Distance and Consensus for Preference Relations Corresponding to Ordered Partitions

Boris Mirkin[1,2] · Trevor I. Fenner[2]

Abstract

Ranking is an important part of several areas of contemporary research, including social sciences, decision theory, data analysis, and information retrieval. The goal of this paper is to align developments in quantitative social sciences and decision theory with the current thought in Computer Science, including a few novel results. Specifically, we consider binary preference relations, the so-called weak orders that are in one-to-one correspondence with rankings. We show that the conventional symmetric difference distance between weak orders, considered as sets of ordered pairs, coincides with the celebrated *Kemeny distance* between the corresponding rankings, despite the seemingly much simpler structure of the former. Based on this, we review several properties of the geometric space of weak orders involving the ternary relation "between," and contingency tables for cross-partitions. Next, we reformulate the consensus ranking problem as a variant of finding an optimal linear ordering, given a correspondingly defined consensus matrix. The difference is in a subtracted term, the partition concentration that depends only on the distribution of the objects in the individual parts. We apply our results to the conventional *Likert scale* to show that the *Kemeny consensus* rule is rather insensitive to the data under consideration and, therefore, should be supplemented with more sensitive consensus schemes.

Keywords Ranking · Tied ranking · Ordered partition · Weak order · Distance · Consensus · Muchnik test

✉ Boris Mirkin
  bmirkin@hse.ru; mirkin@dcs.bbk.ac.uk

  Trevor I. Fenner
  trevor@dcs.bbk.ac.uk

[1] Department of Data Analysis and Machine Intelligence, National Research University Higher School of Economics, 20 Miasnitskaya str., Moscow RF, Russia

[2] Department of Computer Science, Birkbeck University of London, Malet Street, London WC1E 7HX, UK

 Springer

1 Background and Motivation

Analysis of rankings and tied rankings as a topic in data analysis is receiving increasing attention, especially in relation to automated decision-making and search. Considered initially as a model for human judgment in psychology, ranking came to the attention of social scientists as a basis for voting and other social decisions. It was next adopted in economics as a model of utility and rational choice. Rankings are currently used in many areas of computational data analysis. Among abundant examples are the following: ordering alternatives in collaborative filtering, document ranking in information retrieval engines, similarity scoring of protein sequences in bioinformatics, and league tables in sports and in higher education. Important sources of such ranked data are social surveys in which respondents are asked to classify alternatives using an ordered set of categories, such as "strongly agree," "agree," "neutral," "disagree," and "strongly disagree." This type of scale is widely used for scoring the strength of complex phenomena, such as winds, earthquakes, and sentiment expressions in texts.

The topic of comparing tied rankings was initiated by Charles Spearman (a junior collaborator of the founding fathers of multivariate statistics, Francis Galton and Karl Pearson), who was hired to further pursue the golden dream of Galton, a proof that human talent is inherited from one's parents and, partly, from even more distant ancestors. Although ranking is non-quantitative, Spearman proposed using ranks as numerical values, so that the Pearson correlation coefficient could be employed (Spearman 1904). This is straightforward when the observations being compared are linearly ordered. However, different observations can sometimes be assigned the same numerical rank value, which led to the introduction of the term "tied observations," subsequently replaced by "*tied ranking*." Formally, a tied ranking can be represented as an ordered partition $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$, that is, a partition whose parts are linearly ordered by their indices $1, 2, \ldots, p$. We say that an element i *precedes* an element j in the tied ranking $\boldsymbol{R}$ if the part containing i precedes the part containing j. A clear-cut case of an ordered partition is given by the rank features in social surveys, as mentioned above. There is a difference between tied rankings and ordered partitions by rank features: as the number of objects grows, the length of a tied ranking will grow accordingly, whereas the number of categories of a feature and, thus, parts in the corresponding ordered partition, remains constant. However, this difference plays no role in the current paper as we do not consider processes in which the numbers of objects grow.

Despite the use of the Pearson correlation coefficient for non-metric judgements having no theoretical foundation, it took another 35 years to develop a more satisfactory approach. In 1938, Maurice Kendall introduced a different representation for rankings by using the relation of precedence between ranks rather than the ranks themselves (Kendall 1938). Given a tied ranking $\boldsymbol{R}$, we define a square observation-to-observation matrix (the *Kendall matrix*) in which the (i, j) entry is $+1$ if i precedes j in $\boldsymbol{R}$, 0 if i and j have the same rank, or -1 if j precedes i. The Kendall rank correlation coefficient between two tied rankings is the Pearson correlation coefficient between the corresponding Kendall matrices, considered as vectors in an $N \times N$-dimensional space, where N is the number of observations. This accords with the non-quantitative nature of tied rankings. The twentieth century saw intense work on the analysis of the relationship between the Kendall and Spearman correlation coefficients; these were proven to be asymptotically equivalent under conventional statistical assumptions (Lehmann and D'Abrera 2006). This is also compatible with the finding that metrics based on the Spearman and Kendall representations are nearly equivalent, that is, their ratio always lies in a fixed finite interval (Diaconis and Graham 1977). Yet, the similarity between them should

not be overemphasized: in a more recent investigation, for a model in which a bivariate Gaussian distribution is contaminated with a dose of high variation noise, these two rank correlation coefficients were found to behave differently (Xu et al. 2013).

In the 1950s, John Kemeny approached the issue of comparing rankings from a social consensus perspective. Given a set of ordered partitions, a consensus ordered partition should represent the major tendency in the set. A conventional approach, the majority rule, may fail when determined by voting on pairs of alternatives. Specifically, the so-called Condorcet paradox holds: if there are three parties at a meeting, each supporting one of three cyclically related linear orderings of three alternatives, say, (a) $[i, j, k]$, (b) $[j, k, i]$, and (c) $[k, i, j]$, respectively, then the majority rule would lead to a cycle in the precedence relation: i would precede j because this is so for the majority, (a) and (c); similarly, j would precede k, and k would precede i. This contradicts the requirement that the precedence relation corresponding to the majority consensus ranking should be transitive. This paradox is a basis of the celebrated "social choice impossibility theorem" by Kenneth Arrow (see, for example, Aleskerov and Monjardet 2002).

John Kemeny proposed a different definition for consensus ranking using a distance measure between tied rankings. Rather than defining any specific distance measure ab initio, he formulated four axioms that should hold for any admissible distance measure. These axioms led Kemeny to derive the unique distance measure satisfying them. The *Kemeny distance* turned out to be the *L_1-distance* between the Kendall matrices (see Kemeny and Snell 1962 and Heiser and D'Ambrosio 2013).

In the monograph (Mirkin 1979), a joint geometric space of ordered and unordered partitions was considered using the corresponding weak order and equivalence relations on the set of observations. The distance was introduced as the *mismatch*, or symmetric difference, distance between the binary relations as subsets of the Cartesian product of the set of objects with itself. This distance can also be expressed as the L1-distance, or Hamming distance, between the 0 and 0-1 matrices of the relations. Unfortunately, the differences between the mathematical structures of rankings and the corresponding binary relations were glossed over in (Mirkin 1979). The ordered partitions were only considered via the corresponding binary relations. On the one hand, this was useful because of the greater flexibility of binary relations (being just subsets of ordered pairs) over tied rankings (being ordered sets of subsets). For example, the binary relation perspective allowed a reinterpretation of Arrow's impossibility theorem by using the fact that the set-theoretic union of transitive binary relations is not necessarily transitive. On the other hand, the suppression of the ranking aspects led to ignoring the ways international developments were proceeding. For example, the fact that the mismatch and Kemeny distances were equal, although well known to the author (see Mirkin and Cherny 1972), was never proven, nor even formulated, in the book. This perhaps resulted in the distance between preference relations being excluded from the current ranking research discourse. The binary relations related to rankings play a more or less hidden role and are just used as auxiliary constructions in the derivations; (Fagin et al. 2006) it can be considered a representative example of this.

This paper is aimed at bringing the distance between binary relations back onto the scene. Therefore, we formulate the Kemeny approach in terms of binary preference relations, the so-called weak orders, and derive some properties of the *Kemeny consensus* (or *median*) that so far have not appeared in the international literature. These include, in particular, the following:

(i)   A proof that the Kemeny distance between rankings is, in fact, the mismatch distance between the corresponding weak-order binary relations. The importance of this result stems from the fact that the former involves Kendall object-to-object matrices with three

possible values for the entries: 1 for preceding, − 1 for following, and 0 for a tie; whereas the latter involves only two: 1 for the presence and 0 for the absence of a pair in the binary relation. This makes the result rather counter-intuitive.

(ii) Explicit statements of some of the properties of the distance, especially those regarding the relationship between weak orders and their induced equivalence relations, using the ternary relation "between" on the set of binary relations and the notion of "refinement" on the set of tied rankings.

(iii) An explicit reformulation of the Kemeny consensus criterion in terms of the relational summary matrix, analogous to the so-called consensus matrix in the problem of consensus clustering (see, for example, Mirkin 2012). In contrast to the analysis of consensus clustering, however, the (i, j) entry in this consensus matrix is not simply the number of partitions for which elements i and j belong to the same part, but also includes the number of rankings for which i precedes j. The problem, which involves the subtraction of a threshold, is equivalent to maximizing the sum of the consensus matrix entries minus the number of pairs in the corresponding equivalence relation (sometimes referred to as the partition concentration index), weighted with a penalty defined by the threshold. The subtracted part plays the role of a naturally emerged regularizer. Of course, the regularizer does not affect the solution if it is restricted to a class of ranked partitions for which the partition concentration index is constant, like, for example, the class of linear rankings with no ties.

(iv) The sensitivity of the Kemeny median concept is tested by extending the so-called *Muchnik test* from the case of unordered partitions (Mirkin 2012) to the case of ordered partitions. Specifically, we apply the concept of median to the *Likert scales* popular in Psychology (Likert 1932). Given an ordered partition $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$, the Likert scale replaces $\boldsymbol{R}$ by the set of binary ordered partitions $\boldsymbol{S}^t$ $(t = 1, 2, \ldots, p\text{-}1)$ that separate the union of the first t parts of $\boldsymbol{R}$ from the rest. The question then arises as to whether $\boldsymbol{R}$ is a median for the set of binary rankings $\boldsymbol{S}^t$ $(t = 1, 2, \ldots, p\text{-}1)$, as one might expect, or not. Perhaps surprisingly, it turns out that it is one of the "coarse" binary rankings $\boldsymbol{S}^t$ that is a median, rather than $\boldsymbol{R}$ itself.

The remainder of the paper is structured as follows. In Section 2, we describe the mathematical structures of tied rankings (ordered partitions) and the corresponding weak orders. This includes the operation of intersection and a ternary relation "between." In Section 3, we analyze distances between rankings and between weak orders, including a proof that the Kemeny distance between rankings and the mismatch distance between weak orders are equal. Various expressions for the distance in terms of the elements of the contingency table are reviewed, and its representation as the sum of the "ordered" and "unordered" components is precisely formulated. In Section 4, we describe the Kemeny median consensus rankings in terms of a related "triangulation" problem, with a specific emphasis on the penalty coefficient and its effect on the distribution of the median ordered partition. We also demonstrate that the median is rather insensitive to the granularity of ties in the raw rankings. To this end, we utilize the Muchnik test from the theory of consensus clustering (Mirkin 2012) and apply it to the median-ordered partition. Specifically, we analyze what median would emerge as a consensus for dichotomous versions of the Likert scale: unfortunately, as mentioned above, this is nothing more granular than one of the central binary rankings. Section 5 concludes the paper.

2 Ordered Partitions and Preference Relations

2.1 Characterization of the Preference Relation for an Ordered Partition

Given a finite set A of N elements, a collection of its subsets $\boldsymbol{R} = \{R_1, R_2, \ldots, R_p\}$ is referred to as a *partition* if the subsets R_s are all non-empty, non-overlapping, and cover the entire set A, so that each $i \in A$ belongs to a unique subset R_s, $1 \le s \le p$. The subsets are called the parts of the partition $\boldsymbol{R}$. A partition is said to be *ordered* if there is a linear order relation of precedence between its parts, $R_s < R_t$, that is transitive, anti-reflexive, and complete. If the order coincides with the natural order between indices 1, 2, …, p, we use parentheses to denote this, viz. $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$. In Decision Theory, an ordered partition is referred to as a *tied ranking*. In Computer Science, the terms "partial ranking" and "bucket partition" have sometimes been used (Fagin et al. 2006). We consider that the term "partial ranking" should apply according to usage in the mathematical theory of partial orders: when the precedence relation between parts is not complete, so that for some distinct s and t, the precedence between R_s and R_t is not defined. We do not consider here this type of partial ranking.

Each ordered partition $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$ generates a *binary preference relation*

$$\rho = \{(i, j) : i \in R_s, j \in R_t, \text{and } s \le t\}. \tag{1}$$

Usually, two non-overlapping binary relations are defined with respect to a tied ranking $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$: the *strict preference* relation $P = \{(i, j): i \in R_s, j \in R_t, \text{and } s < t\}$ and the *indifference* relation $E = \{(i, j): i, j \in R_s \text{ for some } s\}$. The indifference relation E here is transitive, reflexive, and symmetric, thus E is the equivalence relation corresponding to the unordered partition $\check{\boldsymbol{R}}$ having the same parts as $\boldsymbol{R}$. Obviously, $\rho = P \cup E$, that is, ρ in (1) is a non-strict preference relation in which the strict preference and indifference relations are merged together. Usually, researchers try to avoid such a "mix;" but we will see later that there is no problem with this merger. The next part of this section is a brief reminder of some conventional concepts and facts about preference relations (see, for example, Steele and Stefánsson 2015).

If ρ is a binary relation, its inverse ρ^{-1} is defined as $\rho^{-1} = \{(i, j): (j, i) \in \rho\}$. If ρ is the preference relation corresponding to a tied ranking $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$, then its inverse ρ^{-1} corresponds to the reverse tied ranking $\boldsymbol{R}^{-1} = (R_p, \ldots, R_2, R_1)$. It is easy to see that the indifference relation E corresponding to any tied ranking $\boldsymbol{R}$ satisfies $E = \rho \cap \rho^{-1}$. Thus, the strict preference relation is the difference $P = \rho - E = \rho - \rho^{-1}$.

It is clear that ρ in (1) is

– Reflexive, that is, $(i, i) \in \rho$ for any $i \in A$,
– Transitive, that is, if $(i, j) \in \rho$ and $(j, k) \in \rho$, then $(i, k) \in \rho$ for any $i, j, k \in A$, and
– Complete, that is, $(i, j) \in \rho$ or $(j, i) \in \rho$, or both, for any $i, j \in A$.

Of course, reflexivity can be considered as the special case of completeness for which $i = j$. A binary relation satisfying these properties is usually referred to as a *weak order* (Steele and Stefánsson 2015). In fact, a converse statement also holds:

Theorem 1 A preference relation ρ corresponds to an ordered partition $\boldsymbol{R}$ if and only if it is a weak order.

Proof Let ρ be a binary relation on the set A that is reflexive, transitive, and complete. Consider any $i \in A$ and define the subset $\rho(i) = \{j \in A: (i, j) \in \rho\}$. Then, for any pair $i, k \in A$, if $(i, k) \in \rho$ then $\rho(k) \subseteq \rho(i)$. This holds because whenever $j \in \rho(k)$, i.e., $(k, j) \in \rho$, then $(i, j) \in \rho$ also, because ρ is transitive. Therefore, since ρ is complete, for any pair $i, k \in A$, either $\rho(k) \subseteq \rho(i)$ or $\rho(i) \subseteq \rho(k)$, or both. It follows that the collection of sets $\rho(i)$ is linearly ordered by set-theoretic inclusion, so they can be ordered as a sequence of sets S_t for $t = 1, 2, …, p$, where $S_1 \supset S_2 \supset … \supset S_p$. Then the subsets $R_t = S_t - S_{t+1}$, $t = 1, 2, …, p\text{-}1$, and $R_p = S_p$, form an ordered partition $\mathbf{R} = (R_1, R_2, …, R_p)$. It is quite easy to check that its corresponding preference relation (1) coincides with the given relation ρ. The reverse implication, that the relation (1) corresponding to an ordered partition is reflexive, transitive, and complete, has already been established above. This completes the proof.

Corollary 1 The subsets $R_t = S_t - S_{t+1}$ in the proof each satisfy $R_t = \rho(i) \cap \rho^{-1}(i)$ for some $i \in A$.

Corollary 2 A binary relation ρ is a weak order if and only if its strict part P is anti-reflexive and transitive, its indifference part E is an equivalence relation, and P, P^{-1}, E form a partition of the Cartesian product $A \times A$.

2.2 Refinement and Betweenness

A tied ranking $\mathbf{R'}$ is a *refinement* of a tied ranking $\mathbf{R}$ if it is obtained from the latter by subdividing some of its parts into smaller ones, and some ordering is defined between the smaller parts of each subdivided part of $\mathbf{R}$. The corresponding preference relations, ρ' and ρ, are related by set-theoretic inclusion:

Theorem 2 A tied ranking $\mathbf{R'}$ is a refinement of a tied ranking $\mathbf{R}$ if and only if $\rho' \subset \rho$.

Proof Indeed, if $\mathbf{R'}$ is a refinement of a tied ranking $\mathbf{R}$ then, for some pairs i, j of elements of A such that both $(i, j) \in \rho$ and $(j, i) \in \rho$, only one of these holds for ρ'. Conversely, suppose that ρ and ρ' correspond to tied rankings $\mathbf{R}$ and $\mathbf{R'}$, respectively, and that $\rho' \subset \rho$. Then $\rho'(i) \subseteq \rho(i)$ for any $i \in A$, and, moreover, the inclusion is proper for some $i \in A$. Consider any such i.

Let $\{i1, i2, …, ik\}$ be a maximal subset of A such that $\rho(i) \supset \rho'(i1) \supset \rho'(i2) \supset … \supset \rho'(ik)$. Then, by Corollary 1, every equivalence class $R'_{tu} = \rho'(iu) \cap \rho'^{-1}(iu)$ will be part of the equivalence class $R_t = \rho(i) \cap \rho^{-1}(i)$, which completes the proof.

We say that ρ is *coarser* than ρ', if ρ' is a refinement of ρ.

A binary relation τ on A is said to be *between* binary relations ρ and ρ' if and only if $\rho \cap \rho' \subseteq \tau \subseteq \rho \cup \rho'$ (Mirkin 1979). A tied ranking T is said to be between tied rankings $\mathbf{R}$ and $\mathbf{R}$ if, for any $i, j \in A$, the ordering between them in T is compatible with their ordering in both $\mathbf{R}$ and $\mathbf{R'}$: that is, (i) if i precedes j in both $\mathbf{R}$ and $\mathbf{R'}$ then i precedes j in T; (ii) if i precedes j in one of $\mathbf{R}$ and $\mathbf{R'}$, and i and j are indifferent in the other, then i either precedes j or is indifferent to j in T; (iii) if i and j are indifferent in both $\mathbf{R}$ and $\mathbf{R'}$, then i and j are indifferent in T; lastly, (iv) if i precedes j in $\mathbf{R}$ but j precedes i in $\mathbf{R'}$, then anything can be true of the ordering between i and j in T: i may precede j, or j may precede i, or i and j may be indifferent in T (Kemeny and Snell 1962). It follows that T is between $\mathbf{R}$ and $\mathbf{R'}$ if and only if the same is true for their weak orders, as stated in the following theorem.

Theorem 3 A preference relation τ on A, corresponding to the tied ranking T, is between preference relations ρ and ρ', corresponding to tied rankings $\boldsymbol{R}$ and $\boldsymbol{R}'$, if and only if T is between $\boldsymbol{R}$ and $\boldsymbol{R}'$.

For preference relations ρ and ρ' corresponding to tied rankings $\boldsymbol{R}$ and $\boldsymbol{R}'$, usually neither $\rho \cap \rho'$ nor $\rho \cup \rho'$ corresponds to a tied ranking. There is a class of situations, however, for which the intersection does correspond to a tied ranking. We say that $\boldsymbol{R}$ and $\boldsymbol{R}'$ are *concordant* if there exists a linear ordering that is a refinement of both $\boldsymbol{R}$ and $\boldsymbol{R}'$; so the parts of both are just intervals of the underlying linear ordering. In this case, $\rho \cap \rho'$ does correspond to a tied ranking and the equivalence classes corresponding to $\rho \cap \rho'$ are formed by the non-empty intersections of these intervals of the linear ordering.

In the general case of two arbitrary tied rankings $\boldsymbol{R}$ and $\boldsymbol{R}'$, the relation $\rho \cap \rho'$ is a partial preference relation because there can be $i, j \in A$ such that i strictly precedes j in $\boldsymbol{R}$, whereas j strictly precedes i in $\boldsymbol{R}'$, so that neither (i, j) nor (j, i) belongs to $\rho \cap \rho'$. Such a case, which is not uncommon, is exemplified by a proverbial question: "What is better: being poor but healthy or being rich but ill?" (with a proverbial answer that to be both rich and healthy is better indeed.)

What is appealing about $\rho \cap \rho'$ is that its indifference relation is always an equivalence relation, thus corresponding to the partition that is just the intersection of the unordered partitions $\check{\boldsymbol{R}}$ and $\check{\boldsymbol{R}}'$ that correspond to the ordered partitions $\boldsymbol{R}$ and $\boldsymbol{R}'$, respectively. The intersection $\check{\boldsymbol{R}} \cap \check{\boldsymbol{R}}'$ is the partition of A in which the parts are the intersections $R_s \cap R_t'$ of some part R_s of $\boldsymbol{R}$ and some part R_t' of R' for which R_s and R_t' are not disjoint.

Both ordered and unordered intersections can be visualized as a block matrix in which the blocks are formed by the subsets of rows and columns corresponding to the parts of the ordered partitions $\boldsymbol{R}'$ and $\boldsymbol{R}$, respectively (see Table 1). Of course, the blocks of the intersections are only partially ordered so that, for example, blocks $R_2' \cap R_3$ and $R_3' \cap R_2$ are not comparable. However, a linear order can be imposed naturally by ordering the blocks first by rows and then by columns, so that any block of the first row precedes the blocks in all other rows. This is the so-called lexicographic product $\boldsymbol{R}' \times \boldsymbol{R}$ introduced in (Mirkin 1979). Similarly, an alternative lexicographic product $\boldsymbol{R} \times \boldsymbol{R}'$ is defined by ordering blocks first by columns and then by rows. Curiously, in the ordered series $\boldsymbol{R}'$, $\boldsymbol{R}' \times \boldsymbol{R}$, $\boldsymbol{R} \times \boldsymbol{R}'$, and $\boldsymbol{R}$, the middle term of each triplet is between the other two (Mirkin 1979). A similar statement holds for the corresponding relations ρ', $\rho' \times \rho$, $\rho \times \rho'$, and ρ.

3 Matrix Representation of Preference Relations and Distance Between Them

3.1 Correlation by Spearman and Kendall

Consider the Spearman rank correlation, that is, the Pearson correlation coefficient between ranks taken as numerical values. To deal with the case of tied rankings, each element of an equivalence class of indifference is assigned with the average within-class rank. The average rank of the elements in part R_s of the tied ranking $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$ is $L + (|R_s| + 1)/2$, where L is the cardinality of $R_1 \cup R_2 \cup \ldots \cup R_{s-1}$, and $|\cdot|$ denotes the number of elements in a set. The *Kendall rank correlation* is based on the representation of tied rankings on A by $N \times N$ matrices. Given a tied ranking $\boldsymbol{R}$ and the corresponding preference relation $\rho = P \cup E$, we now define a skew-symmetric matrix $\mathbf{K} = (k_{ij})$, for $i, j \in A$, such that $k_{ij} = 1$ if $(i, j) \in P$, $k_{ij} = 0$ if $(i, j) \in$

Table 1 The intersection of two tied rankings $R'\cap R$, where R' relates to rows and R to columns. It is assumed that the rows and columns are permuted according to the rankings R' and R, respectively

	R_2	R_2	R_3	R_4	R_5
R_1'	$R_1'\cap R_1$	$R_1'\cap R_2$	$R_1'\cap R_3$	$R_1'\cap R_4$	$R_1'\cap R_5$
R_2'	$R_2'\cap R_1$	$R_2'\cap R_2$	$R_2'\cap R_3$	$R_2'\cap R_4$	$R_2'\cap R_5$
R_3'	$R_3'\cap R_1$	$R_3'\cap R_2$	$R_3'\cap R_3$	$R_3'\cap R_4$	$R_3'\cap R_5$
R_4'	$R_4'\cap R_1$	$R_4'\cap R_2$	$R_4'\cap R_3$	$R_4'\cap R_4$	$R_4'\cap R_5$

E, and $k_{ij} = -1$, if $(j, i) \in P$. The Kendall rank correlation coefficient between R and R' is the correlation coefficient between their Kendall matrices, $\mathbf{K}$ and $\mathbf{K}'$, considered as vectors in an N^2-dimensional space. This is compatible with the non-quantitative nature of tied rankings, especially since the mean of a skew-symmetric matrix is always 0.

It should be noted that, soon after the Kendall matrix was defined, a somewhat similar skew-symmetric representation for quantitative features was proposed by Daniels (1944), who proved that, given a quantitative feature x on A, the matrix $\mathbf{X} = (x_{ij})$, where $x_{ij} = x_i - x_j$, can be used to represent the feature in statistical computations. For example, the inner product of the matrices $\mathbf{X}$ and $\mathbf{X}'$ corresponding to features $\mathbf{x}$ and $\mathbf{x}'$ is proportional to the inner product of $\mathbf{x}$ and $\mathbf{x}'$ after they have been centered by subtracting their means, viz. $<\mathbf{X}, \mathbf{X}'> = 2N < \mathbf{x} - m(\mathbf{x})$, $\mathbf{x}' - m(\mathbf{x}')>$, where $m(\mathbf{x})$ is the mean of $\mathbf{x}$. This implies that the correlation coefficient between $\mathbf{X}$ and $\mathbf{X}'$ is equal to the correlation coefficient between $\mathbf{x}$ and $\mathbf{x}'$. Therefore, the Spearman rank correlation coefficient can also be defined as the Pearson correlation coefficient between the corresponding matrices of rank differences (r_{ij}), where $r_{ij} = r_i - r_j$. The Kendall matrix is then just the matrix of signs in the Daniels matrix $\mathbf{X} = (x_{ij})$, where $x_{ij} = x_i - x_j$. The past century saw intense work on the analysis of the relationship between the Kendall and Spearman correlations; these were proven to be asymptotically equivalent under conventional statistical assumptions (see, for example, Lehmann and D'Abrera 2006).

3.2 Kemeny Distance

Rather than defining an ad hoc distance measure, Kemeny formulated four axioms that should hold for any acceptable distance measure $d(R, R')$ between rankings R and R'. These axioms require that the acceptable distance measures should:

A1. Be a mathematical metric, that is, have the following properties:

(a) Symmetry: $d(R, R') = d(R', R)$;
(b) Non-negativity and definiteness: $d(R, R') \geq 0$ and $d(R, R') = 0$ if and only if $R = R'$;
(c) Strict triangle inequality: for any rankings R, R' and R'', $d(R, R'') \leq d(R, R') + (R', R'')$; moreover, equality holds if and only if R' is between R and R''.

A2. If R' is obtained from R by a permutation of the set A and S' from S by the same permutation, then $d(R', S') = d(R, S)$.

To formulate the next axiom, let us say that a subset $B \subset A$ is a *segment* of a tied ranking R if its complement $A - B \neq \varnothing$ and each element $i \in A - B$ either precedes all the elements of B or is situated after all the elements of B. The tied ranking R restricted to a segment B will be denoted by R_B.

A3. If $\mathbf{R}$ and $\mathbf{R}'$ coincide on $A - B$ and B is a segment of both $\mathbf{R}$ and $\mathbf{R}'$, then $d(\mathbf{R}, \mathbf{R}') = d(\mathbf{R}_B, \mathbf{R}_B')$.

A4. Unit of scale: The minimum positive distance is equal to 1.

Kemeny proved that the only distance satisfying all four axioms is the L_1-metric between the corresponding skew-symmetric Kendall matrices divided by 2 (Kemeny 1959), namely:

$$kd(\mathbf{R}, \mathbf{R}') = \frac{1}{2} \sum_{i,j \in A} |k_{ij} - k'_{ij}| \tag{2}$$

We see from (2) that the pairs of elements (i, j) in A can be divided into three subsets:

(a)   those contributing 1 to $kd(\mathbf{R}, \mathbf{R}')$: pairs (i, j) such that i precedes j in either $\mathbf{R}$ or $\mathbf{R}'$ while j precedes i in the other;
(b)   those contributing ½ to $kd(\mathbf{R}, \mathbf{R}')$: pairs (i, j) such that i and j are indifferent in either $\mathbf{R}$ or $\mathbf{R}'$ while one precedes the other in the other ranking;
(c)   those contributing 0 to $kd(\mathbf{R}, \mathbf{R}')$: pairs (i, j), that are similarly related in both rankings— either i precedes j, or j precedes i, or i and j are indifferent.

In Fagin et al. (2006), the family of so-called Kendall metrics is defined using the rules (a), (b), and (c), modified by replacing the contribution ½ in rule (b) by some constant α, $½ \leq \alpha \leq 1$. Therefore, the Kemeny distance between rankings is just the Kendall metric for $\alpha = ½$.

3.3 The Mismatch Distance Between Binary Relations and the Corresponding Binary Matrices

Binary relations considered as subsets of the Cartesian product $A \times A$ may be compared using any of the many measures of dissimilarity between subsets that have been introduced over the years (Baulieu 1989; Snijders et al. 1990; Morlini and Zani 2012). One particularly simple measure is the number of pairs for which they differ, the so-called mismatch distance, i.e., the number of pairs in their symmetric difference:

$$d\left(\rho, \rho'\right) = \left|\left(\rho - \rho'\right) \cup \left(\rho' - \rho\right)\right| \tag{3}$$

The geometric spaces of ordered and unordered partitions with respect to this distance have been described at length in Mirkin (1979). The mismatch distance between unordered partitions was described in earlier publications by B. Mirkin in Russian from 1969 onwards (see, for example, Mirkin and Cherny 1972); it is sometimes referred to as Mirkin's distance (Meilă 2007).

We note that $d(\rho, \rho')$ is a metric on the space of all binary relations on A and satisfies Axiom A1, including the strict triangle inequality, even for binary relations that do not correspond to tied rankings.

The mismatch distance can easily be translated into a distance between $N \times N$ matrices. Given a binary relation $\rho \in A \times A$, we define its binary matrix $\mathbf{r} = (r_{ij})$ by:

$$r_{ij} = \begin{cases} 1 \text{ if } (i,j) \in \rho \\ 0 \text{ if } (i,j) \notin \rho \end{cases}$$

Then the mismatch distance between R and R' is the mismatch (Hamming) distance between the corresponding binary relations ρ and ρ', and is thus given by:

$$d\left(R, R'\right) = d\left(\rho, \rho'\right) = \left|\left(\rho - \rho'\right) \cup \left(\rho' - \rho\right)\right| = \sum_{i,j \in A} |r_{ij} - r'_{ij}| = \sum_{i,j \in A} \left(r_{ij} - r'_{ij}\right)^2 \quad (4)$$

The right-hand equality allows the original L_1-distance to be transformed into the square of the more conventional, Euclidean or L_2-distance because the absolute differences are either 1 or 0.

Obviously, the mismatch distance (4) is much simpler than the Kemeny distance (2) because the only possible non-zero contribution to $d(R, R')$ by an ordered pair (i, j) is 1, and this only occurs when j precedes i in one of the tied rankings but not in the other ranking. This happens when $r_{ij} = 0$ and $r_{ii}' = 1$ or, vice versa, $r_{ij}' = 0$ and $r_{ij} = 1$. It may therefore be somewhat of a surprise that these two distance measures are, in fact, equal.

Theorem 4 The Kemeny distance (2) is equal to the mismatch distance (4).

Proof We first analyze the contributions of pairs of elements $i, j \in A$ to the Kemeny distance between R and R' depending on their relative positions in the rankings R and R'; the various cases are shown in Table 2. We note that the contribution of the pair (j, i) is exactly the same as that of the pair (i, j).

Now, we need to take into account a subtle difference between the concepts of ranking and preference relation. The Kemeny distance is between two rankings—it records disagreements in the relative positions between a pair of elements in the two rankings; the symmetry between i and j accounts for the factor ½ in the expression (2) for the Kemeny distance.

In contrast, the mismatch distance is between binary relations and counts the disagreements between the relations in respect of ordered pairs of elements. We, therefore, must distinguish between the ordered pair (i, j) and the inverse pair (j, i), relative to the corresponding relation, ρ or ρ'. The various cases of the contributions to the mismatch distance are shown in Tables 3 and 4, respectively.

Returning to the analysis of the interrelation between two elements $i, j \in A$, we need to combine Tables 3 and 4 by adding them, which produces Table 5.

If we double the values in Table 2 to account for both ordered pairs (i, j) and (j, i), we observe that the resulting entries are identical to those in Table 5, which completes the proof.

Consider a simple example where A consists of three elements, 1, 2, and 3 that are linearly ordered in R and all tied in R', so that $R = (\{1\}, \{2\}, \{3\})$ and $R' = (\{1, 2, 3\})$. Their respective Kendall matrices are

$$k = \begin{array}{ccc} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{array} \text{ and } k' = \begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \text{ so that the Kemeny distance } kd\left(R, R'\right) = 6/2 = 3.$$

Table 2 The contribution of a pair $(i, j) \in A \times A$ to the Kemeny distance (2) between R and R'

		R		
	Cases	i precedes j	i and j are indifferent	j precedes i
R'	i precedes j	0	½	1
	i and j are indifferent	½	0	½
	j precedes i	1	½	0

Table 3 The contribution of the ordered pair $(i, j) \in A \times A$ to the mismatch distance (4) between R and R'

	Cases	R		
		i precedes j	i and j are indifferent	j precedes i
R'	i precedes j	0	0	1
	i and j are indifferent	0	0	1
	j precedes i	1	1	0

On the other hand, their respective weak order matrices are

$$r = \begin{matrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{matrix} \quad \text{and} \quad r' = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}, \quad \text{so that the mismatch distance } d\left(R, R'\right) = 3 \text{ as well}.$$

3.4 The Mismatch Distance Expressed in Terms of the Contingency Table

Although the following results can be established directly, we now rely on Axiom A1(c), which states that $d(R, R') = d(R, R'') + d(R'', R')$ if and only if R'' is between R and R', that is $\rho \cap \rho' \subseteq \rho'' \subseteq \rho \cup \rho'$ for the corresponding preference relations. By virtue of (4), we may use $d(R, R')$ and $d(\rho, \rho')$ interchangeably.

Let us first consider a tied ranking R and its reverse R^{-1}. Obviously, $\rho \cap \rho^{-1} = E$, where E is the indifference relation of R, which is an equivalence relation, stripped of all ranking information. Similarly, $\rho \cup \rho^{-1} = U$, the universal relation $U = A \times A$, which contains all possible ordered pairs of elements of A. Both E and U are, therefore, between ρ and ρ^{-1} for any weak order ρ.

Let N_s be the number of elements in part R_s of the tied ranking $R = (R_1, R_2, \ldots, R_p)$. Then the mismatch distance between U and E is easily seen to be

$$d(E, U) = N^2 - \Sigma_s N_s^{\ 2}, \tag{5}$$

since the first term on the right is the number of ones in the binary matrix of U and the second term is the number of ones in the binary matrix of E.

Curiously, the mismatch distance between U and R itself is exactly half the distance in (5).

Theorem 5 The mismatch distance between a tied ranking R and the universal relation U is given by

$$d(R, U) = 1/2 \left(N^2 - \Sigma_s N_s^{\ 2}\right). \tag{6}$$

Proof We first notice that $d(R, U) = d(R^{-1}, U)$ and $d(R, E) = d(R^{-1}, E)$. Indeed, neither U nor E depend on the ranking information in R, and, moreover, the number of pairs in ρ and ρ^{-1}, which is the number of ones in their respective matrices r and r^{-1}, is the same. Since both U and E are between R and R^{-1}, we have:

Table 4 The contribution of the ordered pair $(j, i) \in A \times A$ to the mismatch distance (4) between R and R'

	Cases	R		
		i precedes j	i and j are indifferent	j precedes i
R'	i precedes j	0	1	1
	i and j are indifferent	1	0	0
	j precedes i	1	0	0

Table 5 Summary contribution of the ordered pairs (i, j) and (j, i) to the mismatch distance (4)

	Cases	i precedes j	i and j are indifferent	j precedes i
R				
R'	i precedes j	0	1	2
	i and j are indifferent	1	0	1
	j precedes i	2	1	0

$$d\left(\boldsymbol{R}, \boldsymbol{R}^{-1}\right) = d(\boldsymbol{R}, U) + \mathrm{d}\left(U, \boldsymbol{R}^{-1}\right) = 2d(\boldsymbol{R}, U) \text{ and}$$

$$d\left(\boldsymbol{R}, \boldsymbol{R}^{-1}\right) = d(\boldsymbol{R}, E) + d\left(E, \boldsymbol{R}^{-1}\right) = 2d(\boldsymbol{R}, E).$$

This implies that $d(\boldsymbol{R}, U) = d(\boldsymbol{R}, E)$. So, since $\boldsymbol{R}$ is between E and U, $d(E, U) = d(E, \boldsymbol{R}) + d(\boldsymbol{R}, U) = 2d(\boldsymbol{R}, U)$. Equation (6) now follows from this and (5), which completes the proof.

Corollary 1 The distance $d(\boldsymbol{R}, \boldsymbol{R}^{-1})$ is equal to $d(E, U)$ given by (5), whereas the distance $d(\boldsymbol{R}, E)$ is equal to $d(\boldsymbol{R}, U)$, given by (6).

Proof These results are included in the proof of Theorem 5.

Now, we are in a position to prove a formula for the mismatch distance between a tied ranking $\boldsymbol{R}$ and its arbitrary refinement $\boldsymbol{R}'$. Like the previous results in this subsection, this does not depend on the ranking information.

Theorem 6 The mismatch distance between a tied ranking $\boldsymbol{R} = (R_1, R_2, \ldots, R_p)$ and an arbitrary refinement $\boldsymbol{R}' = (R_1', R_2', \ldots, R_q')$ of $\boldsymbol{R}$, where $q > p$, is given by

$$d\left(\boldsymbol{R}, \boldsymbol{R}^{`}\right) = \frac{1}{2}\left(\Sigma_s N_s{}^2 - \Sigma_t N_t^{`}{}^2\right), \tag{7}$$

where N_s and N'_t are the numbers of elements in the parts R_s of $\boldsymbol{R}$ and R_t' of $\boldsymbol{R}'$, respectively.

Proof Indeed, since $\boldsymbol{R}$ is between $\boldsymbol{R}'$ and U, we have $d(\boldsymbol{R}', U) = d(\boldsymbol{R}', \boldsymbol{R}) + d(\boldsymbol{R}, U)$, so $d(\boldsymbol{R}', \boldsymbol{R}) = d(\boldsymbol{R}', U) - d(\boldsymbol{R}, U)$. Both distances $d(\boldsymbol{R}', U)$ and $d(\boldsymbol{R}, U)$ are determined by Eq. (6), adjusted for the corresponding parts of $\boldsymbol{R}'$ and $\boldsymbol{R}$, respectively. This immediately yields (7), completing the proof.

Consider now two ordered partitions, $\boldsymbol{R}$ and $\boldsymbol{R}'$, and their lexicographic products $\boldsymbol{R} * \boldsymbol{R}'$ and $\boldsymbol{R}' * \boldsymbol{R}$. We shall show that the entire ranking component contributing to the distance between $\boldsymbol{R}$ and $\boldsymbol{R}'$ is accounted for by the distance between $\boldsymbol{R} * \boldsymbol{R}'$ and $\boldsymbol{R}' * \boldsymbol{R}$. First, consider the intersection $\boldsymbol{R} \cap \boldsymbol{R}'$, as presented in Table 1. Letting $N_{st} = |R_s \cap R'_t|$, for $s = 1, 2, \ldots, p$ and $t = 1, 2, \ldots, q$, denote the numbers of elements in the parts of the intersection, we can present these cardinalities as the so-called contingency table, or cross-classification, between $\boldsymbol{R}$ and $\boldsymbol{R}'$ – a popular concept in statistics (see, for example, Mirkin 1979, 2012).

The distance between $\boldsymbol{R} * \boldsymbol{R}'$ and $\boldsymbol{R}' * \boldsymbol{R}$ is equal to half of the total of the products of the cardinalities of those parts in the intersection $\boldsymbol{R} \cap \boldsymbol{R}'$ for which the orderings in $\boldsymbol{R}$ and $\boldsymbol{R}'$ are contradictory:

$$d(\boldsymbol{R} * \boldsymbol{R}', \boldsymbol{R}' * \boldsymbol{R}) = \frac{1}{2}\Sigma_{s > s'}\Sigma_{t < t'} N_{st} N_{s't'} \tag{8}$$

Considering the rankings R and R' as unordered partitions, denoted above by $\check{R}$ and $\check{R}'$, respectively, the mismatch distance between the corresponding equivalence relations, E and E', can be expressed as

$$d\left(E, E'\right) = \sum_s N_s^2 + \sum_t N_t'^2 - 2\sum_{s,t} N_{st}^2 \tag{9}$$

where N_s, N_t', and N_{st} are, as above, the numbers of elements in parts R_s of R, R'_t of R' and $R_s \cap R'_t$ of $R \cap R'$, respectively.

> **Theorem 7**. The mismatch distance between tied rankings R and R' can be decomposed into ranking and equivalence parts as follows:
>
> $$d\left(R, R'\right) = \frac{1}{2} d\left(E, E'\right) + d\left(R \times R', R' \times R\right). \tag{10}$$

Proof Consider the corresponding binary relations ρ, ρ', and $\rho \cap \rho'$. Since the intersection $\rho \cap \rho'$ is between ρ and ρ', $d(\rho, \rho') = d(\rho, \rho \cap \rho') + d(\rho \cap \rho', \rho')$. On the other hand, $\rho \times \rho'$ is between $\rho \cap \rho'$ and ρ, and $\rho' \times \rho$ is between $\rho \cap \rho'$ and ρ', so $d(\rho, \rho \cap \rho') = d(\rho, \rho \times \rho') + d(\rho \times \rho', \rho \cap \rho')$, and $d(\rho \cap \rho', \rho') = d(\rho \cap \rho', \rho' \times \rho) + d(\rho' \times \rho, \rho')$. But $\rho \cap \rho'$ is between $\rho \times \rho'$ and $\rho' \times \rho$, so $d(\rho \times \rho', \rho' \times \rho) = d(\rho \times \rho', \rho \cap \rho') + d(\rho \cap \rho', \rho' \times \rho)$. Substituting these in the equation $d(\rho, \rho') = d(\rho, \rho \cap \rho') + d(\rho \cap \rho', \rho')$, we obtain

$$d\left(\rho, \rho'\right) = d\left(\rho, \rho \times \rho'\right) + d\left(\rho \times \rho', \rho' \times \rho\right) + d\left(\rho \times \rho', \rho'\right). \tag{11}$$

Since $\rho \times \rho'$ is a refinement of ρ, and $\rho' \times \rho$ is a refinement of ρ', $d(\rho, \rho \times \rho') = \frac{1}{2}(\sum_s N_s^2 - \sum_{s,t} N_{st}^2)$ and $d(\rho' \times \rho, \rho') = \frac{1}{2}(\sum_t N_t'^2 - \sum_{s,t} N_{st}^2)$ by Theorem 6. This implies, by (9), that $d(\rho, \rho \times \rho') + d(\rho' \times \rho, \rho') = \frac{1}{2}d(E, E')$. Together with (11), this completes the proof.

4 Consensus Among Tied Rankings

4.1 Consensus Matrix

Let us consider the concept of consensus ranking formulated by John Kemeny as follows. Given a set of tied rankings R^1, R^2, …, R^n, find a ranking R, referred to by Kemeny as a median, that minimizes the summary distance $D(R) = \sum_m kd(R, R^m)$. Such a problem, for unordered partitions, has been widely accepted as a consensus rule in the literature on data analysis and machine learning (see, for example, Guénoche 2011).

Given a set of tied rankings R^1, R^2, …, R^n, let us define a ranking consensus matrix $\mathbf{C} = (c_{ij})$ as follows: for any pair (i, j), c_{ij} is the number of those rankings R^m, $1 \le m \le n$, in which i either precedes j or is indifferent to j. This means that $\mathbf{C} = \sum_m \mathbf{r}^m$, where $\mathbf{r}^m$ is the binary matrix of the binary relation ρ^m corresponding to R^m, for $1 \le m \le n$.

This matrix can be used to obtain an explicit criterion for the concept of median.

Theorem 8 A ranking $R = (R_1, R_2, …, R_p)$ is a median of the set of tied rankings R^1, R^2, …, R^n if and only if it maximizes

$$F(\mathbf{R}) = \sum_{i,j \in A} \left(c_{ij} - \frac{n}{2} \right) \mathbf{r}_{ij} = \sum_{s=1}^{p} \sum_{t \geq s} \sum_{i \in R_s} \sum_{j \in R_t} \left(c_{ij} - \frac{n}{2} \right) \tag{12}$$

with respect to a pre-specified set of admissible rankings $\mathbf{R}$.

Proof By Theorem 4, we may use the mismatch distance in the definition of median, and thus $D(\mathbf{R}) = \sum_m d(\mathbf{R}, \mathbf{R}^m)$. Substituting the right-most expression in (4) for $d(\cdot, \cdot)$, we get

$$D(\mathbf{R}) = \sum_{m=1}^{n} \sum_{i,j \in A} \left(\mathbf{r}_{ij} - \mathbf{r}_{ij}^m \right)^2 = \sum_{i,j \in A} \sum_{m=1}^{n} \left(\mathbf{r}_{ij} - \mathbf{r}_{ij}^m - 2\mathbf{r}_{ij}\mathbf{r}_{ij}^m \right) = \sum_{i,j \in A} \left(n\mathbf{r}_{ij} + c_{ij} - 2\mathbf{r}_{ij}c_{ij} \right),$$

since squaring does not change the binary 0/1 values. The last expression follows by the definition of the consensus matrix $\mathbf{C}$. Rearranging this, we obtain $D(\mathbf{R}) = \sum_{i,j \in A} c_{ij} - 2 \sum_{i,j \in A} \left(c_{ij} - \frac{n}{2} \right) \mathbf{r}_{ij}$, which is equal to a constant minus $2F(\mathbf{R})$. Therefore, to minimize $D(\mathbf{R})$, we need to maximize $F(\mathbf{R})$, which completes the proof.

We note that the subtracted value $n/2$ can be moved out of the summation in (12), so that the expression for $F(\mathbf{R})$ becomes

$$F(\mathbf{R}) = \sum_{s=1}^{p} \sum_{t \geq s} \sum_{i \in R_s} \sum_{j \in R_t} c_{ij} - \frac{n}{4} \left(N^2 + \sum_{s=1}^{p} N_s^2 \right), \tag{13}$$

where N_s is the cardinality of R_s, $1 \leq s \leq p$. This follows because $\sum_{i,j \in A} \mathbf{r}_{ij}$ is the number of pairs $(i, j) \in \rho$, that is, the cardinality of the binary relation ρ corresponding to $\mathbf{R}$, which is given by

$$|\rho| = \frac{1}{2} \left(N^2 + \sum_{s=1}^{p} N_s^2 \right). \tag{14}$$

To prove this, we recall from the previous section that $\rho \cap \rho^{-1} = E$ and $\rho \cup \rho^{-1} = U$. Since $|\rho| = |\rho^{-1}|$, using the fact that $|\rho \cup \rho^{-1}| = |\rho| + |\rho^{-1}| - |\rho \cap \rho^{-1}|$, it follows that $|\rho| = \frac{1}{2}(|U| + |E|)$. Equation (14) now follows immediately from the remark following Eq. (5).

Since $nN^2/4$ is a constant, maximization of $F(\mathbf{R})$ in (13) is equivalent to the modified criterion of maximizing

$$G(\mathbf{R}, a) = \sum_{s=1}^{p} \sum_{t \geq s} \sum_{i \in R_s} \sum_{j \in R} c_{ij} - a \sum_{s=1}^{p} N_s^2, \tag{15}$$

where $a = n/4$.

The criterion (12), or equivalently (15), is an extension of the following criterion of linear ordering, sometimes referred to as the triangulation of a square matrix, which was popular a few decades ago: find a simultaneous permutation of the rows and columns that maximizes the sum of the above-diagonal entries in the matrix (see, for example, Grötschel et al. 1983; Charon and Hudry 2007). Obviously, for a linear ordering all parts R_s are singletons and $p = N$ in (15), and thus the right-most term in (15) is a constant. In this case, the solution does not depend on the value $n/2$ that is subtracted from all the entries in (12).

Curiously, the subtracted term on the right in (15) does not depend on the ordering information. It depends just on the indifference relation E or, more precisely, the distribution of elements over tied parts of the median ranking.

The expression $\sum_{s=1}^{p} N_s^2$ for the cardinality of $|E|$ is not unusual in statistics of categorical variables. To explain its meaning, consider the relative frequency $f_s = N_s/N$ of part R_s, and the

entire frequency profile $f(\boldsymbol{R}) = (f_1, f_2, \ldots, f_p)$ of an arbitrary p-part partition $\boldsymbol{R}$. The inner product, $c(\boldsymbol{R}) = \,<f(\boldsymbol{R}), f(\boldsymbol{R})> = \Sigma_s f_s^2$, may be referred to as the concentration index. Its maximum value 1 is attained for the profile in which all the frequencies are 0, except one that is equal to unity. Its minimum value 1/p is attained for the uniform profile in which all the frequencies are equal, so $f_s = 1/\text{p}$ for all s, $1 \le s \le p$. Obviously, $|E| = N^2 c(\boldsymbol{R})$. A precise meaning of $c(\boldsymbol{R})$ can be given in terms of the so-called proportional prediction rule. The proportional prediction rule is a classifier defined according to profile $f(\boldsymbol{R})$. A classifier assigns a category s, $1 \le s \le p$, to any object randomly supplied, according to the distribution $f(\boldsymbol{R})$. The proportional prediction classifier assigns the categories randomly, each category s with probability f_s. It is easy to see that the average accuracy of this classifier is $c(\boldsymbol{R})$. Its complement to unity, $\delta = 1 - c(\boldsymbol{R})$, is well known in social economics as the Gini index, a measure of inequality of social groups in which f_s is the proportion of the sth income group in the population.

Since the concentration part in (15) is subtracted, we can safely claim that criterion (15), other things being equal, leads to a more uniform median ranking than it would have had nothing been subtracted from c_{ij} in (12). In fact, it can be proved that, when the weight a of the subtracted value in (15) is increased above n/4, the concentration of a ranking $\boldsymbol{R}$ that maximizes the expression corresponding to (15) can only decrease. Indeed, the following stronger statement holds.

Theorem 9 If $\boldsymbol{R}$ and $\boldsymbol{R'}$ are maximizers of $G(\boldsymbol{R}, a)$ in (15) for $a = a_1$ and $a = a_2$, respectively, and $a_2 > a_1$ then $c(\boldsymbol{R'}) \le c(\boldsymbol{R})$.

Proof From (15), $G(\boldsymbol{R}, a) = G(\boldsymbol{R}, 0) - aN^2 c(\boldsymbol{R})$ for any $\boldsymbol{R}$ and a. Since $\boldsymbol{R}$ maximizes $G(\boldsymbol{R}, a_1)$ and $\boldsymbol{R'}$ maximizes $G(\boldsymbol{R}, a_2)$, we have $G(\boldsymbol{R'}, a_1) \le G(\boldsymbol{R}, a_1)$ and $G(\boldsymbol{R}, a_2) \le G(\boldsymbol{R'}, a_2)$. Therefore, $G(\boldsymbol{R'}, 0) - a_1 N^2 c(\boldsymbol{R'}) \le G(\boldsymbol{R}, 0) - a_1 N^2 c(\boldsymbol{R})$ and $G(\boldsymbol{R}, 0) - a_2 N^2 c(\boldsymbol{R}) \le G(\boldsymbol{R'}, 0) - a_2 N^2 c(\boldsymbol{R'})$.

Combining these two inequalities, we obtain

$a_1 N^2 c(\boldsymbol{R}) - a_1 N^2 c(\boldsymbol{R'}) \le G(\boldsymbol{R}, 0) - G(\boldsymbol{R'}, 0) \le a_2 N^2 c(\boldsymbol{R}) - a_2 N^2 c(\boldsymbol{R'})$, so $(a_2 - a_1) N^2 c(\boldsymbol{R'}) \le (a_2 - a_1) N^2 c(\boldsymbol{R})$. Since $a_2 > a_1$, it follows that $c(\boldsymbol{R'}) \le c(\boldsymbol{R})$, which completes the proof.

This result supports the view that the second term in (15) is a regularizer of the distribution of elements in the parts of the sought ordered partition $\boldsymbol{R}$: the greater the weight a, the greater the number of (non-empty) parts and the more uniform the distribution of elements among the parts.

4.2 The Muchnik Test Over Likert Scales

Ranking practitioners have noticed that the concept of median is not that sensitive to their interest in a few leaders rather than an entire consensus ranking (Cook 2006). We focus here upon an even less desirable property: insufficient sensitivity to the differences in similar opinions.

Let us consider a test example, akin to the so-called Muchnik test in consensus clustering (Mirkin 2012), in the context of the so-called Likert scale (Likert 1932; Allen and Seaman 2007). This scale is applied when an individual cannot reproduce an entire ranking. A psychologist then specifies a number of attributes, each of which splits the ordering in question into two complementary fragments—the beginning and the end. For each object or observation, the psychologist asks the individual, in respect of a specific attribute, whether the object falls within the beginning or end part of the scale. Mathematically, the situation can be described as follows.

There is a ranking with p tied parts, $\mathbf{R} = (R_1, R_2, \ldots, R_p)$. This ranking is used to produce p-1 binary rankings $S^1, S^2, \ldots, S^{p-1}$. Each binary ranking S^t consists of just two parts, so $S^t = (U_t, V_t)$ for $t = 1, 2, \ldots, p$-1. The first part U_t is the set-theoretic union of the first t parts of $\mathbf{R}$, so $U_t = \overset{t}{\underset{m=1}{\cup}} R_m$, and the second part V_t is the union of the remaining parts of $\mathbf{R}$, so $V_t = A - U_t$. These binary rankings completely determine the underlying ranking $\mathbf{R}$. Therefore, we may pose the problem of finding a consensus ranking or median $\mathbf{S}$ for the set of binary rankings $S^1, S^2, \ldots, S^{p-1}$. In particular, will the underlying ordered partition $\mathbf{R}$ be a solution to this problem?

Figure 1 shows two cases representing the ranking consensus matrix for the set of binary rankings corresponding to a ranking with $p = 7$ tied parts, with the value $(p$-1$)/2 = 3$ subtracted from each of the entries. Rather than showing individual elements of A, we show just the block structure of the matrix, where the (s, t)-th block, $1 \leq s, t \leq p$, corresponds to the pairs $(i, j) \in R_s \times R_t$, with each entry having the constant value $c_{ij} - (p$-1$)/2$. There are only six negative blocks in the matrix. Of course, the number of negative entries (i, j) clearly depends on the sizes of the corresponding parts of $\mathbf{R}$. Let us consider, for the sake of simplicity, a case in which each part $R_1, R_2, \ldots, R_p$ of $\mathbf{R}$ contains the same number of elements. To maximize the criterion (12), we just need to minimize the sum of the entries below the diagonal. It is almost evident that the parts of $\mathbf{R}$ cannot be split in an optimal ranking $\mathbf{S}$—this, of course, can be proven formally. Therefore, a median $\mathbf{S}$ can be obtained by merging some parts of $\mathbf{R}$, that is, $\mathbf{R}$ should be a refinement of $\mathbf{S}$. The best option would, if it were possible, be to merge the parts in such a way that these negative blocks of entries, and only they, should be excluded from the upper part of the matrix in Fig. 1. However, this is not possible, because, however, the parts are aggregated, some positive entries must be present below the diagonal—more precisely, below a borderline delineating the aggregated parts. This is shown in Fig. 1a for the case in which the candidate ranking has three parts obtained by merging (i) R_1 and R_2; (ii) R_3, R_4, and R_5; and (iii) R_6 and R_7. The borderline between the entries included in and excluded from the sum in (12) is shown in bold. We can see there are six positive entries below the borderline, which almost cancel out the negative values. In this sense, the binary ranking S^4 that merges R_1, R_2, R_3, and R_4 into the first part of $\mathbf{S}$, and R_5, R_6, and R_7 into the second part of $\mathbf{S}$, as shown in Fig. 1b, is better as it excludes all the negative entries and only three positive entries. It is not difficult to prove that this ranking is optimal, and that another optimal ranking is the binary ranking S^3. Similarly, in the general case with a different number of equal-size parts in the underlying ranking $\mathbf{R}$, a binary ranking that splits $\mathbf{R}$ into two equal-size parts (or as equal as possible) is an optimal ranking.
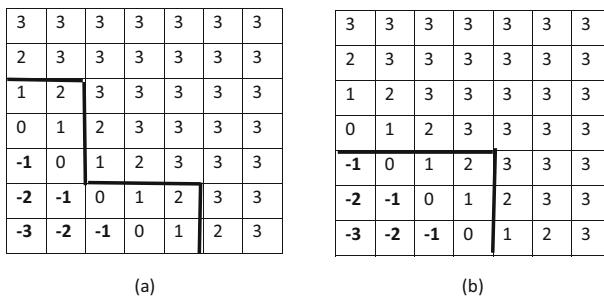
3	3	3	3	3	3	3
2	3	3	3	3	3	3
1	2	3	3	3	3	3
0	1	2	3	3	3	3
-1	0	1	2	3	3	3
-2	-1	0	1	2	3	3
-3	-2	-1	0	1	2	3

(a)

3	3	3	3	3	3	3
2	3	3	3	3	3	3
1	2	3	3	3	3	3
0	1	2	3	3	3	3
-1	0	1	2	3	3	3
-2	-1	0	1	2	3	3
-3	-2	-1	0	1	2	3

(b)

Fig. 1 Values $c_{ij} - (p$-1$)/2$ for the blocks of the consensus matrix for the Likert scale consensus problem when $p = 7$. The bold line represents the boundary separating entries summed in criterion (12), i.e., those above and to the right of it, from the rest. **a** shows the case in which the seven parts are merged into three aggregated parts: (i) the first two parts, (ii) the next three parts, and (iii) the last two parts. **b** shows the case in which the seven parts are merged into two aggregated parts: (i) the first four parts and (ii) the last three parts

As we can see, the median rule for the Likert scale cannot reproduce the original ranking R when $p > 2$. The median S in this case is just a coarser binary version of R, which is a very rough model of the original.

5 Conclusion

Human intuition handles summation of positive numbers rather well. An issue emerges when some of the numbers are negative. This paper can be looked at as an attempt to find some structure in the -1 entries in the Kendall matrices occurring in the formula for the Kemeny distance between tied rankings. These entries appear whenever a pair of elements, i and j, are inversely related so that j precedes i in a ranking. First, we showed in Theorem 4 that the Kemeny distance can be expressed in terms of the mismatch distance between the preference relations (weak orders) corresponding to the tied rankings, in which no negative entries appear. The mismatch distance is defined in terms of the 0–1 matrices of weak orders, rather than the Kendall matrices of the rankings, containing entries 1, 0, and -1. Our next claim is that, in the problem of finding a consensus between tied rankings, all the negative items relate to the subtracted regularizer term in (15), an expression depending only on the distribution of the sizes of the parts in the consensus ranking, not on the precedence relation between them (see Theorems 8 and 9). The structure of the subtracted term relates the problem of finding a consensus tied ranking to the well-known linear ordering problem. Moreover, when applied to the issue of consensus ranking in the Likert scale case, the concept of Kemeny median, or consensus ranking, appears to be less sensitive than one would have hoped, resulting in a solution being one of the central Likert binary rankings rather than the hidden multi-part ranking.

Following this line of narrative, we have explicitly formulated properties of the Kemeny distance, especially those related to its decomposition into "ordered" and "unordered" parts (see Theorem 7) and its computation via the contingency table, popular in statistics (see (8), (9), and (10)).

Among possible directions for further research, two are quite straightforward. First, to try to solve numerically the problem of consensus ranking. For example, the additive structure in (15) suggests that one might first find an optimal linear ordering and then aggregate some of its parts to form a tied ranking. Second, the failure of the Muchnik test on Likert scales suggests some ways for formulating more sensitive criteria for consensus.

There are other approaches as well (see, for example, Amor and Martel 2014, Barthélemy, Leclerc and Monjardet 1986, Critchlow 1980/2012, Dwork et al. 2001, Emond and Mason 2002, Monjardet 1978, Nandi, Fa and Abu-Jamous 2015).

References

Aleskerov, F., & Monjardet, B. (2002). *Utility maximization, choice, and preference, Stud. Econ. Theory 16.* Berlin: Springer-Verlag.
Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress, 40*(7), 64.

Amor, S. B., & Martel, J. M. (2014). A new distance measure including the weak preference relation: application to the multiple criteria aggregation procedure for mixed evaluations. *European Journal of Operational Research, 237*(3), 1165–1169.

Barthélemy, J. P., Leclerc, B., & Monjardet, B. (1986). On the use of ordered sets in problems of comparison and consensus of classifications. *Journal of Classification, 3*(2), 187–224.

Baulieu, F. B. (1989). A classification of presence/absence based dissimilarity coefficients. *Journal of Classification, 6*(1), 233–246.

Charon, I., & Hudry, O. (2007). A survey on the linear ordering problem for weighted or unweighted tournaments. 4OR. *A Quarterly Journal of Operations Research, 5*(1), 5–60.

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

Critchlow, D.E. (1980). Springer-Verlag, Berlin. See also: Critchlow, D.E., 2012. Metric methods for analyzing partially ranked data (34). Springer Science & Business Media.

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika, 33*(2), 129–135.

Diaconis, F., & Graham, R. (1977). Spearman's footrule as a measure of disarray. *Journal of Royal Statistics Society, Series B, 39*, 262–268.

Dwork, C., Kumar, R., Naor, M. and Sivakumar, D. (2001) *Rank aggregation revisited*, http://static.cs.brown.edu/courses/csci2531/papers/rank2.pdf.

Emond, E. J., & Mason, D. W. (2002). A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis, 11*, 17–28.

Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., & Vee, E. (2006). Comparing partial rankings. *SIAM Journal of Discrete Mathematics, 20*(3), 628–648.

Grötschel, M., Jünger, M., & Reinelt, G. (1983). Optimal triangulation of large real world input-output matrices. *Statistische Hefte, 25*(1), 261–295.

Guénoche, A. (2011). Consensus of partitions: A constructive approach. *Advances in Data Analysis and Classification, 5*(3), 215–229.

Heiser, W. J., & D'Ambrosio, A. (2013). Clustering and prediction of rankings within a Kemeny distance framework. In B. Lausen, D. Van der Poel, & A. Ultsch (Eds.), *Algorithms from and for Nature and Life, Studies in Classification, Data Analysis, and Knowledge Organization* (pp. 19–31). Switzerland: Springer International.

Kemeny, J.G. (1959). *Mathematics without numbers,* Daedalus, *88, No. 4*. Quantity and Quality, pp 577–591.

Kemeny, J., & Snell, L. (1962). *Mathematical models in social sciences*. New-York: Blaisdell.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*, 81–93.

Lehmann, E. L., & D'Abrera, H. J. (2006). *Nonparametrics: statistical methods based on ranks*. New York: Springer.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 55.

Meilă, M. (2007). Comparing clusterings—an information based distance. *Journal of Multivariate Analysis, 98*(5), 873–895.

Mirkin, B. (1979) *Group Choice*. Halsted Press, Washington DC (Translated from Russian "Problema Gruppovogo Vybora", Nauka Physics-Mathematics, Moscow, 1974).

Mirkin, B. (2012). *Clustering: A data recovery approach*. Boka-Raton: CRC Press/Chapman & Hall.

Mirkin, B., & Cherny, L. (1972). Some properties of the partition space. In K. Bagrinovsky & E. Berland (Eds.), *Mathematical analysis of economic models III* (pp. 126–147). Novosibirsk: Institute of Economics of the Siberian Branch of the USSR's Academy of the Sciences (in Russian).

Monjardet, B. (1978). An axiomatic theory of tournament aggregation. *Mathematics of Operations Research, 3*(4), 334–351.

Morlini, I., & Zani, S. (2012). A new class of weighted similarity indices using polytomous variables. *Journal of Classification, 29*(2), 199–226.

Nandi, A.K., Fa, R. and Abu-Jamous, B. (2015). *Integrative cluster analysis in bioinformatics*. John Wiley & Sons.

Snijders, T. A., Dormaar, M., Van Schuur, W. H., Dijkman-Caes, C., & Driessen, G. (1990). Distribution of some similarity coefficients for dyadic binary data in the case of associated attributes. *Journal of Classification, 7*(1), 5–31.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*, 72–101.

Steele, K. & Stefánsson, H. O. (2015). *Decision theory, the stanford encyclopedia of philosophy* (Winter 2015 Edition). In E. N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/win2015/entries/decision-theory/>.

Xu, W., Hou, Y., Hung, Y. S., & Zou, Y. (2013). A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Signal Processing, 93*(1), 261–276.