pollution (G), noise (H), risk of crime (I), bad smell (L). In the original data X, the x_{ij} cell is the percentage of people in the i^{th} region, who feel that their city particularly suffers from the j^{th} problem. We re-arranged the data such that within each row rank 1 is assigned to the problem with the highest percentage and so on. In other words, there are 22 judges (the regions) expressing their preferences on 10 elements (problems), where the item that is ranked first is the problem that afflicts the region the most.
The aim is to study the influence of the weighting vector on the resulting consensus ranking. Two weighting vectors will be compared; w_1 which assigns the same weight to each element and w_2 that is based on the item similarity criterion, i.e. swapping similar items should be less penalized than swapping two dissimilar ones.
For this purpose, we found three clusters of items. Cluster number 1 called "Mobility and road conditions" that contains the items: A, B, C, D, E. Cluster number 2 called "Livability" that includes: F, G, H, L. Finally, cluster number 3 contains only element

[2] https://www.istat.it/

I (risk of crime). With w_2, we penalized swapping elements of the same cluster with weight equal to 1, while swapping elements of a different cluster with weight equal to 50.
In this case, the relative weight of each inversion between two generic elements i and j is defined as follow:

$$r_{ij} = \begin{cases} 0.001 & \text{if } i,j \text{ belong to the same cluster} \\ 0.034 & \text{if } i,j \text{ belong to the different clusters} \\ 0 & \text{if } i = j \end{cases} \tag{22}$$

Table 19: Relative weight of each inversion

	A	B	C	D	E	F	G	H	I	L
A	0.000	-	-	-	-	-	-	-	-	-
B	0.001	0.000	-	-	-	-	-	-	-	-
C	0.001	0.001	0.000	-	-	-	-	-	-	-
D	0.034	0.034	0.034	0.000	-	-	-	-	-	-
E	0.001	0.001	0.001	0.034	0.000	-	-	-	-	-
F	0.034	0.034	0.034	0.001	0.034	0.000	-	-	-	-
G	0.034	0.034	0.034	0.001	0.034	0.001	0.000	-	-	-
H	0.034	0.034	0.034	0.001	0.034	0.001	0.001	0.000	-	-
I	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.000	-
L	0.034	0.034	0.034	0.001	0.034	0.001	0.001	0.001	0.034	0.000

The consensus estimated for each weighting vector is shown in Table 20

Table 20: Consensus ranking for each weighting vectors

	1	2	3	4	5	6	6	8	9	10	$\tau_{x,e}$
w_1	E	A	B	C	D	G	H	F	I	L	**0.69**
w_2	E	A	B	C	D	G	H	F	I	L	**0.78**

The consensus ranking shows that elements of cluster 1 "Mobility and road conditions" take up the first five positions. In particular element E (road conditions) worries the citizens the most. The impact of weights is visible, although the optimal solution remains the same the value of $\tau_{x,e}$ increases. The value of the correlation coefficient stands for the representativeness of the optimal solution found by the algorithm. In this case, taking into account the element similarities brings to an increase of the representativeness of the consensus ranking. The positive variation of $\tau_{x,e}$ reveals that most of the times the disagreement among the regions' rankings occur between similar elements, i.e. belonging to the same cluster. Therefore the general weighted agreement, computed with w_2, is higher than the unweighted one computed using w_1.

6.4 Quiz dataset

The quiz dataset (Jacques et al., 2014) contains the answers of 70 students (40 of the third year and 30 of the fourth year) from Polytech'Lille (Statistics Engineering School, France) to the four following quizzes: Literature Quiz, Football Quiz, Mathematics Quiz and Cinema Quiz. In this study the Mathematics Quiz will be analyzed, it consists of ranking four numbers according to increasing order: $A = \frac{\pi}{3}$, $B = \log(1)$, $C = e^2$, $D = \frac{1+\sqrt{5}}{2}$.
Each student provides his ranking without using the calculator such that the data matrix has 70 rows and 4 columns. Differently from the previous examples the exact order of items is known, that is; $\log(1) < \frac{\pi}{3} < \frac{1+\sqrt{5}}{2} < e^2$, i.e $B < A < D < C$.
The QUICK algorithm allows us to find out the unweighted consensus ranking, that is: B, A, D, C with correlation coefficient $\tau_x = 0.85$. Therefore the global solution is the right one, furthermore the degree of concordance between students is high.
Now we assume that the students had no difficulty in realizing that the elements B ($log(1)$) and C (e^2) had to be placed in the first and in last position respectively, and maybe this "easy choice" let the correlation coefficient grows. Therefore we want to test whether the students were good enough to recognize the exact order of elements A ($\frac{\pi}{3}$) and D ($\frac{1+\sqrt{5}}{2}$). A way of doing that is to define a vector of weights $w = (10, 1, 1, 10)$ that emphasises the inversion between A and D and then to compute the weighted consensus ranking and the correspondence correlation coefficient $\tau_{x,e}$. The relative weight of each inversion is reported in Tab.21.

Table 21: Relative weights r_{ij} of each generic inversion

	A	B	C	D
A	0.000	-	-	-
B	0.071	0.000	-	-
C	0.071	0.007	0.000	-
D	**0.709**	0.071	0.071	0.000

The weighted consensus ranking is: B, A, D, C and the corresponding $\tau_{x,e}$ is 0.72. What does it mean? We can say that the value of $\tau_{x,e}$, although decreased, is still quite high, indicating that the unweighted consensus was robust and not mainly influenced by the "easy choice". At the same time, items A and D are indeed the most difficult values to rank: $\tau_{x,e}$ assumes its minimum value, 0.72, with a vector of weights $w = (10, 1, 1, 10)$ and its maximum value, 0.90, with a vector of weights $w = (1, 10, 10, 1)$ (items B and C are the easiest to rank). In this way $\tau_{x,e}$ can also be useful to verify where (i.e. referring to which items) the disagreement between rankings mainly occurs.

7 Concluding remarks

Within the framework of preference data, where individuals express their preferences over a set of items, the main interest lies in evaluating the agreement between them and obtaining a synthesis of their preferences by computing a consensus ranking. Different approaches have been proposed in the literature to cope with this problem, but the most popular one is probably the one related to distances/correlations. Usually, these are not sensitive to the importance of items, since each inversion is considered equally important. In many cases, this assumption could be simplistic. For this reason in this paper, we provided an element weighted rank correlation coefficient $\tau_{x,e}$ for linear, weak and incomplete orderings. We demonstrated the correspondence between $\tau_{x,e}$ and the corresponding weighted Kemeny distance $d_{K,e}$. Finally, we showed that, in the case of equal weights for all items $w_i = C$, the weighted rank distance $d_{K,e}$ is proportional to the well known Kemeny distance d_K, while the correlation coefficient $\tau_{x,e}$ is equal to the Emond and Mason's τ_x. From the simulation study and the real data examples, we demonstrated that the BB algorithm allows us to find the true consensus and to show how the weighting vector affects the representativity of the median ranking. The weighted consensus algorithm's computational effort was investigated by considering some simulations. We progressively increased the sample size (from 200 to 1000) and the number of items (from 3 to 10). Compared with the unweighted algorithm, the weighted consensus algorithm entails a slight increase in computational time which has never exceeded 30%. Future studies could fruitfully explore this issue further by including the element weighting procedure in a cluster analysis of ranking data. When dealing with preference data, the cluster analysis attempts to identify homogeneous groups of rank choices (clusters). The use of weights allows taking into account the importance of alternatives minimizing the distances between cluster members.

Acknowledgement

References

Amodio, S., D'Ambrosio, A., and Siciliano, R. (2016). Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the kemeny axiomatic approach. *European Journal of Operational Research*, 249(2):667–676.

Borda, J.-C. d. (1781). Mémoire sur les élections au scrutin: Histoire de l'académie royale des sciences. *Paris, France*, 12.

Can, B. (2014). Weighted distances between preferences. *Journal of Mathematical Economics*, 51:109–115.

Cook, W. D. (2006). Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of operational research*, 172(2):369–385.

Cook, W. D., Kress, M., and Seiford, L. M. (1986). An axiomatic approach to distance on partial orderings. *RAIRO-Operations Research*, 20(2):115–122.

D'Ambrosio, A., Amodio, S., and Iorio, C. (2015). Two algorithms for finding optimal solutions of the kemeny rank aggregation problem for full rankings. *Electronic Journal of Applied Statistical Analysis*, 8(2):198–213.

D'Ambrosio, A., Amodio, S., and Mazzeo, G. (2016). Consrank, compute the median ranking (s) according to the kemeny's axiomatic approach. r package version 2.0. 0.

D'Ambrosio, A. and Heiser, W. J. (2016). A recursive partitioning method for the prediction of preference rankings based upon kemeny distances. *psychometrika*, 81(3):774–794.

D'Ambrosio, A., Iorio, C., Staiano, M., and Siciliano, R. (2019). Median constrained bucket order rank aggregation. *Computational Statistics*, 34(2):787–802.

D'Ambrosio, A., Mazzeo, G., Iorio, C., and Siciliano, R. (2017). A differential evolution algorithm for finding the median ranking under the kemeny axiomatic approach. *Computers & Operations Research*, 82:126–138.

DeConde, R. P., Hawley, S., Falcon, S., Clegg, N., Knudsen, B., and Etzioni, R. (2006). Combining results of microarray experiments: a rank aggregation approach. *Statistical applications in genetics and molecular biology*, 5(1).

Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. (2001). Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622.

Emond, E. J. and Mason, D. W. (2000). *A new technique for high level decision support.* Department of National Defence Canada, Operational Research Division.

Emond, E. J. and 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.

Fagot, R. F. (1994). An ordinal coefficient of relational agreement for multiple judges. *Psychometrika*, 59(2):241–251.

Fligner, M. A. and Verducci, J. S. (1990). Posterior probabilities for a consensus ordering. *Psychometrika*, 55(1):53–63.

García-Lapresta, J. L. and Pérez-Román, D. (2010). Consensus measures generated by weighted kemeny distances on weak orders. In *2010 10th International Conference on Intelligent Systems Design and Applications*, pages 463–468. IEEE.

Heiser, W. J. and D'Ambrosio, A. (2013). Clustering and prediction of rankings within a kemeny distance framework. In *Algorithms from and for Nature and Life*, pages

19–31. Springer.

Jacques, J., Grimonprez, Q., and Biernacki, C. (2014). Rankcluster: An r package for clustering multivariate partial rankings.

Kemeny, J. G. (1959). Mathematics without numbers. *Daedalus*, 88(4):577–591.

Kemeny, J. G. and Snell, L. (1962). Preference ranking: an axiomatic approach. *Mathematical models in the social sciences*, pages 9–23.

Kendall, M. G. (1948). Rank correlation methods.

Kumar, R. and Vassilvitskii, S. (2010). Generalized distances between rankings. In *Proceedings of the 19th international conference on World wide web*, pages 571–580.

Plaia, A., Buscemi, S., and Sciandra, M. (2019). A new position weight correlation coefficient for consensus ranking process without ties. *Stat*, 8(1):e236.

Plaia, A., Buscemi, S., and Sciandra, M. (2020). Consensus measures among preference rankings: a new weighted correlation coefficient for linear and weak orderings. *Submitted*.

Plaia, A., Sciandra, M., and Buscemi, S. (2018). Consensus measures for preference rankings with ties: an approach based on position weighted kemeny distance. *Advances in Statistical Modelling of Ordinal Data*, page 171.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.