Axiomatic aggregation of incomplete rankings

Erick Moreno-Centeno and Adolfo R. Escobedo

Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA

ABSTRACT

In many different applications of group decision-making, individual ranking agents or judges are able to rank only a small subset of all available candidates. However, as we argue in this article, the aggregation of these incomplete ordinal rankings into a group consensus has not been adequately addressed. We propose an axiomatic method to aggregate a set of incomplete rankings into a consensus ranking; the method is a generalization of an existing approach to aggregate complete rankings. More specifically, we introduce a set of natural axioms that must be satisfied by a distance between two incomplete rankings; prove the uniqueness and existence of a distance satisfying such axioms; formulate the aggregation of incomplete rankings as an optimization problem; and propose and test a specific algorithm to solve a variation of this problem where the consensus ranking does not contain ties; and show that the consensus ranking obtained by our axiomatic approach is more intuitive than the consensus ranking obtained by other approaches.

1. Introduction

Group decision-making settings where ranking agents or judges are unable to rank all candidates of a universal object set $V$ are ubiquitous. In organizational decision-making, the managerial hierarchy and departmental divisions may explicitly or implicitly prevent certain managers from ranking all alternatives (Goddard, 1983); in electoral candidate ranking, voting may be capricious when there is a high number of candidates (Levin and Nalebuff, 1995); in R&D sponsorship processes, decision-makers rank only a fraction of all proposed projects due to the submissions being numerous and diverse (Cook et al., 2007); in the National Science Foundation (NSF) proposal review process, review panel members judge a subset of all submissions. Practicality, feasibility, and judiciousness are among the most prominent reasons why a judge performs an evaluation of a proper subset of $V$ or an incomplete ranking instead of a full evaluation of all elements of $V$ or a complete ranking. Furthermore, incomplete rankings are often explicitly required in order to obtain meaningful results and to prevent over-ranking fatigue. In the context of eliciting consumer opinions, firms may prudently opt to restrict the number of choices per consumer to no more than a handful, since ranking capabilities rapidly deteriorate as the number of products increases (Gibbons, 1971). Moreover, in academic paper competitions such as the INFORMS George Nicholson Annual Student Paper Competition, submissions typically number in the hundreds; however, it is infeasible to assign each participating judge to rank more than even a handful of entries. This is due to the broad range of research areas the submissions encompass, the limited area of expertise of an individual judge, and the sheer effort required to properly evaluate the merit of an entry. In short, the motivations and applications of incomplete rankings vary widely. A critical point of emphasis is that in all of the aforementioned situations, as well as in the ensuing narrative, the fact that a judge does not rank a subset of objects does not provide any information at all about the judge’s preferences, or lack thereof, among his/her unranked objects or between his/her unranked objects and his/her ranked objects.

The axiomatic distance-based ranking aggregation method developed in this article is designed to solve the general problem of aggregating a collection of individual rankings of potentially distinct subsets of $V$ (possibly differing in size as well) into a judicious complete consensus evaluation. In other words, this approach provides the decision maker with the flexibility to assign a reasonable number of objects (perhaps randomly) to individual judges in order to limit ranking manipulation and bias and to prevent over-ranking fatigue. Despite the potential incompleteness and unevenness in size of the input rankings, however, the decision maker can still obtain an unbiased collective ranking, provided that certain input requirements are satisfied (see Section 3.3). Before proceeding, we clarify that throughout this article rankings refers to ordinal evaluations (e.g., first place, second place, etc.).

The ability of the proposed aggregation method to handle individual rankings of possibly different sizes and subsets of $V$ is critical as it is difficult/impractical to satisfy the requirement that every judge be assigned the same number of objects. Three reasons for this are limited judge expertise, conflict of interest, and overlap requirements for obtaining meaningful results (see Section 3.3). For example, in a Ph.D. student presentation competition that took place in the author’s home department, faculty members were assigned to judge
individual presentations based on their availability, area of expertise, and relationship to the presenters; the latter criterion is critical in preventing partisanship and conflict of interest in the evaluation process. Based on the advanced educational level of each participant, the presentations were judged principally on their technical content. This meant that assigning the same number of objects to each judge would have required either that some faculty members evaluate technical content outside their area of expertise or that some presenters be judged by their own advisor (not to mention that the scheduling problem involved in this process is difficult in itself). The more reasonable choice in this and other situations, therefore, is to assign an uneven number of objects to each judge, so that more critical constraints can be enforced.

In spite of the widespread incidence of incomplete rankings in group decision-making, axiomatic distance-based ranking aggregation methods (see, for example, Kemeny and Snell (1962), Bogart (1973, 1975), and Cook et al. (1986)) have either neglected to consider them explicitly or, as we shall argue, taken them into account in an inadequate manner. Moreover, when the axiomatic methods that do not allow for the possibility of incomplete rankings are reasonably adapted to consider them, they yield aggregate rankings or consensus rankings that are unintentionally partial toward particular types of solutions (this issue is explored in Section 2 and Section 5). The fidelity of a consensus ranking depends equally on its ability to derive the most representative evaluation from a collection of individual evaluations and on the reasonableness of its assumptions. Hence, in order to expand the breadth of distance-based ranking aggregation methods, this article develops an axiomatic incomplete-ranking framework as well as solution procedures to derive a corresponding consensus ranking.

There exist other ranking aggregation methodologies in addition to axiomatic distance-based approaches, and some of them handle incomplete rankings. Compared with the axiomatic method herein presented, these alternative methods fall into one of the following three cases: they produce biased and/or restrictive aggregate rankings (Case 1); they have limited scope (Case 2); or they solve a problem that is similar but significantly different to the one addressed in this work (Case 3). In particular, Case 3 includes the situation where the objects not ranked in an evaluation contain information about preference—for instance, when the unrated objects are assumed to be not as good as the ranked objects. In order to clearly distinguish our work from alternative aggregation methods, we discuss the approaches that fall under Case 1 in Section 2.2 and relegate those that fall under Cases 2 and 3 to Section B of the online supplement to this article. Finally, we point out that although the term partial rankings has been previously used as a synonym for incomplete rankings—such as in Critchlow (1985) or in González-Pachón and Romero (2001)—it commonly refers to rankings that contain ties;—see Fagin et al. (2006) or Bansal and Fernández-Baca (2009), for example. To avoid confusion, this article avoids the use of the term partial rankings altogether.

The main contribution of this article is to generalize the Kemeny–Snell axiomatic distance in the space of complete rankings (Kemeny and Snell, 1962) in order to define an axiomatic distance in the space of incomplete rankings. Specifically, this article (i) introduces a set of natural axioms that must be satisfied by a distance between two incomplete rankings; (ii) proves the uniqueness and existence of a distance satisfying such axioms; and (iii) shows that this distance is equivalent to the Kemeny–Snell distance on the subspace of complete rankings. Another contribution of the current work is that it demonstrates the inadequacy of other evaluation aggregation methods in obtaining a consensus of incomplete rankings and contrasting it with the heightened intuitiveness of the aggregate ranking obtained by the proposed method. In addition, this article provides a mathematical model and algorithms for finding a consensus ranking. Since the herein-defined distance is equivalent to the Kemeny–Snell distance on the subspace of complete rankings, the proposed algorithms can also be used in the case where the given rankings are complete. Thus, our methodology provides a unifying framework for ranking aggregation: it can be used to aggregate input rankings that are complete, incomplete, contain ties, or do not contain ties. We supplement this flexible aggregation framework with an optimization algorithm that is capable of finding the (no-ties) complete consensus ranking quickly for instances of practical size.

This article is organized as follows: Section 2 presents the literature review. Section 3 provides a precise definition of a consensus ranking (Section 3.1), develops the notation to represent incomplete rankings (Section 3.2), and states basic assumptions for aggregating incomplete rankings (Section 3.3). Section 4 develops the incomplete-ranking aggregation method. Specifically, it introduces the Kemeny–Snell complete-ranking aggregation model (Section 4.1); then, expanding from this model, it develops a set of natural axioms that must be satisfied by any distance between two incomplete rankings (Section 4.2), and it proves that these axioms lead to the existence of a unique distance function (Section 4.3). In Section 5, other evaluation aggregation approaches previously suggested in the literature are extended to handle incomplete rankings (Sections 5.1 to 5.3), and their consensus rankings are compared with those of the proposed method (Sections 5.4 to 5.6). Section 6 shows that finding a consensus ranking is NP-hard; then it proposes and tests an exact optimization methodology based on the implicit hitting set approach (Moreno-Centeno and Karp, 2013) to find a consensus ranking that does not contain ties. Finally, Section 7 provides the concluding remarks, and Section 8 discusses possible avenues of future research.

2. Literature review

2.1. Axiomatic distance-based methods

A number of axiomatic distance-based methods have been developed to address the group decision-making ranking problem, with each method solving a different variant of the problem. The difference between these methods is the type of evaluations (complete rankings, complete ratings, strict partial orders, non-strict partial orders, etc.) being aggregated. Regardless of the type of evaluations, all of these methods are typically developed in three successive parts as follows: (i) provide a set of axioms any distance in the evaluation space should satisfy; (ii) prove that the axioms lead to the existence of a unique distance function; and (iii) provide solution procedures to find a consensus
evaluation, otherwise known as a median ranking in distance-based terms (see Section 3.1 for a formal definition). We remark, however, that the third part of this general process is not provided in most studies.

Kemeny and Snell (1962) were the first researchers to propose a distance-based method. They examined the problem where the evaluations are given as non-strict complete rankings. The Kemeny–Snell distance can be interpreted as follows: the distance between two complete ranking vectors \(a\) and \(b\) is given by the total number of rank reversals between them. A rank reversal is incurred whenever two objects have a different relative order in rankings \(a\) and \(b\). Similarly, half a rank reversal is incurred whenever two objects are tied in one ranking but not in the other ranking. Since our axioms (given in Section 4.2) for a distance between incomplete rankings are generalizations of Kemeny and Snell’s axioms, Section 4.1 thoroughly reviews their work. Kemeny and Snell did not provide a method for finding a median ranking. Subsequently, Bartholdi et al. (1989) showed that finding a median ranking is NP-hard.

Bogart (1973, 1975) looked at the problems of aggregating strict partial orders and asymmetric relations, respectively. A strict partial order is a binary relation that is irreflexive, asymmetric, and transitive, whereas an asymmetric relation is asymmetric and irreflexive but not necessarily transitive. Both of Bogart’s distances, defined in the space of strict partial orders and asymmetric relations, respectively, are equivalent to the Kemeny–Snell distance in the subspace of complete rankings. Bogart neither provided a method for finding median evaluations nor analyzed the complexity of such problems. Section 5.1 briefly reviews Bogart’s distance (Bogart, 1973) and shows that with a minor modification, it can be used as a distance between incomplete rankings. The unintentionally negative consequence of using Bogart’s distance (Bogart, 1973) to aggregate incomplete rankings, however, is the median ranking tends to favor rankings not containing ties (see Section 5).

Cook et al. (1986) examined the problem of aggregating non-strict partial orders. A non-strict partial order is a binary relation that is reflexive, antisymmetric, and transitive. In particular, a non-strict partial order allows three levels of comparison between each pair of objects, namely, strict preference, tied preference, and no comparison. The Cook–Kress–Seiford’ distance presented in Cook et al. (1986), defined in the space of non-strict partial orders, is also equivalent to the Kemeny–Snell distance in the subspace of complete rankings. Cook et al. (1986) did not provide a method for finding a median non-strict partial order, nor did they analyze the complexity of such a problem. Section 5.2 briefly reviews the Cook–Kress–Seiford’ distance, which can also be used as a distance between incomplete rankings. The inadvertent drawback of using the Cook–Kress–Seiford’s distance to aggregate incomplete rankings, however, is the resulting median rankings disproportionally favor rankings containing ties (see Section 5).

Cook and Kress (1985, p. 27) developed what they called an “ordinal ranking with intensity of preference.” However, careful inspection reveals that Cook and Kress were referring to what we (and most other literature in this subject) define as a complete rating or a complete cardinal evaluation. Therefore, the scope of their work differs from that of the proposed method, and we do not allude to it for the remainder of this article (we note, however, that any cardinal evaluation implicitly yields a corresponding ordinal evaluation based on the preference information conveyed by the ratings of multiple objects). Along the same line of reasoning, the aggregation method developed by Stengel (2013), which builds and solves a simultaneous system of equations, is not applicable to the current problem, as it deals with aggregating incomplete cardinal evaluations. More important, this non-axiomatic method makes the restrictive assumption that the number of judges equals at most the number of objects to evaluate. Their method cannot be extended to the case where judges outnumber objects as this would result in an over-determined system of equations. We refer the reader to the online supplement to this article for an extended discussion of alternative axiomatic distances and of restrictive and unrelated ranking aggregation methods.

The Kemeny–Snell distance can be naïvely extended to a distance between incomplete rankings. The distance between two incomplete rankings, \(a\) and \(b\), is the total number of rank reversals between them, where the rank reversals are only summed over the object pairs ranked in both \(a\) and \(b\). For reasons that will become clear later, we call this distance the Projected Kemeny–Snell distance, or PKS distance. The PKS distance was used by Dwork et al. (2001) and Cook et al. (2007). Moreover, we note that both Dwork et al. (2001) and Cook et al. (2007) considered only strict rankings. Unlike the other distances reviewed in this section, neither of those authors provided a set of axioms uniquely defining the PKS distance. This by itself is not so important; indeed, in Section 5.3, we prove that such a set of axioms exists. The trouble with using the PKS distance to aggregate incomplete rankings is that the median ranking tends to be disproportionally closer to rankings that compare more objects, relatively speaking (see Section 5). Indeed, this limitation can be easily tracked down to the set of axioms that we show uniquely define the PKS distance.

Dwork et al. (2001) and Cook et al. (2007) developed a heuristic algorithm and a branch-and-bound algorithm, respectively, for the following problem. Given a set of \(K\) strict incomplete rankings, find the strict complete ranking with the minimum sum of PKS distances to the given rankings. The optimization algorithm developed in this article can be applied to solve the above problem. Specifically, the featured algorithm has no restriction on the input rankings being strict or complete.

2.2. Alternative approaches

Ranking aggregation methods without an underlying axiomatic distance are referred to in the literature as ad hoc procedures, with axiomatic distance-based methods being regarded as the more formal methodologies based on their intuitive appeal and social choice–related axiomatic properties (Cook, 2006). Due to its popularity and customizability, we first discuss a purely ad hoc method known as the Borda Count (De Borda, 1781). This ranking aggregation method works by scoring each object in an individual evaluation based on its relative position utilizing a specific scoring function. Each object’s individual scores are summed and then the cumulative scores of all the objects are compared to determine the object’s position in the final ranking. It is not difficult for the Borda Count to handle incomplete
rankings. However, as we will illustrate in Section 5.6 via simple examples with predictable consensus rankings, none of the six Borda Count variations herein conceived is able to attain the correct solution with acceptable consistency. As we shall argue, this is attributable to the assumptions implied (intentionally or otherwise) by the scores assigned to unrated objects by the Borda Count and by similar approaches.

A hybrid distance-based ad hoc ranking aggregation model that utilizes $p$-metric distances was formulated in González-Pachón and Romero (1999). The family of distance functions is not founded upon a formal axiomatic framework. For $p = 1$ and $p = \infty$, the model is equivalent to an Archimedean linear goal programming problem and to a Chebyshev goal programming problem, respectively; their corresponding solutions yield a consensus that minimizes cumulative disagreement and a consensus that optimizes the largest disagreement. As González-Pachón and Romero (1999) explained, the first consensus can be very biased with respect to one individual ranking, whereas the other one places greater importance on the point of view of minority groups. González-Pachón and Romero (2001) converted the aforesaid models into interval goal programming models and handled incomplete and indecisive rankings as inputs by expressing the ranking of an object as an interval. There are inherent drawbacks to utilizing either of these models. First, as previously mentioned, both corresponding consensus measures are biased. Second, based on a constraint requiring that the sum of the elements of the aggregate ranking equal the sum of integers 1 to $m = |V|$, the aggregate rankings obtained with these models cannot express every possible complete ranking. These include the all-tied ranking when $m$ is even and the three-object ranking where a pair of objects is tied and strictly better or worse than a third item. Due to these significant biases and limitations, this article makes no further reference to this approach.

3. Preliminaries for aggregating incomplete rankings

3.1. Consensus ranking definition

In Kemeny and Snell (1962), Bogart (1973, 1975), and Cook et al. (1986), a consensus evaluation is defined as a median evaluation of the judges’ evaluations; that is, a median non-strict complete ranking, a median strict partial order, a median asymmetric relation, and a median partial order, respectively. We remark that their use and our use of the strict (non-strict) qualifier denotes a relationship without (with) ties. In general, a median is defined as follows: given a finite collection $\Omega$ of points in a metric space, a median of $\Omega$ is a point in the space with a minimum sum of distances to the points in $\Omega$. Accordingly, given a set of incomplete rankings, we define a consensus ranking as a complete ranking with the minimum sum of distances to the given rankings. In a strict sense, we require a consensus ranking to lie in a subspace of the evaluation space. The requirement of a consensus ranking to be complete captures the nature of the group decision-making problem, where it would be unacceptable to obtain a consensus ranking that is incomplete—after all, the group’s goal is to evaluate all objects. Therefore, regardless of the completeness (or lack thereof) of the given rankings, throughout this work we require a consensus ranking to be a complete ranking. With a slight abuse of notation, however, we refer to our consensus ranking as a median ranking. Henceforth, we use the terms consensus ranking and median ranking interchangeably.

3.2. Representation of incomplete rankings

Consider the universe $V$ of $n$ objects to be ranked and without loss of generality, assign a unique identifier to each object in $V$ so that $V = \{1, 2, \ldots, n\}$. In the vector representation, a complete ranking is a vector of the form $a = (a_1, \ldots, a_n)$, where $a_i$ is the rank of object $i$. A natural way to represent an incomplete ranking as a vector is by setting $a_i$ equal to a special symbol $\bullet$ if object $i$ is not ranked in $a$. Given an incomplete ranking $a$, we denote as $V_a$ the set of objects ranked in $a$. For example, for $V = \{1, 2, 3, 4\}$, the incomplete ranking $a$ with object 2 as first, object 3 as second, and objects 1 and 4 not ranked is represented as $a = (\bullet, 1, 2, \bullet)$; here, $V_a = \{2, 3\}$. Similarly, the incomplete ranking $b$ with object 4 as first, objects 1 and 3 tied as second, and object 2 not ranked is represented as $b = (2, \bullet, 2, 1)$; here, $V_b = \{1, 3, 4\}$. All of the theory and algorithms in this article are independent of the convention used to report ties; all we require is that all objects tied for the same ranking be assigned the same number (higher than the previous ranking on the list but lower than the next ranking on the list). Thus, for example, ranking $b$ can also be represented as $b = (2.5, \bullet, 2.5, 1)$.

Throughout this article, we will mostly use the ranking vector representation; however, it will be convenient to define the representation used by Kemeny and Snell (1962). They represent the ranking $a$ by an $n \times n$ matrix $(a_{ij})$, where $a_{ij} = 1$ if $i$ is preferred to $j$, $a_{ij} = -1$ if $j$ is preferred to $i$, and $a_{ij} = 0$ if $i$ and $j$ are tied. Clearly, any ranking can be represented by such a matrix; however, for a given matrix to represent a ranking, it must satisfy the properties of a ranking. The preference relationship $a_{ij} = 1$ must be asymmetric and transitive, and the tie relationship $a_{ij} = 0$ must be an equivalence relation.

A natural way to extend the Kemeny–Snell representation to include incomplete rankings is to set $a_{ij} = a_{ji} = \bullet$ for $j = 1, \ldots, n$ when object $i$ is not ranked in $a$. It should be clear from our notation when we use one subscript we are using the vector representation, and when we use two subscripts we are using the Kemeny–Snell representation. Clearly, the Kemeny–Snell representation can be obtained from the vector representation as follows:

$$a_{ij} = \begin{cases} \text{sign}(a_j - a_i) & \text{if } a_i \neq \bullet \text{ and } a_j \neq \bullet \\ \bullet & \text{otherwise} \end{cases}$$

Before proceeding, we briefly discuss how to represent the preference relationships of a general ranking $a$ (strict, non-strict, complete, or incomplete) via a directed graph known as its preference graph $G_a = (V, A)$ (also known as a directed beat-graph). Each $a \in V$ represents an object, and the arc $(i, j) \in A$ if and only if $i$ is ranked at least as high as $j$ in $a$ (ties imply $(j, i) \in A$ as well); if $i$ and $j$ are not both compared to each other in $a$, then $(i, j), (j, i) \notin A$. Accordingly, the cumulative preference graph $G = (V, A)$ of the set of rankings $\{a_k\}_{k=1}^{K}$ is formed by the union of the individual preference graphs $G_{a_k}$ for all $k$, where $K$ denotes the total number of rankings (judges).
3.3. Conditions for obtaining meaningful results

In order to aggregate incomplete rankings meaningfully, the input rankings should fulfill transitive comparability. In essence, this means that for every pair of objects \(i\) and \(j\) in \(V\), the input rankings must imply a transitive relationship. This property is formally verified by constructing the cumulative preference graph \(G = (V, A)\) of the set of input rankings \(\{a^k\}_{k=1}^K\), obtaining its transitive closure, and then checking either \((i, j) \in A\) or \((j, i) \in A\) (or both) for all \(i, j \in V\), where \(i \neq j\); when neither of these arcs exists between objects \(i\) and \(j\), the objects are said to lack transitive comparability. A second condition for aggregating incomplete rankings meaningfully, denoted as connective comparability, requires only that \(G\) be connected in the undirected sense. Although connective comparability is implied by transitive comparability, its advantage is that it can be established in the pre-ranking stage via the appropriate object-to-judge allocations, whereas transitive comparability must be verified \textit{a posteriori}. We refer the reader to Yates (1936), Cochran and Cox (1957), Cook et al. (2005), and Hochbaum and Levin (2006) for possible solutions to the problem of allocating the subsets of objects to be ranked to individual reviewers in order to satisfy these conditions.

4. Aggregation of incomplete rankings

4.1. The Kemeny–Snell distance between complete rankings

Kemeny and Snell (1962) proposed a set of axioms describing a distance function between complete rankings. In order to present Kemeny and Snell's axioms (denoted as KS-Axioms, for short), the following concepts are defined.

Definition 1. Ranking \(b\) is between rankings \(a\) and \(c\) if, for each pair of objects \(i\) and \(j\), the preference judgment of \(b\) either (i) agrees with \(a\) or (ii) agrees with \(c\), or (iii) \(a\) prefers \(i\), \(c\) prefers \(j\), and \(b\) ties \(i\) and \(j\).

Definition 2. A nonempty set \(S\) of objects is a segment of a given ranking \(a\) if \(S\) (the complement of \(S\)) is not empty and if the rank of \(a_i\), every element \(i\) in \(S\) is either higher than that of every element in \(S\) or lower than that of every element in \(S\).

Kemeny and Snell argued that a distance, \(d(\cdot, \cdot)\), between two complete rankings, \(a\) and \(b\), should satisfy the following axioms.

\textbf{KS-Axiom 1:} (Non-negativity.) \(d(a, b) \geq 0;\) and \(d(a, b) = 0\) if and only if \(a\) and \(b\) are the same ranking.

\textbf{KS-Axiom 2:} (Commutativity.) \(d(a, b) = d(b, a)\).

\textbf{KS-Axiom 3:} (Triangular inequality.) \(d(a, b) + d(b, c) \geq d(a, c)\), and the equality holds if and only if \(b\) is between \(a\) and \(c\).

\textbf{KS-Axiom 4:} (Anonymity.) If \(a'\) results from \(a\) by a permutation of the objects in \(V\), and \(b'\) results from \(b\) by the same permutation, then \(d(a, b) = d(a', b')\).

\textbf{KS-Axiom 5:} (Extension.) If two rankings \(a\) and \(b\) agree except for a set \(S\) of \(k\) elements, which is a segment of both, then \(d(a, b)\) may be computed as if these \(k\) objects were the only objects being ranked.

\textbf{KS-Axiom 6:} (Scaling.) The minimum positive unit is one.

KS-Axioms 1 and 3 are self-explanatory. KS-Axiom 4 ensures that the distance does not depend on the particular labeling of the objects. KS-Axiom 5 states that if the two rankings are in complete agreement at the beginning and at the end of the list (i.e., if they differ only in the ranking of \(k\) objects they rank in the middle), then this distance is the same as if these \(k\) objects were the only objects under consideration. Notice that this axiom includes situations when two rankings are in complete agreement only at one end of the list; in this case, the remaining objects are considered to be in the middle of the list, with the subset of objects in agreement at the other end of the list simply being the empty set. KS-Axiom 6 is just a matter of convention: Kemeny and Snell set the minimum positive unit to one; however, as explained below, their distance has a nicer interpretation if the minimum positive unit is one-half.

Kemeny and Snell proved KS-Axioms 1 to 6 are simultaneously satisfied by only one distance, which is denoted by the following equivalent equations (here, we set the minimum positive unit to one-half):

\[
d_{KS}(a, b) = \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij} - b_{ij}| \tag{1a}
\]

\[
d_{KS}(a, b) = \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} |\text{sign}(a_i - a_j) - \text{sign}(b_i - b_j)|, \tag{1b}
\]

where Equation (1a) uses the Kemeny–Snell representation of rankings, and Equation (1b) uses the vector representation of rankings.

The Kemeny–Snell distance \(d_{KS}(\cdot, \cdot)\), with the minimum positive unit being one-half, has the following interpretation. The distance between two rankings is given by the total number of rank reversals between them. A rank reversal is incurred whenever two objects have a different relative order in the rankings \(a\) and \(b\). Similarly, half a rank reversal is incurred whenever two objects are tied in one ranking but not in the other ranking. Hereafter, when we refer to the Kemeny–Snell distance, we will assume that the minimum positive distance is set to one-half.

4.2. Axioms for a distance between incomplete rankings

In this section, we modify Kemeny and Snell's axioms in order to obtain a set of axioms appropriate for a distance between incomplete rankings. Moreover, we want to define a distance, \(d(\cdot, \cdot)\), suitable for solving the incomplete-ranking aggregation problem. Given the incomplete rankings of \(K\) judges, \(\{a^1, \ldots, a^K\}\), the distance-based consensus ranking \(r\) is the optimal solution to the incomplete-ranking aggregation optimization problem:

\[
\min_r \sum_{k=1}^{K} d(a^k, r) \tag{2}
\]

where the minimum is over all strict and non-strict complete rankings.
The aim of the axioms proposed in this section is to describe a distance between incomplete rankings such that when this distance is used to solve Problem (2), the resulting consensus ranking minimizes the disagreement of the judges. We now specify the axioms our distance must satisfy. Throughout this section, the axioms are denoted using the number of the corresponding Kemeny and Snell’s axioms. In addition, the axioms denoted by a prime are those modified slightly to apply to distances between incomplete rankings. For example, Axiom 0 has no corresponding KS-Axiom and Axiom 1’ is a slight variant of KS-Axiom 1.

First, \( d(\cdot, \cdot) \) must capture our intuition of disagreement between incomplete rankings: Given two incomplete rankings \( a \) and \( b \), \( d(a, b) \) must consider only the disagreement between \( a \) and \( b \) in the objects ranked by both \( a \) and \( b \) (i.e., only the objects in \( V_a \cap V_b \)). That is, \( d(\cdot, \cdot) \) must measure the disagreement in the relative order of the ranked objects, not the disagreement on which the objects were ranked (or not ranked) by each judge. For example, given the incomplete rankings \( a = (1, 2, 3) \) and \( b = (3, 2, 1) \), the distance should only consider the objects in \( V_a \cap V_b = \{2, 4\} \). Specifically, we do not say \( a \) disagrees with \( b \) because \( b \) does not rank object 1; likewise, we do not say \( b \) disagrees with \( a \) because \( a \) does not rank object 3. However, it is clear that rankings \( a \) and \( b \) disagree in their preference between objects 2 and 4, with each object preferred over the other object by \( a \) and \( b \), respectively. This condition is self-explanatory; however, to make it mathematically precise, we need the following definition.

**Definition 3.** Given a ranking \( a \) and a subset \( S \) of the object universe \( V \), the projection of \( a \) on \( S \), denoted as \( a|_{S} \), is the ranking of the objects in \( S \) preserving the relative order of the objects specified by \( a \) on the objects in \( S \).

The desired condition \( d(a, b) \) must only consider the objects in \( V_a \cap V_b \) is expressed as follows:

**Axiom 0:** (Relevance.) \( d(a, b) = d(a|_{V_a \cap V_b}, b|_{V_a \cap V_b}) \).

**Axiom 1’:** (Non-negativity.) \( d(a, b) \geq 0 \); and \( d(a, b) = 0 \) if and only if \( a|_{V_a \cap V_b} \) and \( b|_{V_a \cap V_b} \) are the same ranking.

**Axiom 2:** (Commutativity.) Identical to KS-Axiom 2.

The modification of Axiom 1’ with respect to KS-Axiom 1 has the same justification as Axiom 0. That is, we want \( d(\cdot, \cdot) \) to measure the disagreement in the relative order of the ranked objects, not the disagreement on which objects were ranked by each judge. Therefore, the rankings \( a \) and \( b \) have no disagreement \( d(a, b) = 0 \) whenever they have the same relative preferences among the objects ranked by both. Given two arbitrary incomplete rankings \( a \) and \( b \), notice the set \( V_a \cap V_b \) might be empty or contain just one object. In the case when \( V_a \cap V_b = \emptyset \), we have

\[
d(a, b) = d(a|_{\emptyset}, b|_{\emptyset}) = d(\emptyset, \emptyset) = 0,
\]

where the first equality follows from Axiom 0; the second inequality follows since a ranking of an empty set of objects is itself the empty set (interpreting a ranking as a partial order); and the third inequality follows from Axiom 1’. A similar analysis shows that when \( |V_a \cap V_b| = 1 \), these axioms imply \( d(a, b) = 0 \). This follows since two incomplete rankings that do not rank the same objects or that rank only one object in common have no disagreement. In any case, for our purposes, we will only use \( d(\cdot, \cdot) \) in the context of Problem (2). In particular, we only need to calculate the disagreement between a complete ranking and an incomplete ranking. Thus, the observation articulated in this paragraph is not important since we can assume, without loss of generality, that all of the given rankings \( a^k \) rank at least two objects.

KS-Axiom 3 requires the distance to satisfy the triangle inequality \( d(a, b) + d(b, c) \geq d(a, c) \). This condition cannot be imposed directly on the distance, as this condition is inconsistent with Axioms 0 and 1’. This is illustrated in the following example.

Let \( a = (1, 2, \bullet) \), \( b = (\bullet, 2, 1) \), and \( c = (2, 1, \bullet) \). Since \( |V_a \cap V_b| = 1 \) and \( |V_b \cap V_c| = 1 \), we have \( d(a, b) = 0 \) and \( d(b, c) = 0 \). On the other hand, from Axiom 1’ and the fact that \( a \) and \( c \) are not the same ranking when projected to \( V_a \cap V_c \), we have \( d(a, c) > 0 \). This clearly violates the triangle inequality. We nevertheless require our distance satisfy a relaxed version of the triangle inequality (Axiom 3’ below). Strictly speaking, \( d(\cdot, \cdot) \) will not be even a pseudometric because it violates the (unrelaxed) triangle inequality.

**Axiom 3’:** (Relaxed triangular inequality.)

\[
d(a|_{V_a \cap V_b}, b|_{V_a \cap V_b}) + d(b|_{V_a \cap V_c}, c|_{V_a \cap V_c}) \geq d(a|_{V_a \cap V_c}, c|_{V_a \cap V_c}).
\]

**Axiom 4:** (Anonymity.) Identical to KS-Axiom 4.

**Axiom 5:** (Extension.) Identical to KS-Axiom 5.

The sixth and last Kemeny and Snell’s axiom, KS-Axiom 6, is the scaling axiom. For the case of complete rankings, this axiom is a mere convention. However, the scaling axiom is of central importance for Problem 2. The idea behind the scaling axiom is implicit in the definition of the incompleteranking aggregation problem; that is, all of the judges’ rankings, \( a^k \), have the same importance. However, Problem (2) minimizes a sum of distances, each of which may be evaluated over spaces of different dimensionality, due to Axiom 0 requiring the distance to be evaluated by projecting the two incomplete rankings into the set of objects ranked by both. Therefore, since distances in higher-dimension spaces tend to be larger than those in lower-dimension spaces, the objective function of Problem (2) will be dominated by the distances from \( r \) to the incomplete rankings ranking a larger number of objects.

In light of the discussion in the previous paragraph, the following is argued. The distances between incomplete rankings should be normalized so that all of the distances in Problem (2) are comparable. This can be achieved by normalizing the distances so that they are between zero and one ( inclusively), where a distance of zero will indicate total agreement and a distance of one will indicate total disagreement. Intuitively, given a ranking \( a \), the ranking disagreeing the most with \( a \) is its reverse ranking. A ranking \( a \) is the reverse ranking of \( b \) if, for every pair of objects \( i \) and \( j \), either \( a \) prefers \( i \) and \( b \) prefers \( j \) or \( a \) prefers \( j \) and \( b \) prefers \( i \). Instead of having a scaling axiom like Kemeny and Snell, we impose the following normalization axiom on our distance.
Axiom 6': (Normalization.) \( d(a, b) \leq 1 \); and \( d(a, b) = 1 \) if and only if \( b|_{V_a \cap V_b} \) is the reverse ranking of \( a|_{V_a \cap V_b} \).

4.3. Existence and uniqueness of the distance

Here, we show that the distance between the incomplete rankings given in Equation (3), here called the normalized projected Kemeny–Snell distance, or simply NPKS distance, satisfies Axioms 0 to 6. We also prove that the NPKS distance is the unique distance simultaneously satisfying these axioms:

\[
d_{\text{NPKS}}(a, b) = \begin{cases} 
\frac{d_{\text{KS}}(a|_{V_a \cap V_b}, b|_{V_a \cap V_b})}{(|V_a \cap V_b| - |V_a \cap V_b|)/2} & \text{if } |V_a \cap V_b| \geq 2 \\
0 & \text{otherwise}
\end{cases}
\]

(3)

Lemma 1. (Kemeny and Snell, 1962). Given two complete rankings \( a \) and \( b \) on a set of objects \( V \), \( d_{\text{KS}}(a, b) \) attains its maximum of \(||V| - |V|)/2 \) when \( b \) is the reverse ranking of \( a \).

Lemma 2. The NPKS distance (Equation (3)) satisfies Axioms 0 to 6.

Proof. The NPKS distance, \( d_{\text{NPKS}}(\cdot, \cdot) \), satisfies Axiom 0 directly from its definition, Equation (3). The non-negativity of \( d_{\text{NPKS}}(\cdot, \cdot) \) follows from Equation (3) and the non-negativity of \( d_{\text{KS}}(\cdot, \cdot) \). To see \( d_{\text{KS}}(\cdot, \cdot) \) satisfies the second part of Axiom 1', we consider two cases:

Case 1: \(|V_a \cap V_b| \leq 1\): In this case, \( d_{\text{NPKS}}(a, b) = 0 \) and both \( a|_{V_a \cap V_b} \) and \( b|_{V_a \cap V_b} \) are rankings over either an empty set of objects or a set with a single element; thus, they are, by definition, the same ranking. Therefore, the second part of Axiom 1' is satisfied.

Case 2: \(|V_a \cap V_b| > 1\): In this case, \( d_{\text{NPKS}}(\cdot, \cdot) \) satisfies the second part of Axiom 1' as a consequence of Equation (3) and \( d_{\text{KS}}(\cdot, \cdot) \) satisfying the second part of KS-Axiom 3. Axioms 2, 4, and 5 follow from Equation (3) and from the fact that \( d_{\text{KS}}(\cdot, \cdot) \) satisfies KS-Axioms 2, 4, and from the fact that 5, respectively.

Axiom 3' follows from Equation (3); the fact \( a|_{(V_a \cap V_b) \setminus V_a} \) and \( b|_{(V_a \cap V_b) \setminus V_a} \) are complete rankings; and from the fact that \( d_{\text{KS}}(\cdot, \cdot) \) satisfies KS-Axiom 3. Finally, from Lemma 1 and Equation (3), \( d_{\text{NPKS}}(\cdot, \cdot) \) satisfies Axiom 6'.

The following corollary follows directly from Lemma 2.

Corollary 1. Axioms 0 to 6 are consistent.

Next, we show that Axioms 0 to 6 uniquely determine the NPKS distance.

Theorem 1. The distance \( d_{\text{NPKS}}(\cdot, \cdot) \) is the unique distance satisfying Axioms 0 to 6 simultaneously.

Proof. The fact that \( d_{\text{NPKS}}(\cdot, \cdot) \) satisfies Axioms 0 to 6 was established in Lemma 2. Thus, we only need to show that no other distance simultaneously satisfies Axioms 0 to 6. Let \( d(\cdot, \cdot) \) be a generic distance satisfying Axioms 0 to 6. We will prove the theorem by showing for any two rankings \( a \) and \( b \), \( d(a, b) = d_{\text{NPKS}}(a, b) \). We divide our analysis into two cases based on the cardinality of the ranked objects in common.

Case 1: \(|V_a \cap V_b| \leq 1\). As argued in Section 4.2, for any distance function satisfying Axioms 0 and 1', and for any two rankings \( a \) and \( b \) such that \(|V_a \cap V_b| \leq 1\), the distance must be equal to zero.

Case 2: \(|V_a \cap V_b| \geq 2\). This case is further split into two cases depending on the individual cardinalities of \( V_a \) and \( V_b \).

Case 2.1: \(|V_a| = |V_b| = |V| \) (i.e., both rankings are complete). Since in this case we are restricting our attention to complete rankings, it follows that Axioms 1 to 5 are identical to KS-Axioms 1 and 5. Therefore, since by assumption \( d(a, b) \) satisfies Axioms 1 to 5, \( d(a, b) \) satisfies KS-Axioms 1 and 5.

As explained in Section 4.1, the sole purpose of KS-Axiom 6 is to fix the measurement unit; in other words, KS-Axioms 1 to 5 uniquely determine a distance function up to a scaling factor. Therefore, since both \( d(a, b) \) and \( d_{\text{KS}}(a, b) \) satisfy KS-Axioms 1 to 5, \( d(a, b) = \alpha d_{\text{KS}}(a, b) \) for some constant \( \alpha \) that may depend only on \(|V|\).

From Equation (3), and since in Case 2 \(|V| \geq 2\) and \(|V_a \cap V_b| = V \), we have \( d_{\text{NPKS}}(a, b) = (|V|^2 - |V|)/2)^{-1} d_{\text{KS}}(a, b) \).

Let the ranking \( r \) be the ranking on which the objects are ranked according to their index; that is, \( r = (1, 2, \ldots, |V|) \). Let the ranking \( -r \) be the reverse ranking of \( r \); that is, \( -r = (|V|, |V| - 1, \ldots, 1) \). From Axiom 6, we have \( d(r, r) = 0 \) and \( d_{\text{NPKS}}(r, r) = 0 \), and from Axiom 6 it follows \( d(r, -r) = 1 \) and \( d_{\text{NPKS}}(r, -r) = 1 \). Therefore, it must be the case \( \alpha = (|V|^2 - |V|)/2)^{-1} \). Thus, in Case 2.1, as in Case 1, \( d(a, b) = d_{\text{NPKS}}(a, b) \) for any two rankings.

Case 2.2: \(|V_a| < |V| \) or \(|V_b| < |V| \) (i.e., at least one of the rankings is incomplete). In this case, as in Case 1 and Case 2.1, \( d(a, b) = d_{\text{NPKS}}(a, b) \) for any two rankings \( a \) and \( b \). This is shown by the following sequence of equalities:

\[
d(a, b) = d(a|_{V_a \cap V_b}, b|_{V_a \cap V_b}) \\
= \frac{|V_a \cap V_b|^2 - |V_a \cap V_b|}{2} d_{\text{KS}}(a|_{V_a \cap V_b}, b|_{V_a \cap V_b}) \\
= d_{\text{NPKS}}(a|_{V_a \cap V_b}, b|_{V_a \cap V_b}) = d_{\text{NPKS}}(a, b).
\]

The first and last equalities follow from Axiom 0. The second equality follows from our analysis of Case 1 and \( a|_{V_a \cap V_b} \) and \( b|_{V_a \cap V_b} \) being complete rankings over the set \( V_a \cap V_b \). The third equality follows from Equation (3) and since \(|V_a \cap V_b| \geq 2\). Hence, the result holds for Case 2. \(\square\)
5. Comparison with other approaches

There are other distances that may be used to solve the Incomplete-ranking Aggregation Problem (IAP). These distances include Bogart’s distance (Bogart, 1973), the Cook–Kress–Seiford distance (Cook et al., 1986), and the unnormalized version of the NPKS distance (proposed in Dwork et al. (2001)). This section shows that the consensus ranking obtained when using the NPKS distance is more intuitive than the consensus rankings obtained with these alternative distances. The section is organized as follows: Sections 5.1 to 5.3 present each of these distances and comment on their drawbacks when applied to solve the IAP; Section 5.4 gives specific examples with predictable consensus rankings to illustrate these drawbacks, and Section 5.5 extrapolates these examples by generating random instances with similarly predictable consensus rankings. Section 5.6 demonstrates the inadequacy of the Borda Count for aggregating incomplete rankings when no assumptions should be made regarding unranked items. Finally, the drawback of each of the aforementioned aggregation methodologies is summarized at the end of this section.

5.1. Review and discussion of Bogart’s distance

Bogart’s distance between two given partial orders, \(P\) and \(Q\), is given by \(d_B(P, Q) = \|I(P) - I(Q)\|\), where \(\|\cdot\|\) denotes the matrix \(L_1\)-norm (i.e., the sum of all the matrix entries’ absolute values), and \(I(P)\) is the incidence matrix of the partial ordering \(P\) defined as follows: \(I(P)_{ij} = 1 \text{ if } (i, j) \in P\) and \(I(P)_{ij} = 0\) otherwise, where \((i, j) \in P\) means object \(j\) is not preferred to object \(i\) (that is, either object \(i\) is preferred to object \(j\) or the objects are tied). Since an incomplete ranking is a partial order, we can use Bogart’s distance to find the distance between incomplete rankings (i.e., \(I(P)_{ij} = 0\) for every unranked object \(j\) and ranked/unranked object \(i\)).

The median ranking obtained by using Bogart’s distance will tend to be a ranking containing an artificially low number of ties. That is, even objects that are expected to be tied by the consensus ranking will not be tied by it. This is illustrated via an example in Section 5.4 and discussed in greater depth in Section A.2 of the online supplement. We close this section with the following remark. Our purpose is not to criticize Bogart’s distance (a legitimate distance between partial orders); rather, it is to show that Bogart’s distance is not appropriate to solve the IAP (unlike our NPKS distance, this distance was not designed to solve the IAP).

5.2. Review of the Cook–Kress–Seiford distance

The Cook–Kress–Seiford (Cook et al., 1986) distance between two given partial orders, \(Q\) and \(R\), on a set of \(n\) objects, is given by

\[
d_C(Q, R) = \sum_{j=1}^{n} \sum_{k=1}^{n} \left[ \frac{1}{2} |J(Q)_{ij} - J(R)_{ij}| + |P(Q)_{ij} - P(R)_{ij}| \right],
\]

where \(J\), referred to as the information matrix, and \(P\), referred to as the preference matrix, are defined as follows: \(J_{ij} = 1\) if \(i\) and \(j\) are compared (strict preference or tie) and \(J_{ij} = 0\) otherwise; \(P_{ij} = 1\) if \(i\) is strictly preferred to \(j\) and \(P_{ij} = 0\) otherwise.

In general, the behavior of the Cook–Kress–Seiford distance can be generalized similarly to the behavior exhibited by Bogart’s distance. The difference is that the median ranking obtained by using the Cook–Kress–Seiford distance will tend to be a ranking containing an artificially high number of ties. That is, even objects that are expected to not be tied by the consensus ranking will be tied by it. This is illustrated via an example in Section 5.4 and discussed in greater depth in Section A.2 of the online supplement. We close this section with the following remark. Our purpose is not to criticize the Cook–Kress–Seiford distance (a legitimate distance between partial orders); our intention is to show that it is not appropriate to use this distance to solve the IAP (unlike our NPKS distance, this distance was not designed to solve the IAP).

5.3. Review and discussion of the (unnormalized) PKS distance

Dwork et al. (2001) extended the Kendall Tau distance (a distance between strict complete rankings equivalent to the Kemeny–Snell distance in the space of strict complete rankings) to a distance between strict incomplete rankings. Dwork et al. (2001) referred to this distance as the induced Kendall Tau distance, but for reasons that will become clear below, we refer to it as the PKS distance. Strictly speaking, the induced Kendall Tau distance does not allow non-strict rankings, however the PKS distance does; if the rankings are strict, then the two distances are equivalent. Unlike Bogart (1973) and Cook et al. (1986), Dwork et al. (2001) did not articulate a set of axioms characterizing the PKS distance; here, we show that the PKS distance can be derived axiomatically. Given a universe of objects \(V\) and two incomplete rankings \(a\) and \(b\), the PKS distance is

\[
d_{PKS}(a, b) = d_{KS}(a|_{V_a \cap V_b}, b|_{V_a \cap V_b}). \tag{4}
\]

Notice that the PKS distance is simply an unnormalized version of the NPKS distance defined in Section 4.3. On a related note, this implies, that the NPKS distance reduces to the PKS distance (times a constant) whenever the rankings contain the same number of ranked objects. Moreover, based on the connection between the two distances and Theorem 1, we obtain the following result.

**Corollary 2.** The distance \(d_{PKS}(\cdot, \cdot)\) is the unique distance satisfying simultaneously Axioms 0 to 5 and KS-Axiom 6.

The only difference between the set of axioms uniquely satisfied by \(d_{PKS}(\cdot, \cdot)\) and the set of axioms uniquely satisfied by \(d_{NPKS}(\cdot, \cdot)\) is the scaling axiom; in particular, \(d_{PKS}(\cdot, \cdot)\) satisfies KS-Axiom 6, whereas \(d_{NPKS}(\cdot, \cdot)\) satisfies Axiom 6’.

The median ranking derived via the PKS distance tends to favor incomplete rankings with higher completeness; that is, incomplete rankings that rank relatively more objects than other rankings. The reason for this phenomenon is the following. Recall that a median ranking is a complete ranking with the minimum sum of distances to the given incomplete rankings. Since the PKS distance between rankings \(a\) and \(b\) is calculated by projecting the rankings to the set \(V_a \cap V_b\), the distances in
the objective function are taken over object spaces of different dimensions. Therefore, since distances in higher-dimensional spaces tend to be larger than distances in lower-dimensional spaces, the objective function will be dominated by the distances from the median ranking to rankings with higher completeness. This is illustrated further in the following section.

5.4. Comparison of the consensus rankings

This section evaluates the consensus rankings yielded by the NPKS distance and by other selected methods via three examples. The comparisons highlight the increased intuitiveness of utilizing the NPKS distance to solve the IAP. The examples compare the consensus rankings obtained by the NPKS distance ($d_{\text{NPKS}}$) and by the three axiomatic distances discussed in Sections 5.1 to 5.3: Bogart's distance ($d_B$), the Cook–Kress–Seiford distance ($d_C$), and the PKS distance ($d_{\text{PKS}}$).

5.4.1. Comparison 1: Inadequacy of Bogart’s distance to solve the IAP

Let $a$, $b$, and $c$ be rankings of three objects such that each corresponding judge ties a different pair of objects and leaves the remaining object unranked. These input rankings and the median rankings obtained via the four distinct distances are depicted in Table 1. Based on the given information, it is reasonable to expect the unique consensus ranking ties all objects. However, using Bogart's distance, any possible strict ranking (that is, any possible permutation of the objects) is a median ranking. In other words, using Bogart's distance, all possible strict rankings have the same sum of distances to rankings $a$, $b$, and $c$, and such sum of distances is smaller than the sum of distances from the ranking tying all objects to rankings $a$, $b$, and $c$. This suggests that the median ranking obtained with Bogart's distance tends to contain an artificially low number of ties. That is, even objects that are expected to be tied in the median ranking are not tied. Note that all other median rankings, including the one obtained using our NPKS distance, agree with the intuitive consensus ranking.

5.4.2. Comparison 2: inadequacy of the Cook–Kress–Seiford distance to Solve the IAP

Let $a$, $b$, and $c$ be rankings of four objects such that each corresponding judge prefers object $i$ over object $i + 1$, but judge $a$ ranks only objects 1 and 2, and judge $c$ ranks only objects 3 and 4. These input rankings and the median rankings obtained via the four distinct distances are depicted in Table 2. Based on the given information, it is reasonable to expect the unique consensus ranking is $(1, 2, 3, 4)$. However, using the Cook–Kress–Seiford distance, the median ranking ties all of the objects. This suggests that the median ranking obtained using this distance tends to contain an artificially high number of ties. That is, even objects that are expected not to be tied by the consensus ranking are tied by it. Note that all other median rankings, including the one obtained using our NPKS distance, agree with the intuitive consensus ranking.

5.4.3. Comparison 3: inadequacy of the PKS distance to solve the IAP

Consider the five-object incomplete rankings $a$ to $j$ characterized as follows: rankings $a$ and $b$ prefer object 1 to object 2; rankings $c$ and $d$ prefer object 2 to object 3; rankings $e$ and $f$ prefer object 3 to object 4; rankings $g$ and $h$ prefer object 4 to object 5; ranking $i$ prefers object 3 to object 4, and it also prefers object 2 to both objects 3 and 4; and ranking $j$ is the contrarian ranking $(5, 4, 3, 2, 1)$ (i.e., in reverse-order with respect to the overwhelming majority). We note that the unstated objects in each ranking are assumed to be unranked. These input rankings and the median rankings obtained via the four distinct distances are depicted in Table 3. Since, for every pair of objects $l$ and $l + 1$, two or more judges prefer $l$ over $l + 1$, whereas only one judge prefers $l + 1$ over $l$, it is reasonable to expect the unique consensus ranking is $(1, 2, 3, 4, 5)$. However, using the PKS distance, the median ranking is $(5, 2, 3, 4, 1)$. This ranking contradicts what rankings $a$ to $i$ collectively imply that object 1 is the best and object 5 is the worst. Furthermore, it contradicts the majority of judges preferring object $l$ over object $l + 1$. This suggests that the median ranking obtained via the PKS distance tends to agree with the ranking with higher completeness, which is the minority contrarian ranking in this example (this also applies to Bogart's distance, as it is also an unnormalized distance). Note the only median ranking agreeing with the intuitive consensus ranking is obtained using our NPKS distance.

---

Table 1. Bogart's median ranking may contain an artificially low number of ties

| Judge's rankings | Median ranking using distance |
|------------------|------------------------------|
|                  | $d_B$ | $d_C$ | $d_{\text{PKS}}$ | $d_{\text{NPKS}}$ |
| Object 1         | •     | 1     | 1                | 1                 |
| Object 2         | 1     | •     | 1                | 1                 |
| Object 3         | 1     | 1     | •                | 1                 |

Table 2. The median ranking obtained using the Cook–Kress–Seiford distance may contain an artificially high number of ties

| Judge's rankings | Median ranking using distance |
|------------------|------------------------------|
|                  | $d_B$ | $d_C$ | $d_{\text{PKS}}$ | $d_{\text{NPKS}}$ |
| Object 1         | 1     | 1     | •                | 5                 |
| Object 2         | 2     | 2     | 1                | 2                 |
| Object 3         | •     | 2     | 2                | 3                 |
| Object 4         | •     | •     | 2                | 4                 |
| Object 5         | •     | •     | •                | 1                 |

Table 3. The PKS median ranking tends to agree with rankings with higher completeness

| Judge's rankings | Median ranking using distance |
|------------------|------------------------------|
|                  | $d_B$ | $d_C$ | $d_{\text{PKS}}$ | $d_{\text{NPKS}}$ |
| Object 1         | 1     | 1     | •                | 5                 |
| Object 2         | 2     | 2     | 1                | 2                 |
| Object 3         | •     | 2     | 2                | 3                 |
| Object 4         | •     | •     | 2                | 4                 |
| Object 5         | •     | •     | •                | 1                 |
5.5. Extended comparison tests

This section verifies the insights of the previous examples by comparing the solutions of the respective distance-based ranking aggregation methods when they are applied to solve random incomplete-ranking problem instances generated from three complete-ranking data templates; the incompleteness of an instance is introduced by randomly removing values from the template ranking vectors. Simply put, these templates are generalizations of the examples in Section 5.4: Type I has every judge tie all objects, Type II has every judge rank each object based on its index (i.e., yielding a set of identity–permutation rankings), and Type III has every judge follow the Type II template, except for a pre-specified minority of contrarians who rank each object based on its reverse-order index (i.e., yielding a subset of reverse rankings with respect to the identity–permutation ranking).

It is straightforward to discern that the unique consensus ranking of the Type I incomplete-ranking instances should be the all-tied ranking and the unique consensus ranking of the Type II and Type III incomplete-ranking instances should be the identity permutation ranking. Indeed, the templates were chosen so that their respective individual rankings unequivocally imply these rankings—in fact, every method applied to each of the three complete-ranking templates obtains the respective correct unique ranking. Before continuing, we clarify that the incompleteness and differences in the individual instances are created by randomly removing a uniform number of values (between zero and \( n - 2 \)) in each of the individual complete-ranking vectors; however, the Type III reverse rankings are left as complete rankings.

The experiments described in this section counted the number of times a median ranking of each of the highlighted ranking aggregation methods exactly matches the expected consensus ranking. In addition, since a median ranking of these methods is not guaranteed to be unique based on the non-convexity of the solution space, the experiments tallied the number of alternative median rankings obtained for each instance. Table 4 summarizes the results of nine different experiments (three per template), each consisting of 50 randomly generated instances. The reported metrics are the total number of exact matches to the expected consensus ranking (“Exact matches”), the number of times the median ranking is unique and it matches the expected consensus ranking (“Unique exact matches”), and the average and maximum numbers of alternative median rankings (“Ave number alt. medians” and “Max number alt. medians,” respectively). The “Unique exact matches” metric is of special interest as this particular outcome guarantees that the expected consensus ranking would be obtained via a standard run of a mathematical programming solver (in general, it is not possible to predict which optimal solution a solver will return when there are multiple optimal solutions).

To obtain all of the median rankings associated with a ranking aggregation method, the experiments performed a full solution enumeration for each instance. The number of objects ranked in all of the experiments is seven, due to the enumeration for each instance. The number of objects represented in the table is the total number of exact matches to the expected consensus ranking (“Exact matches”), thenumber of times the median ranking is unique and it matches the expected consensus ranking (“Unique exact matches”), and the average and maximum numbers of alternative median rankings (“Ave number alt. medians” and “Max number alt. medians,” respectively). The “Unique exact matches” metric is of special interest as this particular outcome guarantees that the expected consensus ranking would be obtained via a standard run of a mathematical programming solver (in general, it is not possible to predict which optimal solution a solver will return when there are multiple optimal solutions).

The instances without contrarian rankings contain no disagreements and yet, as evidenced by the first horizontal block of Table 4, Bogart’s distance and the Cook–Kress–Seiford distance obtained the expected consensus ranking in the Type I and Type II individual experiments, respectively, a maximum of 12 out 50 times. By contrast, the PKS and NPKS distances attained it in all Type I and Type II instances, and it was also the unique median ranking for both in at least 90% of all of the related experimental instances—in fact, both distances performed identically over these two types of experiments. As the third and fourth horizontal blocks of the table show, a further downside of using Bogart’s distance when the incomplete rankings contain ties (i.e., the Type I instances) is its propensity for admitting an excessive number of alternative median rankings. The
existence of numerous alternative median rankings is problematic due to it reducing the usefulness of the method as a decision-making tool as well as the likelihood that the expected consensus ranking would be obtained via the standard solution of the corresponding mathematical program. Thus, based on their performance over these simple instances, Bogart’s and the Cook–Kress–Seiford distances are inadequate for dealing with incomplete rankings.

Type III instances contain a minority of contrarian judges, specifically 20%, who rank every object and who categorically oppose the collective views of the overwhelming majority. In all three related experiments, the Cook–Kress–Seiford distance was not competitive. Bogart’s distance and the PKS distance attained the expected consensus ranking for the same instances, and the matching median ranking was unique for a smaller identical subset of these instances. The NPKS distance matched their results for the two aforementioned subsets of instances and it performed better with regard to both metrics over the remaining instances. Moreover, as the third and fourth blocks of Table 4 show, the NPKS distance allowed significantly fewer maximum and average numbers of alternative median rankings than the PKS distance in nearly every experiment. This suggests that the NPKS distance reduces median ranking symmetry relative to the PKS distance in the space of incomplete rankings, which enhances its usefulness as a decision-making tool. In conclusion, the NPKS distance outperformed every other distance in achieving the correct consensus ranking and in allowing fewer alternative median rankings.

As the preceding analysis suggests, in certain instances the NPKS distance yielded the expected consensus ranking as well as other alternative median rankings; specifically, this held true for 24 of the 450 instances (i.e., all Type I, II, III instances combined). To ascertain the quality of the alternative median rankings in these situations, we calculated the average number of rank reversals (i.e., the average Kemeny–Snell distance) between the expected consensus ranking and each of the possible median rankings yielded by the NPKS distance. Remarkably, in 12 (i.e., half) of the instances the average number of rank reversals was at most one, and in six (i.e., a quarter) of the instances it was at most one-half (recall that half a rank reversal between two rankings is incurred whenever two objects are tied in one ranking but not in the other ranking); the average number of rank reversals for all 24 instances was 1.45. Based on these relatively low averages, we conjecture that when there are non-unique median rankings with respect to the NPKS distance, these alternative median rankings tend to be relatively close to one another (i.e., there are relatively few rank reversals between them). This result further enhances the reliability of the proposed method since, as the previous two paragraphs explained, the NPKS method tends to produce significantly more unique exact matches than the other distance-based methods.

5.6. Inadequacy of the Borda Count to solve the IAP

This section demonstrates the inadequacy of the Borda Count for aggregating incomplete rankings by applying the non-distance-based method to solve the random incomplete-ranking instances of the previous section. Recall that the Borda Count determines the aggregate ranking by adding each object’s scores from all of the individual input rankings and then comparing the cumulative scores of all objects. Hence, the Borda Count is guaranteed to yield only one consensus ranking. As the Borda Count is not explicitly defined over the space of incomplete rankings, we test three variations of the method for dealing with unranked objects (labeled as $B_1$, $B_2$, and $B_3$); the corresponding scoring functions are detailed in Section C of the online supplement. For further analysis, the experiments also test three additional variations of the Borda Count (labeled $W_1$, $W_2$, and $W_3$), which are obtained from $B_1$, $B_2$, and $B_3$ by taking their respective weighted averages (i.e., divide each object’s cumulative score by the number of times the object was ranked). Table 5 gives the number of exact matches to the expected consensus ranking for each of the six Borda Count variations. Even though six different scoring functions were tested, none of the respective Borda Count variations found the expected consensus with acceptable consistency. These underwhelming results are a direct consequence of the assumptions implied (intentionally or otherwise) by the Borda Count and similar approaches about the unranked objects in an individual evaluation. Based on this reasoning, it is our belief that such methods are unreliable for aggregating incomplete rankings when no assumptions should be made regarding unranked items. The drawbacks of each of the discussed methodologies are listed in Table 6.

6. Solving the consensus ranking problem

This section studies the optimization problem (introduced as Problem (2) in Section 4.2) that needs to be solved in order to

Table 5. The Borda Count rarely finds the expected consensus ranking

| $|A|$ | Type I | Type II | Type III |
|---|---|---|---|
| 5 | 10 | 15 | 5 | 10 | 15 | 5 | 10 | 15 |
| $B_1$ | 0 | 0 | 0 | 11 | 27 | 43 | 6 | 9 | 13 |
| $W_1$ | 0 | 0 | 0 | 28 | 48 | 48 | 6 | 18 | 27 |
| $B_2$ | 0 | 0 | 0 | 14 | 27 | 38 | 6 | 8 | 15 |
| $W_2$ | 0 | 0 | 0 | 13 | 26 | 37 | 7 | 12 | 20 |
| $B_3$ | 50 | 50 | 50 | 5 | 7 | 3 | 1 | 0 | 0 |
| $W_3$ | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |

Individual table entries in bold denote that, for the corresponding scoring function, the Borda Count achieved the optimal number of exact matches over the full set of instances in the respective experiment.
find a consensus ranking. In the Complete-ranking Aggregation Problem (CAP), each judge’s ranking is a complete ranking (strict or non-strict), and the distance in Problem (2) is the Kemeny–Snell distance, $d_{KS}(\cdot, \cdot)$. Bartholdi et al. (1989) showed that the CAP is NP-hard when restricting the input to be strict complete rankings. Since the CAP contains the strict CAP as a special case, the CAP is also NP-hard.

In the IAP, each judge’s ranking in Problem (2) is a possibly incomplete ranking (strict or non-strict). Since Bogart’s distance, the Cook–Kress–Seiford distance, the PKS distance, and our NPKS distance are all equivalent up to scaling factors to the Kemeny–Snell distance in the space of complete rankings, it follows that the IAP, using any of these distances, is also NP-hard.

The ensuing paragraphs develop an exact method for solving a special case of the IAP that applies to any of the axiomatic distances heretofore discussed. The strict-IAP is a problem identical to the IAP; except that in the strict-IAP, the solution space is restricted to the set of all strict complete rankings (the input rankings can be strict or non-strict). Denoting $r(\cdot)$ as the optimal solution to the strict-IAP, the method works by reducing the problem of finding $r(\cdot)$ to the minimum weight Feedback Arc Set Problem (FASP), otherwise known as the linear ordering problem, and then solving the transformed problem; we note that FASP is an NP-hard problem as well (Garey and Johnson, 1979).

**Definition 4.** A feedback arc set for a directed graph $G = (V, A)$ is a subset $A' \subseteq A$ such that $A'$ contains at least one arc from every directed cycle in $G$. The minimum FASP is defined as follows: Given a directed graph $G(V, A)$ with a weight assigned to each arc, find a feedback arc set, $A^*$, with a minimum sum of weights.

Given a set of objects to be ranked, $V = \{1, \ldots, n\}$, and a set of incomplete rankings, $\{a^k\}^K_{k=1}$, we find $r(\cdot)$ by solving the FASP on the weighted cumulative preference graph $G = (V, A)$ obtained from the union of the individual preference graphs $G_{a^k}$ for all $k$ (see Section 3.2). The weight of each arc $(i, j)$ in the cumulative preference graph is defined as follows:

$$w_{ij} = \sum_{k=1}^{K} w_{ij}^k,$$

(5a)

where

$$w_{ij}^k = \begin{cases} \frac{1}{\sqrt{|V_a| - |V_{a^k}|}} & \text{if } i \text{ is preferred to } j \text{ in } a^k \\ \frac{1}{\sqrt{|V_{a^k}|-|V_a|}} & \text{if } i \text{ and } j \text{ are tied in } a^k \end{cases}$$

(5b)

Figure 1 provides an example of this input transformation procedure. The three left-hand connected graphs represent the three-object individual rankings $a^1 = (1, 2, 3), a^2 = (2, 1, 3),$ and $a^3 = (1, 2, 3)$ each evaluating objects A, B, and C; the right-hand connected graph represents the corresponding weighted cumulative preference graph. For simplicity purposes and since all input rankings in the example are complete, the weights in Fig. 1 omit the common denominator $(|V|^2 - |V|)/2$.

The preferences in $r(\cdot)$ are represented by the arcs on the complement of $A^*$—the minimum weight feedback arc set in $G$. To verify that this statement is true, we note from the interpretation of the Kemeny–Snell distance (see Section 4.1 and the definition of the NPKS distance given by Equation (3)) that the NPKS distance can be interpreted as follows. The NPKS distance between two incomplete rankings $a$ and $b$ is given by the sum of the weights of the rank reversals between them, where each rank reversal has a weight of $[((|V_a| \cap |V_b|)^2 - |V_a| \cap |V_b|)/2]^{-1}$. Therefore, $r(\cdot)$ is the complete strict ranking with the minimum sum of rank reversals’ weights between $r(\cdot)$ and each of the given incomplete rankings $\{a^k\}^K_{k=1}$. Now, since $(i, j) \in A^*$ implies $r(\cdot)$ prefers $j$ over $i$ (i.e., it can be proven $A^*$ contains exactly one of $(i, j)$ and $(j, i)$ because $G$ is a complete graph), $r(\cdot)$ has a rank reversal with every given incomplete ranking preferring $i$ over $j$. As a result, the sum of the rank reversals’ weights is exactly the weight of arc $(i, j)$ and, therefore, minimizing these arc weights is equivalent to minimizing the rank reversals’ weights between $r(\cdot)$ and the given incomplete rankings. Based on this equivalence, we remark that, since the FASP is approximable within $O(|V| \log |V| \log \log |V|)$ (Even et al., 1995), the strict-IAP is equivalently approximable.

Moreno-Centeno and Karp (2013) applied the Implicit Hitting Set (IHS) approach to solve the maximum-weight trace problem, which is closely related to the FASP. Since solving the FASP is equivalent to solving the strict-IAP, we modified their methodology appropriately and then assessed how their observed practicality translates to the solution of strict-IAP instances of moderate sizes. To this end, we performed two experiments that transform a set of incomplete rankings into their weighted cumulative preference graph and then solve the corresponding FASP via the IHS approach. The experiments were performed on a machine equipped with 32 GB of RAM memory and a quad-core 3.6 GHz Intel E5-1620 processor. The operating system of the machine was CentOS release 5.4 Linux. The code was written in C++ using a Concert Technology C++ interface for the MIP solver CPLEX (version 12.4). Concert and CPLEX are registered trademarks of IBM, Inc.

The first experiment fixed the number of objects to 32 and generated incomplete-ranking instances for an increasing number of judges; the second fixed the number of judges to 32 and generated incomplete-ranking instances for an increasing number of objects. Both tested two types of instances: Type III, described in Section 5.5, and Type IV, which are constructed
from a random set of complete-sized strict-IAP problems efficiently: (a) experiment 1: 32-object instances and (b) experiment 2: 32-judge instances.

Figure 2. The IHS approach solves practical-sized strict-IAP problems efficiently: (a) experiment 1: 32-object instances and (b) experiment 2: 32-judge instances.

We introduced a unifying framework for distance-based ranking aggregation by defining an axiomatic incomplete-ranking aggregation method. In addition to aggregating strict and non-strict (without ties and with ties) incomplete rankings adequately, the framework is equivalent to Kememy and Snell's ranking aggregation method on the subspace of complete rankings. Thus, it can also aggregate strict and non-strict complete rankings adequately. Although other existing distance-based methods can be adapted to aggregate incomplete rankings, we showed that they are unintentionally biased toward certain types of consensus solutions. In addition, they were inadequate in attaining the correct consensus ranking of random instances with predictable consensus rankings. The comprehensiveness of the proposed method also extends to the featured algorithm that finds a strict median ranking. Particularly, the algorithm can be utilized to aggregate strict and non-strict rankings that are complete or incomplete utilizing the herein-defined distance as well as other distances.

The featured framework can also be applied to situations beyond those discussed in the body of this article. In particular, its application can be readily extended to scenarios where the number of possible rankings is strictly smaller than the number of objects to be ranked. This is accomplished by creating a one-to-one relationship between the available ranking categories and their corresponding ordinal values. For example, in the context of the NSF proposal review process, the judgments “fund,” “revise,” and “do not fund” would map to the ordinal values 1, 2, and 3, respectively. Another application of the highlighted framework involves aggregating incomplete rankings where the unranked objects are assumed to be strictly worse than the ranked objects. For every such incomplete ranking, one can construct a corresponding complete ranking where the initially unranked objects are all tied for position $|V|$ (i.e., last place) and, consequently, the traditional Kememy–Snell distance would yield equivalent results up to a scaling factor (see Section 6). That said, in this case, finding the strict median ranking can be done with the implicit hitting set approach as proposed in Section 6.
8. Future research

The consensus ranking obtained via the NPKS distance may not be unique. However, this is also the case for aggregating complete rankings via the PKS distance or for finding a unique median among a set of points, in general (Kemeny and Snell, 1962). Thus, this feature is inherited from the general distance-based framework rather than from our method. Nevertheless, as was demonstrated via a comprehensive set of experiments in Section 5.5, the NPKS distance achieved a unique and correct median ranking in more instances than any other of the featured ranking aggregation methods. We opine that a topic that merits its own study is the development of an efficient process for determining the total number of alternative median rankings and for assessing their respective benefits.

Notes on contributors

Erick Moreno-Centeno is an Assistant Professor at Texas A&M University’s Department of Industrial and Systems Engineering. He received an M.S. and a Ph.D. in Industrial Engineering & Operations Research (2006 and 2010, resp.) and a M.S. in Computer Science (2010), all from the University of California at Berkeley. He received a B.S. in Industrial Physics Engineering from the Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), Mexico (2002). His research interests are in the general area of mathematical programming, particularly in the subareas of decision-making, design and analysis of combinatorial optimization models and algorithms, and roundoff-error-free optimization methods.

Adolfo R. Escobedo is a Ph.D. candidate at the Texas A&M University’s Department of Industrial and Systems Engineering (expected graduation date of May 2016). He received a B.A. degree in Mathematics from California State University, Los Angeles, in 2009. His research interests lie in the areas of mathematical programming and computing and their application to decision-making, power systems, and optimization solvers.

References

Bansal, M. S. and Fernández-Baca, D. (2009) Computing distances between partial rankings. Information Processing Letters, 109(4), 238–241.
Bartholdi, J., Tovey, C. A. and Trick, M. A. (1989) Voting schemes for which it can be difficult to tell who won the election. Social Choice and Welfare, 6, 157–165.
Bogart, K.P. (1973) Preferences structures I: distances between transitive preference relations. Journal of Mathematical Sociology, 3, 49–67.
Bogart, K.P. (1975) Preference structures II: distances between asymmetric relations. SIAM Journal on Applied Mathematics, 29(2), 254–262.
Cochran, W.G. and Cox, G.M. (1957) Experimental Design, Wiley, New York, NY.
Cook, W.D. (2006) Distance-based and ad hoc consensus models in ordinal preference ranking. European Journal of Operations Research, 172, 369–385.
Cook, W.D., Golany, B., Kress, M., Penn, M. and Raviv, T. (2005) Optimal allocation of proposals to reviewers to facilitate effective ranking. Management Science, 51, 655–661.
Cook, W.D., Golany, B., Penn, M. and Raviv, T. (2007) Creating a consensus ranking of proposals from reviewers’ partial ordinal rankings. Computers and Operations Research, 34, 954–965.
Cook, W.D. and Kress, M. (1985) Ordinal ranking with intensity of preference. Management Science, 31, 26–32.
Cook, W.D., Kress, M. and Seiford, L.M. (1986) Information and preference in partial orders: a bimatrix representation. Psychometrika, 51, 197–207.
Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2001) Introduction to Algorithms, second edition, The MIT Press, Cambridge, MA.
Critchlow, D. (1985) Metric Methods for Analyzing Partially Ranked Data, Springer-Verlag, Berlin, Germany.
De Borda, J.C. (1785) Mémoire sur les élections au scrutin, in Historie de l’Académie Royale des Sciences.
Dwork, C., Kumar, R., Naor, M. and Sivakumar, D. (2001) Rank aggregation methods for the web, in Proceedings of the 10th International Conference on the World Wide Web, ACM, New York, NY, pp. 613–622.
Even, G., Naor, J., Schieber, B. and Sudan, M. (1995) Approximating minimum feedback sets and multi-cuts in directed graphs, in Proceedings of the Fourth International Conference on Integer Programming and Combinatorial Optimization, Springer-Verlag, Berlin, pp. 14–48.
Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D. and Vee, E. (2006) Comparing partial rankings. SIAM Journal on Discrete Mathematics, 20(3), 628–648.
Garey, M.R. and Johnson, D.S. (1979) Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Co., New York, NY.
Gibbons, J. (1971) Nonparametric Statistical Inference, McGraw-Hill, New York, NY.
Goddard, S.T. (1983) Ranking in tournaments and group decisionmaking. Management Science, 29(12), 1384–1392.
González-Pachón, J. and Romero, C. (1999) Distance-based consensus methods: a goal programming approach. Omega, 27(3), 341–347.
González-Pachón, J. and Romero, C. (2001) Aggregation of partial ordinal rankings: an interval goal programming approach. Computers & Operations Research, 28(8), 827–834.
Hochbaum, D.S. and Levin, A. (2006) The k-allocation problem and its variants, in Proceedings of the 4th Workshop on Approximation and Online Algorithms, Lecture Notes in Computer Science, vol. 4368, Springer-Verlag Berlin Heidelberg, Germany, pp. 253–264.
Kemeny, J.G. and Snell, L.J. (1962) Preference ranking: an axiomatic approach, in Mathematical Models in Social Science, MIT Press, Boston, MA, pp. 9–23.
Levin, J. and Nalebuff, B. (1995) An introduction to vote-counting schemes. The Journal of Economic Perspectives, 9(1), 3–26.
Moreno-Centeno, E. and Karp, R.M. (2013) The implicit hitting set approach to solve combinatorial optimization problems with an application to multigenome alignment. Operations Research, 61(2), 453–468.
Stengel, D.N. (2013) Aggregating incomplete individual ratings in group resource allocation decisions. Group Decision and Negotiation, 22(2), 235–258.
Yates, F. (1936) Incomplete randomized blocks. Annals of Eugenics, 7, 121–140.