Bilevel integer linear models for ranking items and sets

Martine Labbé, Mercedes Landete, Juan F Monge

Operations Research Perspectives 10 (2023) 100271

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

Bilevel integer linear models for ranking items and sets

Martine Labbé [a,b], Mercedes Landete [c], Juan F. Monge [c,*]

[a] Computer Science Department, Université Libre de Bruxelles, 1050 Brussels, Belgium
[b] INOCS Team, INRIA, Lille, France
[c] Center of Operations Research, University Miguel Hernandez of Elche, 03202 Elche, Spain

ARTICLE INFO

Keywords:
Linear Ordering Problem
Rank aggregation problem
Kendall-distance

ABSTRACT

Item and set orderings help with data management. Depending on the context, it is just as important to order a list of items (customers from different provinces, companies from different sectors, players from different teams) as it is to order a list of sets of these items (provinces, sectors, teams). It is evident that the order that is chosen for the items is not independent of the order that is chosen for the sets. It is possible that several set orders are sensible for the same item order and vice versa, that several item orders are sensible for the same set order. In this work, we propose a bilevel model to calculate an adequate order of items when an order of sets is available and another bilevel model to calculate an adequate order of sets when an order of items is available. In addition, it is shown how to reduce both bilevel models to single level models. Two illustrative computational studies are presented, the first with collected on 25 tennis players and ATP statistics and the second with Biomedical data. Both examples illustrate the good behavior of the models and the interest of their application in a real case scenario

1. Introduction

The ordering of elements and sets is a hot topic due to the heyday of machine learning algorithms and the increase in data accessibility. Consequently, there are many different methods in the literature to rank items or sets.

The rank aggregation problem consists of obtaining a consensus ranking that combines multiple rankings of items, see for example [1,2] and the references therein.

The Kemeny ranking problem also combines multiple rankings by identifying the ranking at the smallest Kendall-τ distance [3,4].

The Linear Ordering Problem (LOP) assumes that for each pair of items in the list of items to be ranked there is a benefit of ranking one before the other and it consists in finding a linear ranking that maximizes the sum of benefits. The squared matrix M which has a row for each item and whose values m_{ij} are the benefit of ranking item i before item j is called the preference matrix. The preference matrix values can be obtained in many different ways: by expert knowledge, by computing the number of rankings in a sample in which one item is ranked before the other or by following other different algorithms, see [5] for different computing approaches. LOP returns the simultaneous permutation of rows and columns of the preference matrix that maximizes the sum of values above the diagonal and therefore the sum of the *preferences* satisfied by the ranking. Equivalently, LOP finds an acyclic tournament in a complete weighted digraph with maximum weight, where items are the nodes in the graph and the preferences are the weights. The book [6] is a good text to find out more about its properties and potential. On the other hand, the solution of Kemeny problem can be obtained by solving the LOP (see [7] for the equivalence).

There are many contexts in which subset rankings are important and many methods for ranking multiple subsets from a ranking of items. Chapter [8] focus on these criteria and methods. The paper [9] uses the Borda rule with this goal. And in [10], the authors propose a Kendall distance-based method and compare it with the mean and median rank procedures: they prove some robustness properties of their Kendall distance-based method and compare it to the classical mean and median rank methods.

In [11], the authors introduce the Generalized Linear Ordering Problem where the items to rank are arranged in clusters. The chapter [12] is a comprehensive survey on rank aggregation procedures and the papers [13,14] analyze the impact of ties in ranking procedures.

In this paper we pursue two goals. The first goal is to define a linear order of items from a linear order of sets. It could be the case when from the evaluation of a set of projects developed by groups of researchers we want to rank the researchers. The second goal is to define a linear order of sets from a linear order of items. It could be the case when researchers are ranked and several combinations of them apply for a project.

* Corresponding author.
E-mail addresses: martine.labbe@ulb.be (M. Labbé), landete@umh.es (M. Landete), monge@umh.es (J.F. Monge).

https://doi.org/10.1016/j.orp.2023.100271
Received 22 July 2022; Received in revised form 24 January 2023; Accepted 16 February 2023
Available online 24 February 2023

It would be desirable that if from a ranking of sets we obtain a ranking of items, then this would be such that the order of sets obtained from it coincides with the initial ranking. Similarly, if you start from a ranking of items. While this property is desirable, in general this ranking may not exist. A desirable objective is to find the order that, when used to order again, gives an order that approximates the original (initial) order as much as possible.

In order to achieve our goals, we propose two bi-objective problems. First, it is assumed that a linear order of sets of items is given. Then, we propose the problem of obtaining a ranking of items that achieves two different objectives: on one hand to be coherent with the preference matrix that follows from the input ranking of sets, and on the other hand, to induce a ranking of sets close to the input ranking. Secondly, it is assumed that a linear order of items is given and we propose an analogous problem to calculate the ranking of items. Instead of designing an algorithm to calculate the efficient frontier for both objectives, in this work we propose two bilevel models with single objectives that are a linear combination of the two mentioned objectives.

Since the problem of ordering items/sets by taking care of the future order of sets/items has not been addressed so far, the content of this paper addresses this research gap by introducing and evaluating two bi-level models.

The main contributions of this work can be detailed as follows:

i. A maximization non-linear bilevel mixed-integer optimization model is proposed for obtaining a ranking of items from a ranking of a sets of these items.
ii. A maximization non-linear bilevel mixed-integer optimization model is proposed for obtaining a ranking of sets of items from a ranking of the items.
iii. Linear one-level reformulations are proposed for both models.
iv. Properties of the optimal solutions are analyzed and valid inequalities for the models are proposed.
v. A relaxation method is proposed for solving the one-level models, which have an excessive number of constraints.
vi. Two applications are used to illustrate the performance of the models and the properties of our optimal rankings.

In the following section, the preliminaries necessary for the development of the work are introduced. Section 3 reviews existing models for ordering items from sets and vice versa. The limitation of the existing models is highlighted and different ways to generalize these results are proposed. In Section 4, the new models under study are introduced. In the same section, how to solve each of the models is analyzed, relaxing the bi-level condition and linearizing the non-linear constraints. Section 4 concludes with a generalization of the introduced models. In Section 5, the bilevel models are applied to two different kind of data, ATP data and biomedical data. Some conclusions are discussed in the last section.

2. Preliminaries

A ranking is a permutation. If we say ranking of items, it is because it is a permutation of items and if we say ranking of sets, it is because it is a permutation of sets. Hereinafter, we will use the letter π to refer to a ranking of items $(a, b, c, \dots)$ and the letter ρ to refer to a ranking of sets $(V_1, V_2, V_3, \dots)$. If the ranking of items is the input data, we add a tilde to the letter, i.e. we write $\tilde{\pi}$. If the ranking of sets is the input data, we also write $\tilde{\rho}$. Letters without the tilde symbol represent output data or variables. A preference matrix is a non-symmetric square matrix whose values indicate preferences: the value in row i and column j indicates how much better i is than j. Given a set of objects to be ordered, whether these objects are items or sets, and a preference matrix for the objects, the classic way to order it is to solve the LOP.

The Kendall-τ distance between two permutations λ_1 and λ_2 of n items (items or sets) is given by:

$$d_{K\text{-}\tau}(\lambda_1, \lambda_2) = |\{(i,j) : i < j, (\lambda_1(i) < \lambda_1(j) \wedge \lambda_2(i) > \lambda_2(j)) \vee ((\lambda_1(i) > \lambda_1(j) \wedge \lambda_2(i) < \lambda_2(j)))\}|$$

where, $\lambda_1(i)$ and $\lambda_2(i)$ are the positions of item i in λ_1 and λ_2 respectively.

The distance $d_{K\text{-}\tau}(\lambda_1, \lambda_2)$ is the number or pairwise disagreements between the two permutations λ_1 and λ_2. The larger the distance, the more different the permutations are. For instance, if we had three items, the distance from permutation 123 to permutations 132, 231 and 321 is 1, 2 and 3 respectively. The maximum number of disagreements that may occur between two permutations λ_1 and λ_2 of n items is $n(n-1)/2$, and in this case λ_1 gives the reverse order to λ_2. This bound for the maximum number of disagreements allows to normalize the Kendall-τ distance in the following way:

$$\frac{d_{K\text{-}\tau}(\lambda_1, \lambda_2)}{n(n-1)/2}.$$

In [15], the author uses the relative distance of a permutation to define the Concordance Coefficient; it is used to measure the ordinal association between quantity and quality measures when two or more samples are considered. The author proposes this new measure as an alternative to the non-parametric mean rank-based methods to compare two or more samples, and compares it with the classical Kruskal–Wallis method. Note that a mean rank is another way to rank sets from an order of items.

3. Known models for rankings of sets and items and its updates

One way of obtaining a linear order of items from a linear order of sets is to compute a preference matrix of the items that follows from the ranking of the sets and then to solve the LOP for this preference matrix. Likewise, a way of obtaining a linear order of sets from a linear order of items, is to compute a preference matrix of the sets that follows from the ranking of the items and then to solve the LOP for this preference matrix.

Let $V = \{1, \dots, n\}$ the items under analysis and let V_1, V_2, …, V_b different subsets of V. If $\tilde{\pi}$ is an order of the items in V, we define the set preference matrix $M^{\tilde{\pi}}$ as the $b \times b$ matrix with $m^{\tilde{\pi}}_{rs}$ equal to the number of times an item in V_r is ranked before an item in V_s in $\tilde{\pi}$. The preference matrix $M^{\tilde{\pi}}$ records the number of times that than an item in the subset V_r is ranked before an item in the subset V_s. Likewise, if $\tilde{\rho}$ is an order of the sets $V_1, V_2, \dots, V_b$, the item preference matrix $M^{\tilde{\rho}}$ is the $n \times n$ matrix with $m^{\tilde{\rho}}_{ij}$ equal to the number of times that item i precedes item j in $\tilde{\rho}$.

The following examples show how to obtain these preference matrices and how they are used to obtain the rankings. Example 1 illustrates the computing of a ranking of items from a ranking of sets and Example 2 illustrates the computing of a ranking of sets from a ranking of items.

Example 1. Let $\tilde{\rho} = (V_1|V_2|V_3|V_4|V_5)$ be a ranking of subsets of items in V, where $V_1 = (ac)$, $V_2 = (bd)$, $V_3 = (bc)$, $V_4 = (be)$ and $V_5 = (ad)$ are subsets of $V = \{a, b, c, d, e\}$ and the items in V have to be ranked. The preference matrix $M^{\tilde{\rho}}$ for items from the ranking $\tilde{\rho}$ is the following:

$$M^{\tilde{\rho}} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{pmatrix} - & 3 & 1 & 2 & 1 \\ 3 & - & 1 & 3 & 2 \\ 2 & 4 & - & 3 & 2 \\ 1 & 2 & 1 & - & 1 \\ 1 & 0 & 0 & 1 & - \end{pmatrix} \end{array},$$

The optimal value for the corresponding LOP is 23 and it is achieved by these five different alternative optimal solutions : $\pi_1 = (c, a, b, d, e)$, $\pi_2 = (c, a, b, e, d)$, $\pi_3 = (c, b, a, d, e)$, $\pi_4 = (c, b, a, e, d)$ and $\pi_5 = (c, b, e, a, d)$. Moreover, $\sum_{ij} m^{\tilde{\rho}}_{ij} - 23 = 11$ is the Kendall-τ distance from these orders of items π_i $(i = 1, \dots, 5)$ and the partial order

Table 1
Ranking of game performance corresponding to the 25 players for whom we have available data for all these factors.

	C1	C2	C3	C4	C5	C6	C7	C8	C9
1	Pete Sampras	Rafael Nadal	Rafael Nadal	Roger Federer	Pete Sampras	Rafael Nadal	Andre Agassi	Rafael Nadal	Rafael Nadal
2	Roger Federer	Novak Djokovic	Roger Federer	Pete Sampras	Roger Federer	Novak Djokovic	Rafael Nadal	Novak Djokovic	Novak Djokovic
3	Tomas Berdych	J.M. del Potro	Novak Djokovic	Novak Djokovic	Rafael Nadal	Andy Murray	Andy Murray	Andy Murray	Andy Murray
4	Robin Soderling	Andre Agassi	Andre Agassi	Rafael Nadal	Novak Djokovic	Stefan Edberg	Novak Djokovic	Andre Agassi	Andre Agassi
5	Patrick Rafter	Jim Courier	P. Kohlschreiber	J.M. del Potro	Andre Agassi	Roger Federer	Stefan Edberg	Stefan Edberg	Stefan Edberg
6	Todd Martin	Roger Federer	Stan Wawrinka	Patrick Rafter	Todd Martin	Tim Henman	Jim Courier	Robin Soderling	Roger Federer
7	Nicolas Almagro	Radek Stepanek	Jim Courier	Tomas Berdych	Patrick Rafter	Andre Agassi	Daniil Medvedev	Jim Courier	Jim Courier
8	Andy Murray	Patrick Rafter	J.M. del Potro	Jim Courier	Stan Wawrinka	M. Washington	Tomas Berdych	M. Washington	Tim Henman
9	J.M. del Potro	Richard Gasquet	Richard Gasquet	Andre Agassi	J.M. del Potro	Richard Gasquet	Todd Martin	J.M. del Potro	M. Washington
10	Tim Henman	Todd Martin	Nicolas Almagro	Robin Soderling	Stefan Edberg	James Blake	J.M. del Potro	James Blake	J.M. del Potro
11	F. Gonzalez	Javier Frana	Pete Sampras	Todd Martin	F. Gonzalez	J.M. del Potro	M. Washington	Tomas Berdych	James Blake
12	Tommy Haas	Stefan Edberg	Tomas Berdych	F. Gonzalez	Javier Frana	Stan Wawrinka	Robin Soderling	Roger Federer	Daniil Medvedev
13	Jim Courier	Robin Soderling	Tommy Haas	Stan Wawrinka	Tomas Berdych	Radek Stepanek	Pete Sampras	Richard Gasquet	Richard Gasquet
14	Daniil Medvedev	F. Gonzalez	Andy Murray	Stefan Edberg	Jim Courier	F. Gonzalez	Roger Federer	Tim Henman	Robin Soderling
15	Stan Wawrinka	P. Kohlschreiber	Daniil Medvedev	Nicolas Almagro	Robin Soderling	Jim Courier	James Blake	Pete Sampras	Pete Sampras
16	Novak Djokovic	Pete Sampras	Stefan Edberg	Richard Gasquet	P. Kohlschreiber	Robin Soderling	Richard Gasquet	Javier Frana	Tomas Berdych
17	Richard Gasquet	Daniil Medvedev	Robin Soderling	Andy Murray	Tim Henman	Daniil Medvedev	Javier Frana	Radek Stepanek	Tommy Haas
18	Stefan Edberg	M. Washington	Patrick Rafter	P. Kohlschreiber	Andy Murray	Nicolas Almagro	Tim Henman	Daniil Medvedev	Stan Wawrinka
19	Javier Frana	Tommy Haas	M. Washington	Tommy Haas	Tommy Haas	P. Kohlschreiber	Patrick Rafter	F. Gonzalez	Radek Stepanek
20	Andre Agassi	Andy Murray	F. Gonzalez	Daniil Medvedev	Daniil Medvedev	Patrick Rafter	Tommy Haas	Todd Martin	Patrick Rafter
21	Radek Stepanek	Nicolas Almagro	Javier Frana	Tim Henman	Nicolas Almagro	Tomas Berdych	P. Kohlschreiber	P. Kohlschreiber	Todd Martin
22	James Blake	Tomas Berdych	James Blake	Radek Stepanek	Radek Stepanek	Tommy Haas	Radek Stepanek	Tommy Haas	F. Gonzalez
23	P. Kohlschreiber	Stan Wawrinka	Todd Martin	Javier Frana	James Blake	Javier Frana	F. Gonzalez	Patrick Rafter	P. Kohlschreiber
24	M. Washington	James Blake	Radek Stepanek	M. Washington	Richard Gasquet	Pete Sampras	Nicolas Almagro	Stan Wawrinka	Javier Frana
25	Rafael Nadal	Tim Henman	Tim Henman	James Blake	M. Washington	Todd Martin	Stan Wawrinka	Nicolas Almagro	Nicolas Almagro

C1: Career 1st Serve Points Won On All Surfaces From All Countries.
C2: Career 1st Serve On All Surfaces From All Countries.
C3: Career 2nd Serve Points Won On All Surfaces From All Countries.
C4: Career Service Games Won On All Surfaces From All Countries.
C5: Career Break Points Saved On All Surfaces From All Countries.
C6: Career 1st Serve Return Points Won On All Surfaces From All Countries.
C7: Career 2nd Serve Return Points Won On All Surfaces From All Countries.
C8: Career Break Points Converted On All Surfaces From All Countries.
C9: Career Return Games Won On All Surfaces From All Countries.
Data from http://www.atpworldtour.com/.

$\tilde{\rho}$, for example, $d_{K\text{-}\tau}(\pi_1, \tilde{\rho})$ is the Kendall distance between permutations $(ccaabbbdde)$ and $\tilde{\rho} = (V_1|V_2|V_3|V_4|V_5) = (ac|db|bc|be|ad)$, where $(ccaabbbdde)$ is obtained from π_1 by replicating each item the number of times it appears in the subsets $V_1, \ldots, V_5$. Note that $\tilde{\rho}$ is a partial order of items in V, i.e., $\tilde{\rho}$ is also represented by permutation $\tilde{\rho} = (ca|db|cb|be|ad)$, where the items in each set can be permuted to their position. The number of disagreements of $\tilde{\rho} = (ca|db|cb|be|ad)$ from $(ccaabbbdde)$ is exactly 11.

Example 2. Let $\tilde{\pi} = (a, b, c, d, e)$ be a ranking of the items in V, and $V_1 = (ad)$, $V_2 = (ae)$, $V_3 = (bc)$, $V_4 = (bd)$ and $V_5 = (ce)$ subsets of V which have to be ranked. Thus, the preference matrix $M^{\tilde{\pi}}$ for sets from the ranking $\tilde{\pi}$ is the following:

$$M^{\tilde{\pi}} = \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{array} \begin{array}{c} \begin{array}{ccccc} V_1 & V_2 & V_3 & V_4 & V_5 \end{array} \\ \begin{pmatrix} - & 2 & 2 & 2 & 3 \\ 1 & - & 2 & 2 & 2 \\ 2 & 2 & - & 2 & 3 \\ 1 & 2 & 1 & - & 3 \\ 1 & 1 & 0 & 1 & - \end{pmatrix} \end{array},$$

The optimal value for the corresponding LOP is 23 and it is achieved by five different alternative optimal solutions: $\rho_1 = (V_1|V_2|V_3|V_4|V_5)$, $\rho_2 = (V_1|V_3|V_2|V_4|V_5)$, $\rho_3 = (V_1|V_3|V_4|V_2|V_5)$, $\rho_4 = (V_3|V_1|V_2|V_4|V_5)$ and $\rho_5 = (V_3|V_1|V_4|V_2|V_5)$. Moreover, $\sum_{rs} m^{\tilde{\pi}}_{rs} - 23 = 12$ is the Kendall-τ distance from these orders of sets ρ_i $(i = 1, \ldots, 5)$ and the order $\tilde{\pi}$, for example, $d(\rho_1, \tilde{\pi})$ is the Kendall distance between permutations $\rho_1 = (V_1|V_2|V_3|V_4|V_5) = (ad|ae|bc|bd|ce)$ and $(aa|bb|cc|dd|ee)$, which comes from $\tilde{\pi}$ replicating each item the number of times it appears in $V_1, \ldots, V_5$.

Obtaining a linear order of items from an order of sets is equivalent to the problem introduced by Aparicio et al. [10], under some assumptions. The next example illustrates this feature.

Example 3 (*Example 1 Continuation*)**.** Ranking $\tilde{\rho} = (ac|bd|bc|be|ad)$ is a partial order of items in multisets $A = (a, a)$, $B = (b, b, b)$, $C = (c, c)$, $D = (d, d)$ and $E = \{e\}$, where some items in $\tilde{\rho}$ are not comparable. Applying [10], the order of multisets from the partial order of items $\tilde{\rho}$ is the solution of the linear ordering problem (LOP) with a matrix $M^{\tilde{\rho}} = m_{rs}$ where $m^{\tilde{\rho}}_{rs}$ is the number of times item r precedes item s in $\tilde{\rho}$. The resulting preference matrix and thus the resulting rankings are those in Example 1.

4. New models for ranking sets and items

In this section, we propose new procedures for computing ranking of sets from ranking of items and vice-versa. We focus on the following two objectives

- Objective 1. From an input ranking of items/sets, maximize the upper diagonal sum of the preference matrix. The solution is a ranking of sets/items.
- Objective 2. Minimize the Kendall-τ distance between the input ranking of items/sets and the ranking of items/sets induced by the solution of Objective 1.

Both objectives are relevant: if the second goal is omitted, the number of alternative optimal solutions may be large (see Examples 1 and 2); if the first goal is omitted, the resulting ranking may be incoherent with the input data. If the final ranking is the LOP solution for the preference matrix, the second goal can be seen as a way of discriminating between alternative optimal solutions.

Table 2
Ranking of items.

Player	Mean-rank ranking	Player	Bi-level ranking
Novak Djokovic	4,22	Rafael Nadal	1
Rafael Nadal	4,33	Novak Djokovic	2
Roger Federer	5,56	Roger Federer	3
Andre Agassi	6,44	Andre Agassi	4
J.M. del Potro	8,22	Jim Courier	5
Jim Courier	9,11	J.M. del Potro	6
Andy Murray	9,89	Pete Sampras	7
Stefan Edberg	9,89	Andy Murray	8
Pete Sampras	10,89	Stefan Edberg	9
Robin Soderling	11,89	Robin Soderling	10
Tomas Berdych	12,56	Tomas Berdych	11
Patrick Rafter	14,00	Patrick Rafter	12
Richard Gasquet	14,00	Richard Gasquet	13
Todd Martin	14,56	Todd Martin	14
Daniil Medvedev	15,56	Daniil Medvedev	15
Tim Henman	16,00	Stan Wawrinka	16
Stan Wawrinka	16,00	Tim Henman	17
F. Gonzalez	16,22	MaliVai Washington	18
MaliVai Washington	16,22	F. Gonzalez	19
P. Kohlschreiber	17,89	Tommy Haas	20
James Blake	18,00	P. Kohlschreiber	21
Tommy Haas	18,11	Radek Stepanek	22
Nicolas Almagro	18,44	James Blake	23
Javier Frana	18,44	Javier Frana	24
Radek Stepanek	18,56	Nicolas Almagro	25

We distinguish two parallel subsections. In Section 4.1, the approach for obtaining a linear order of items from a linear order of sets is presented and in Section 4.2, the approach for the symmetric case is discussed.

4.1. Linear order of items from a linear order of sets

To start the section, the two different objectives are illustrated with the data in Example 1. In the previous section, all alternative optimal solutions for the preference matrix had been obtained. In this section, the ranking of sets associated with each alternative optimal solution is calculated, as well as its distance from the initial set order. The solution of the bi-objective problem is illustrated and the interest of the two objectives is shown. The example shows that it is not sensible that the only objective is the proximity of the two ranking sets because we could propose a ranking of items that is not consistent with the initial ranking of sets. It also illustrates that the problem with Objective 1 alone can have a lot of alternative solutions.

Example 4 (*Example 1 Continuation*). The ranking of items π_1 obtained from $\tilde{\rho}$ can be used to obtain a linear order of a set. The preference matrix for the sets of V from π_1 is

$$M^{\pi_1} = \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{array} \begin{array}{c} \begin{array}{ccccc} V_1 & V_2 & V_3 & V_4 & V_5 \end{array} \\ \begin{pmatrix} - & 4 & 2 & 4 & 3 \\ 0 & - & 0 & 2 & 1 \\ 1 & 3 & - & 3 & 3 \\ 0 & 1 & 0 & - & 1 \\ 0 & 2 & 1 & 3 & - \end{pmatrix} \end{array},$$

and this matrix induces the order of sets $\rho_1^1 = (V_1|V_3|V_5|V_2|V_4)$, and this order is at distance 3 from $\tilde{\rho}$, i.e. $d_{K-\tau}(\rho_1^1, \tilde{\rho}) = 3$

The ranking of items π_2 establishes a preference matrix for the sets of V. The preference matrix is

$$M^{\pi_2} = \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{array} \begin{array}{c} \begin{array}{ccccc} V_1 & V_2 & V_3 & V_4 & V_5 \end{array} \\ \begin{pmatrix} - & 4 & 2 & 4 & 3 \\ 0 & - & 0 & 1 & 1 \\ 1 & 3 & - & 3 & 3 \\ 0 & 2 & 0 & - & 2 \\ 0 & 2 & 1 & 2 & - \end{pmatrix} \end{array},$$

and this matrix induces the orders of sets $\rho_2^1 = (V_1|V_3|V_4|V_5|V_2)$ and $\rho_2^2 = (V_1|V_3|V_5|V_4|V_2)$, and these orders need 3 and 4 disagreements from $\tilde{\rho}$, respectively.

The ranking of items π_3 establishes a preference matrix for the sets of V. The preference matrix is

$$M^{\pi_3} = \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{array} \begin{array}{c} \begin{array}{ccccc} V_1 & V_2 & V_3 & V_4 & V_5 \end{array} \\ \begin{pmatrix} - & 3 & 1 & 3 & 3 \\ 1 & - & 0 & 2 & 2 \\ 2 & 3 & - & 3 & 4 \\ 1 & 1 & 0 & - & 2 \\ 0 & 1 & 0 & 2 & - \end{pmatrix} \end{array},$$

and this matrix induces the orders of sets $\rho_3^1 = (V_3|V_1|V_2|V_4|V_5)$ and $\rho_3^2 = (V_3|V_1|V_2|V_5|V_4)$, and these orders need 2 and 3 disagreements from $\tilde{\rho}$, respectively.

The ranking of items π_4 establishes a preference matrix for the sets of V. The preference matrix is

$$M^{\pi_4} = \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{array} \begin{array}{c} \begin{array}{ccccc} V_1 & V_2 & V_3 & V_4 & V_5 \end{array} \\ \begin{pmatrix} - & 3 & 1 & 3 & 3 \\ 1 & - & 0 & 1 & 2 \\ 2 & 3 & - & 3 & 4 \\ 1 & 2 & 0 & - & 3 \\ 0 & 1 & 0 & 1 & - \end{pmatrix} \end{array},$$

and this matrix induces the order of sets $\rho_4^1 = (V_3|V_1|V_4|V_2|V_5)$, and this order needs 3 disagreements from $\tilde{\rho}$.

The ranking of items π_5 establishes a preference matrix for the sets of V. The preference matrix is

$$M^{\pi_5} = \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{array} \begin{array}{c} \begin{array}{ccccc} V_1 & V_2 & V_3 & V_4 & V_5 \end{array} \\ \begin{pmatrix} - & 3 & 1 & 2 & 3 \\ 1 & - & 0 & 1 & 2 \\ 2 & 3 & - & 3 & 4 \\ 2 & 2 & 0 & - & 4 \\ 0 & 1 & 0 & 0 & - \end{pmatrix} \end{array},$$

and this matrix induces the orders of sets $\rho_5^1 = (V_3|V_1|V_4|V_2|V_5)$ and $\rho_5^2 = (V_3|V_4|V_1|V_2|V_5)$, and these orders need 3 and 4 disagreements from $\tilde{\rho}$, respectively.

Fig. 1 presents the sequence of problems we have solved. Therefore, order π_3 is the best rank of items from $\tilde{\rho}$.

If the first goal was not taken into account, other orders like (a, e, b, d, c), (a, b, c, e, d), which are not optimal solutions for Example 1 could be proposed since both induce set orders with 2 disagreements from $\tilde{\rho}$. However, it would be very weird since in (a, e, b, d, c), item e is the second and it does not fit with $\tilde{\rho}$ and in (a, b, c, e, d) item c is the third when it must clearly be the first.

In this paper, we propose that if a ranking of sets is given and a ranking of items is required, then this ranking of items should be an optimal solution to the model with two weighted objectives, the adequacy of the ranking of items to the ranking of sets and the proximity of the ranking of sets that would be obtained from the ranking of items to the initial ranking of sets.

Let V be a set of items and P be a set of subsets of V. Let $\tilde{\rho}$ be a linear order of sets in P and let $\tilde{m}_{ij}$ be the number of times that i appears before j in $\tilde{\rho}$. Let y_{rs} and x_{ij} be two families of binary variables: for each $r, s \in P$, $y_{rs} = 1$ iff set r goes before set s and for each $i, j \in V$, $x_{ij} = 1$ iff i goes before j. Specifically, we propose the following bi-level combinatorial optimization model.

$$(\text{BIP})^1 \max \sum_i \sum_j \tilde{m}_{ij} x_{ij} - \epsilon\, d_{K\text{-}\tau}(\rho(y), \tilde{\rho}) \tag{1}$$

$$\text{s.t. } m_{rs} = \sum_{i \in r} \sum_{j \in s} x_{ij} \qquad r, s \in P : r \neq s \tag{2}$$

Table 3
Ranking of game factor including some of the best three players GOAT.

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	Pete Sampras	Rafael Nadal	Rafael Nadal	Roger Federer	Pete Sampras	Rafael Nadal	Andre Agassi	Rafael Nadal	Rafael Nadal
2	Roger Federer	Novak Djokovic	Roger Federer	Pete Sampras	Roger Federer	Novak Djokovic	Rafael Nadal	Novak Djokovic	Novak Djokovic
3	Tomas Berdych	J.M. del Potro	Novak Djokovic	Novak Djokovic	Rafael Nadal	Andy Murray	Andy Murray	Andy Murray	Andy Murray
4	Robin Soderling	Andre Agassi	Andre Agassi	Rafael Nadal	Novak Djokovic	Stefan Edberg	Novak Djokovic	Andre Agassi	Andre Agassi
5	Patrick Rafter	Jim Courier	P. Kohlschreiber	J.M. del Potro	Andre Agassi	Roger Federer	Stefan Edberg	Stefan Edberg	Stefan Edberg
6	Todd Martin	Roger Federer	Stan Wawrinka	Patrick Rafter	Todd Martin	Tim Henman	Jim Courier	Robin Soderling	Roger Federer
7	Tommy Haas	Stefan Edberg	Tomas Berdych	F. Gonzalez	Javier Frana	Stan Wawrinka	Robin Soderling	Roger Federer	Daniil Medvedev
8	Daniil Medvedev	F. Gonzalez	Andy Murray	Stefan Edberg	Jim Courier	F. Gonzalez	Roger Federer	Tim Henman	Robin Soderling
9	Novak Djokovic	Pete Sampras	Stefan Edberg	Richard Gasquet	P. Kohlschreiber	Robin Soderling	Richard Gasquet	Javier Frana	Tomas Berdych
10	Rafael Nadal	Tim Henman	Tim Henman	James Blake	MaliVai Washington	Todd Martin	Stan Wawrinka	Nicolas Almagro	Nicolas Almagro

Rafael Nadal rank 2.66.
Novak Djokovic rank 3.44.
Roger Federer rank 4.33.

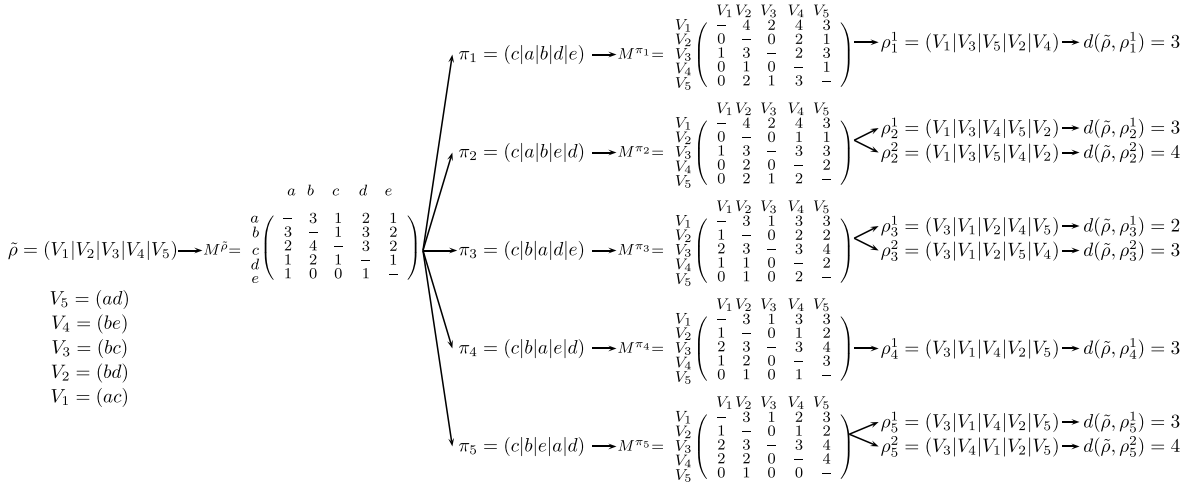

Fig. 1. Order of items from a ranking of sets.

$$x_{ij} + x_{ji} = 1 \qquad i, j \in V : i < j \tag{3}$$

$$x_{ij} + x_{jk} + x_{ki} \leq 2 \qquad i, j, k \in V : i, j, k \text{ pwd} \tag{4}$$

$$x_{ij} \in \{0, 1\} \qquad i, j \in V : i \neq j \tag{5}$$

$$m_{rs} \in \mathbb{N} \qquad r, s \in P : r \neq s \tag{6}$$

$$\max \sum_{r} \sum_{s} m_{rs} y_{rs} \tag{7}$$

$$\text{s.t. } y_{rs} + y_{sr} = 1 \qquad r, s \in P : r < s \tag{8}$$

$$y_{rs} + y_{st} + y_{tr} \leq 2 \qquad r, s, t \in P : r, s, t \text{ pwd} \tag{9}$$

$$y_{rs} \in \{0, 1\} \qquad r, s \in P : r \neq s \tag{10}$$

where ϵ is a parameter that weights the two objectives and $\rho(y)$ is the order defined by the y-variables. The first term in the objective function is the LOP objective function for the preference matrix $M^{\tilde{\rho}} = \{\tilde{m}_{ij}\}_{|V|\times|V|}$. The second term is the distance between the ranking ρ defined by the values of the y-variables and $\tilde{\rho}$. Constraints (2) allow to compute the preference matrix values for the lower level problem. Constraints (3)–(5) are the LOP constraints: (3) states that item i goes before item j or item j goes before item i; (4) states that if item i goes before item j and item j before item k, then item k cannot go before item i. "pwd" in constraint (4) stands for "pairwise disjoint", i.e. $i \neq j, i \neq k$ and $j \neq k$. Analogously, constraints (8)–(10) are the LOP constraints for the sets: (9) states that set r goes before set s or set s goes before set r; (9) states that if set r goes before set s and set s before set t, then set t cannot go before set s. Constraints (5), (6) and (10) are the domain constraints, both x-variables and y-variables are binary while m-variables are positive and integer.

Remark 4.1. The objective function of (BIP)^1 is the addition of two distances. From Examples 1 and 2, the first term of the objective function minus the sum of all the preference matrix entries is the distance between the solution rankings and the ranking that follows from the original ranking by adding as many copies of the items as sets belongs to. Thus, the objective of (BIP)^1 is also the minimization of two ranking distances.

The previous model constitutes a correct way to approach the challenge of finding a ranking of items that minimizes the two objectives of interest. However, in order to solve real problems with this model, it is necessary to overcome three important drawbacks. First of all, it is necessary to rewrite the second term of the objective function in a linear way, since by definition, the Kendall-τ distance is the sum of absolute values. Secondly, it is necessary to decide how the two-level optimization is approached, to analyze if it is possible to transform this model into an equivalent one-level model or if we must resort to some suitable resolution algorithm. Finally, the product of two variables appearing in the objective function of the second level must be linearized. The following results answer these questions.

Proposition 4.1. *Let $\tilde{y}_{rs}$, for all $r, s \in P$, be the binary values that describe the ranking $\tilde{\rho}$. Then, the second term in (1) can be re-written as*

$$\sum_{rs:\tilde{y}_{rs}=0} y_{rs} - \sum_{rs:\tilde{y}_{rs}=1} y_{rs}.$$

Proof. Since $\tilde{y}$ is a linear order, the following equalities hold:

$$d_{K\text{-}\tau}(\rho, \tilde{\rho}) = \frac{1}{2} \sum_{rs} |\tilde{y}_{rs} - y_{rs}| =$$

$$\frac{1}{2}\left(\sum_{rs}\bar{y}_{rs}+\sum_{rs:\bar{y}_{rs}=0}y_{rs}-\sum_{rs:\bar{y}_{rs}=1}y_{rs}\right)=$$
$$\frac{1}{2}\left(\frac{|P|(|P|-1)}{2}+\sum_{rs:\bar{y}_{rs}=0}y_{rs}-\sum_{rs:\bar{y}_{rs}=1}y_{rs}\right).\ \square$$

A way of solving the bi-level problem is to observe that the second-level problem could be seen as the best solution among all the feasible LOP solutions. Let the statement "$\bar{y}_{rs} \in LOP$" represent all the vector parameter describing feasible permutations of P items, i.e., feasible solutions of the LOP problem for a $|P|\times|P|$ matrix. Thus, the previous bi-level problem is equivalent to the following single level problem.

$$\begin{aligned}
(\text{IP})^1 \max\ & \sum_i\sum_j \tilde{m}_{ij}x_{ij} \\
& -\epsilon\left(\sum_{rs:\bar{y}_{rs}=0}y_{rs}-\sum_{rs:\bar{y}_{rs}=1}y_{rs}\right) \\
\text{s.t. } & m_{rs}=\sum_{i\in r}\sum_{j\in s}x_{ij} && r,s\in P: r\neq s \\
& x_{ij}+x_{ji}=1 && i,j\in V: i<j \\
& x_{ij}+x_{jk}+x_{ki}\leq 2 && i,j,k\in V: i,j,k \text{ pwd} \\
& x_{ij}\in\{0,1\} && i,j\in V: i\neq j \\
& \sum_r\sum_s m_{rs}y_{rs}\geq\sum_r\sum_s m_{rs}\bar{y}_{rs} && \bar{y}_{rs}\in LOP \qquad (11) \\
& y_{rs}+y_{sr}=1 && r,s\in P: r<s \\
& y_{rs}+y_{st}+y_{tr}\leq 2 && r,s,t\in P: r,s,t \text{ pwd} \\
& y_{rs}\in\{0,1\} && r,s\in P: r\neq s
\end{aligned}$$

Constraints (11) state that the optimal values for the y-variables will correspond with the order of sets with the largest upper diagonal sum. Although there is a huge number of constraints of type (11), a good separation algorithm for them will be presented. This approach to writing a bilevel model as a single level model has been previously proposed in the literature, see for instance [16].

The product of variables $m_{rs}y_{rs}$ can be linearized by defining a new family of variables

$$z_{rsij}=y_{rs}x_{ij},$$

replacing $m_{rs}y_{rs}$ by $\sum_{i\in r}\sum_{j\in s}z_{rsij}$ in (11) (the replacement follows from Eqs. (2)) and adding the following constraints:

$$z_{rsij}\leq y_{rs} \qquad r,s\in P: r\neq s \qquad (12)$$
$$z_{rsij}\leq x_{ij} \qquad i,j\in V: i\neq j \qquad (13)$$
$$z_{rsij}\geq y_{rs}+x_{ij}-1 \qquad i,j\in V: i\neq j, r,s\in P: r\neq s, r\ni i, s\ni j \qquad (14)$$

Finally, it is interesting to point out that small values of ϵ entail that the optimal solution of BIP is the optimal solution of the LOP problem for the preference matrix. It can be proved that it is the case when ϵ is smaller than $1/(|P|(|P|-1))$.

Proposition 4.2. *Let X^* be the set of all optimal solutions of*

$$\max\ \sum_i\sum_j \tilde{m}_{ij}x_{ij}$$
$$s.t.\ (3)\text{–}(5).$$

If $\epsilon \leq 1/(|P|(|P|-1))$, then the optimal solutions of BIP^1 (or IP^1) belong to X^*.

Proof. Suppose that $\epsilon \leq 1/(|P|(|P|-1))$ and that $\bar{x}$ is an optimal solution of BIP^1 (or IP^1) and $\bar{x}\notin X^*$. We are going to prove that this assumption leads to a false inequality. The (optimal) value of BIP^1 for $\bar{x}$ is $\sum_i\sum_j \tilde{m}_{ij}\bar{x}_{ij}-\epsilon d_{K\text{-}\tau}(\rho(\bar{x}),\tilde{\rho})$, where $\rho(\bar{x})$ is the ranking of sets induced by $\bar{x}$ in model BIP^1. Moreover $\sum_i\sum_j \tilde{m}_{ij}\bar{x}_{ij}-\epsilon d_{K\text{-}\tau}(\rho(\bar{x}),\tilde{\rho})\geq\sum_i\sum_j \tilde{m}_{ij}x^*_{ij}-\epsilon d_{K\text{-}\tau}(\rho(x^*),\tilde{\rho})$ and $\sum_i\sum_j \tilde{m}_{ij}x^*_{ij}>\sum_i\sum_j \tilde{m}_{ij}\bar{x}_{ij}$ for all $x^*\in X^*$. Therefore, $-\epsilon d_{K\text{-}\tau}(\rho(\bar{x}),\tilde{\rho})\geq-\epsilon d_{K\text{-}\tau}(\rho(x^*),\tilde{\rho})+1$ or equivalently, $0\leq\epsilon d_{K\text{-}\tau}(\rho(\bar{x}),\tilde{\rho})\leq\epsilon d_{K\text{-}\tau}(\rho(x^*),\tilde{\rho})-1$. ϵ is bounded by $1/(|P|(|P|-1))$ and $d_{K\text{-}\tau}(\rho(x^*),\tilde{\rho})$ is bounded by $(|P|(|P|-1))/2$, then the last inequality becomes $0\leq(1/(|P|(|P|-1)))((|P|(|P|-1))/2)-1=-0.5$, which is false. $\square$

4.2. Linear order of sets from a linear order of items

This section presents the approach for ranking sets from a linear order of items analogous to the approach in the previous section. The two different objectives are illustrated with the data in Example 2. In the previous section, all alternative optimal solutions for the preference matrix had been obtained. The solution of the bi-objective problem is illustrated and the interest of the two objectives is again shown.

Example 5 (*Example 2 Continuation*)**.** The ranking of items ρ_1 obtained from $\tilde{\pi}$ can be used to obtain a linear order of items. A preference matrix for the items of V from ρ_1 is

$$M^{\tilde{\rho}^1}== \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array}\begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{pmatrix} - & 4 & 4 & 2 & 3 \\ 0 & - & 2 & 1 & 2 \\ 0 & 1 & - & 1 & 1 \\ 1 & 2 & 3 & - & 3 \\ 0 & 2 & 2 & 1 & - \end{pmatrix} \end{array},$$

and this matrix induces the orders of items $\pi^1_1=(a|d|b|e|c)$ and $\pi^1_2=(a|d|e|b|c)$, and these orders present 3 and 4 disagreements with $\tilde{\pi}$.

Fig. 2 presents all the combinations in this example. Therefore, order π_4 is the best order of sets.

If a ranking of items is given and a ranking of sets is required, we then propose that this ranking of sets should be an optimal solution to the model with two weighted objectives: the adequacy of the ranking of sets to the ranking of items and the proximity of the ranking of items that would be obtained from the ranking of sets to the initial ranking of items.

Let V be a set of items and P be a set of subsets of V. Let $\tilde{\pi}$ be a linear order of items in V and let $\tilde{m}_{rs}$ be the number of times that r appears before s in $\tilde{\pi}$. The following bi-level combinatorial optimization model is proposed.

$$\begin{aligned}
(\text{BIP})^2 \max\ & \sum_r\sum_s \tilde{m}_{rs}y_{rs}-\epsilon\, d_{K\text{-}\tau}(\pi,\tilde{\pi}) && (15) \\
\text{s.t. } & m_{ij}=\sum_{r\ni i}\sum_{s\ni j}y_{rs} && i,j\in V: i\neq j \qquad (16) \\
& y_{rs}+y_{sr}=1 && r,s\in P: r<s \qquad (17) \\
& y_{rs}+y_{st}+y_{tr}\leq 2 && r,s,t\in P: r,s,t \text{ pwd} \qquad (18) \\
& y_{rs}\in\{0,1\} && r,s\in P: r\neq s \qquad (19) \\
& \max \sum_{i,j\in V: i\neq j} m_{ij}x_{ij} && (20) \\
\text{s.t. } & x_{ij}+x_{ji}=1 && i,j\in V: i<j \qquad (21) \\
& x_{ij}+x_{jk}+x_{ki}\leq 2 && i,j,k\in V: i,j,k \text{ pwd} \qquad (22) \\
& x_{ij}\in\{0,1\} && i,j\in V: i\neq j \qquad (23)
\end{aligned}$$

For each $i,j\in V$, let $\tilde{x}_{ij}$ be the constants that represent ranking $\tilde{\pi}$. Then, again the Kendall-τ distance between π and $\tilde{\pi}$ is the sum of the absolute value of the differences,

$$d_{K\text{-}\tau}(\pi,\tilde{\pi})=\frac{1}{2}\sum_{ij}|\tilde{x}_{ij}-x_{ij}|$$

and the objective function can be re-written. Also the second problem can be replaced by a set of constraints. In particular, the model becomes the following model:

$$(\text{IP})^2 \max\ \sum_r\sum_s \tilde{m}_{rs}y_{rs}$$

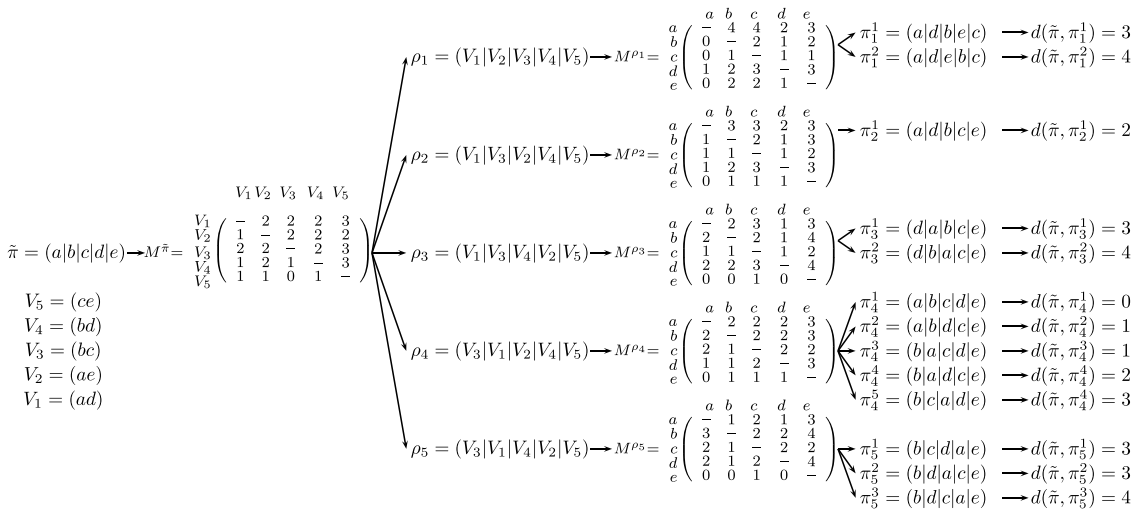

Fig. 2. Order of sets from a ranking of items.

$$
\begin{aligned}
& -\epsilon\left(\sum_{ij:\bar{x}_{ij}=0} x_{ij} - \sum_{ij:\bar{x}_{ij}=1} x_{ij}\right) && \\
\text{s.t. } & m_{ij} = \sum_{r \ni i} \sum_{s \ni j} y_{rs} && i,j \in V : i \neq j \\
& y_{rs} + y_{sr} = 1 && r,s \in P : r < s \\
& y_{rs} + y_{st} + y_{tr} \leq 2 && r,s,t \in P : r,s,t \text{ pwd} \\
& \sum_{i,j \in V : i \neq j} m_{ij} x_{ij} \geq \sum_{i,j \in V : i \neq j} m_{ij} \bar{x}_{ij} && \forall \bar{x}_{ij} \in LOP \qquad (24) \\
& x_{ij} + x_{ji} = 1 && i,j \in V : i < j \\
& x_{ij} + x_{jk} + x_{ki} \leq 2 && i,j,k \in V : i,j,k \text{ pwd} \\
& y_{rs} \in \{0,1\} && r,s \in P : r \neq s \\
& x_{ij} \in \{0,1\} && i,j \in V : i \neq j
\end{aligned}
$$

The product of variables $m_{ij}x_{ij}$ can be linearized by defining a new family of variables

$$z_{rsij} = y_{rs}x_{ij},$$

replacing $m_{ij}x_{ij}$ by $\sum_{r \ni i} \sum_{s \ni j} z_{rsij}$ in (24) (the replacement follows from Eqs. (16)) and adding constraints (12)–(14).

The last result in this section is about the influence of the magnitude of the weight ϵ. It is similar to Proposition 4.2 in the previous section.

Corollary 4.1. *Let Y^* be the set of all optimal solutions of*

$$\max \sum_r \sum_s \tilde{m}_{rs} y_{rs}$$
$$s.t.\ (17)\text{–}(19).$$

If $\epsilon \leq 1/(n(n-1))$, then the optimal solutions of BIP^2 (or IP^2) belong to Y^*.

Proof. Follows from Proposition 4.2. □

4.3. Separation algorithm

The two linear and single level models that have been proposed in the previous sections suffer from the same difficulty, a family of valid inequalities with very many inequalities. To address the resolution of these models, we propose to relax the model by eliminating this family of inequalities and iteratively and intelligently incorporating some of the inequalities of the family until the optimum of the problem is guaranteed. The algorithms in this section explain the corresponding relaxation algorithms that we propose for the two linear problems in question.

Algorithms 1 and 2 explain how to solve (IP)^1 and (IP)^2. At any iteration, a new constraint (11) (or (24)) is added to problem (IP)^1 or (IP)^2.

For example, in Algorithm 1, we solve to optimality the problem (IP)^1 without (11) constraints. If the optimal solution is optimal for the second level problem, the algorithm ends, and the optimal solution of the bilevel problem is found. Otherwise, the solution is not optimal for the second level problem, so the cut needs to be added (line 5 in Algorithm 1) to the problem (IP)^1 in order to prevent this solution.

Algorithm 1: Separation Algorithm for (IP)^1 problem

1 (IP) = model((IP)^1 without (11))
2 **repeat**
3 (x^*, m^*, y^*) = solve((IP))
4 $(\bar{y})$ = solve$\left(\max\{\sum_r \sum_w m^*_{rs} y_{rs}\ s.t.\ (8)\text{–}(10)\}\right)$
5 Add constraint $\left(\sum_{r,s} m_{rs} y_{rs} \geq \sum_{r,s} m_{rs} \bar{y}_{rs}\right)$ to (IP)
6 **until** $\left(\sum_{r,s} m^*_{rs} y^*_{rs} = \sum_{r,s} m^*_{rs} \bar{y}_{rs}\right)$;

Algorithm 2: Separation Algorithm for (IP)^2 problem

1 (IP) = model((IP)^2 without (24))
2 **repeat**
3 (x^*, m^*, y^*) = solve((IP))
4 $(\bar{x})$ = solve$\left(\max\{\sum_i \sum_j m^*_{ij} x_{ij}\ s.t.\ (21)\text{–}(23)\}\right)$
5 Add constraint $\left(\sum_{i,j} m_{ij} x_{ij} \geq \sum_{i,j} m_{ij} \bar{x}_{ij}\right)$ to (IP)
6 **until** $\left(\sum_{i,j} m^*_{ij} x^*_{ij} = \sum_{i,j} m^*_{ij} \bar{x}_{ij}\right)$;

4.4. Aggregating rankings

A natural generalization of the problem presented in Section 4.1 is the problem of determining an order of items from several rankings of sets. If several rankings of sets $\tilde{\rho_1}, \ldots, \tilde{\rho_k}$ are considered as the input of the problem, we can combine those rankings in a single partial ranking $\tilde{\rho} = (\tilde{\rho_1}|\cdots|\tilde{\rho_k})$ (the partial ranking of sets induced by rankings $\tilde{\rho_1}, \ldots, \tilde{\rho_k}$). A partial ranking of sets is a linear order in which some sets are not comparable to each other; in this case, two sets of $\tilde{\rho}$ are comparable if they belong to the same input ranking. Partial orders play a very important role in the problem of finding a consensus ranking between different rankings.

Let $\mathbf{1_{s,r}}$ be the indicator function that takes the value 1 if s and r are comparable in the partial ranking of sets $\tilde{\rho}$, for all $r, s \in P$. Therefore, the (BIP)[1] model can be written for partial orders of sets as follows:

$$(\text{BIP})^3 \max \sum_i \sum_j \tilde{m}_{ij} x_{ij} - \epsilon\, d_{K\text{-}\tau}(\rho(y), \tilde{\rho}) \tag{25}$$

$$\text{s.t. } m_{rs} = \sum_{i \in r} \sum_{j \in s} x_{ij} \qquad r, s \in P : r \neq s, \mathbf{1_{s,r}} = 1 \tag{26}$$

$$x_{ij} + x_{ji} = 1 \qquad i, j \in V : i < j \tag{27}$$

$$x_{ij} + x_{jk} + x_{ki} \leq 2 \qquad i, j, k \in V : i, j, k \text{ pwd} \tag{28}$$

$$x_{ij} \in \{0, 1\} \qquad i, j \in V : i \neq j \tag{29}$$

$$m_{rs} \in \mathbb{N} \qquad r, s \in P : r \neq s, \mathbf{1_{s,r}} = 1 \tag{30}$$

$$\max \sum_r \sum_s m_{rs} y_{rs} \mathbf{1_{s,r}} \tag{31}$$

$$\text{s.t. } y_{rs} + y_{sr} = 1 \qquad r, s \in P : r < s, \mathbf{1_{s,r}} = 1 \tag{32}$$

$$y_{rs} + y_{st} + y_{tr} \leq 2 \qquad r, s, t \in P : r, s, t \text{ pwd}, \quad \mathbf{1_{s,r}} = \mathbf{1_{r,t}} = \mathbf{1_{t,s}} = 1 \tag{33}$$

$$y_{rs} \in \{0, 1\} \qquad r, s \in P : r \neq s, \mathbf{1_{s,r}} = 1 \tag{34}$$

The problem of determining an order of sets from several rankings could also be considered.

5. Results of experiments

5.1. Application to the ATP ranking

5.1.1. Linear order of items from a linear order of sets

The Association of Tennis Professionals (ATP) regularly publishes the rank of tennis players for different variables (game performance). We have chosen nine of these variables and we have saved the ranking list for each of them. The ATP dataset can be found in http://www.atpworldtour.com/. We have saved the information in Table 1. Each column shows the ranking for the corresponding variable: in column C1, we have the list of the best tennis players in *Career 1st serve points won on all surfaces from all countries*: the best in this variable is Pete Sampras, the second Roger Federer and so on. In column C2 we see that best tennis player in C2 is Rafa Nadal, the second Novak Djokovic and so on. In [17], the authors present a combinatorial approach on data envelopment analysis to compare the players, an approach that allows obtaining an efficiency score for each player based on their game performance.

We interpret the first row of Table 1 as the best set of nine tennis players (an ideal player with the best game factors from each of the best players), the second row as the second best set of nine players. Even if each row has nine tennis players, the table contains 25 tennis players. We would like to have a linear order of the 25 tennis players from this linear order of sets of size 9.

The solution for the bi-level problem (BIP^1) is given in Table 2, column Bi-level ranking. The column Mean-rank ranking is the linear order we would obtain if we compute the average position for each player. For instance, the value for Novak Djokovic is $(16 + 2 + 3 + \cdots + 2)/9 = 4.22$. We observe that both linear orders differ.

Comparing the Mean-rank ranking and the Kendall-τ ranking, the second is more robust in the sense that it is less influenced by the rest of the items on the list. We can check it with an experiment. Let consider the subtable of Table 1 which has the rows containing the best three players (the Greatest Of All Time (GOAT)), Rafael Nadal, Novak Djokovic and Roger Federer (we remove the rows which do not have any of them). This subtable is Table 3. If we compute the Mean-rank ranking for this table, then Rafael Nadal is the first. It evidences that the Mean-rank ranking is highly influenced by external items which do not belong to the best items in the list. So, our rank is more robust.

Table 4
Rank of best players.

2	Roger Federer
4	Rafael Nadal
5	Novak Djokovic
8	Pete Sampras
10	Andre Agassi
14	Andy Murray
16	Stefan Edberg
33	Jim Courier
34	Juan Martin del Potro
54	Stan Wawrinka
58	Tommy Haas
59	Richard Gasquet
73	Tomas Berdych
76	Nicolas Almagro
89	Patrick Rafter
90	Fernando Gonzalez
91	Tim Henman
97	Robin Soderling
105	James Blake
123	Todd Martin
130	Philipp Kohlschreiber
136	Daniil Medvedev
182	Radek Stepanek
201	MaliVai Washington
259	Javier Frana

Data from http://www.atpworldtour.com/.

5.2. Linear order of sets from a linear order of items

In this section we want to illustrate the inverse procedure. Table 4 shows a linear order of tennis players based on the number of matches they have won (the data have taken from http://www.atpworldtour.com/). Table 5 shows the first ten rows of Table 1. The goal is obtain a linear ordering of the variables (sets) C1, C2, C3, C4, C5, C6, C7, C8 and C9 from the linear ordering of tennis players in Table 4. This rank of variables (sets) indicates the relevance of each of the variables. Table 6 shows the solution of the bilevel problem.

5.3. Biomedical example

The BioConsert[1] project (Biological Consensus Ranking with Ties) aims to make the most of the data obtained through several ranking methods by generating a consensus ranking. Some project results are in the paper [18]. In particular, the authors propose the Fagin approach as a way of obtaining a complete ranking of items from a list of ranking of sets.

In order to illustrate our approach, we will compare our method with Fagin's method for data obtained from Bioconsert. Table 7 shows different input rankings, where GS = 'Gold Standard', BI = 'Bioggle', IE = 'In Edge', PR = 'Page Rank' and PC = 'Path Count' ([18] for more details). Each item (number) in the rankings corresponds with a gen associated with the disease ADHD (Attention Deficit Hyperactivity Disorder).

In [18], authors claim that the Fagin's approach gives the following consensus ranking:

$\pi(Fagin) := (7|2|3|1|4|5|31|41|8|6|27|43|42|17|40)$

The distance of $\pi(Fagin)$ to rankings (BI, IE, PR, PC) and (GS, BI, IE, PR, PC) computed as the Generalized Kendall Tau distance [18] is 93 and 145, respectively.

[1] https://www.bioguide-project.net/bioconsert/

Table 5
Ranking of game factor corresponding to the 25 players for whom we have available data for all these factors.

C1	C2	C3	C4	C5	C6	C7	C8	C9
Pete Sampras	Rafael Nadal	Rafael Nadal	Roger Federer	Pete Sampras	Rafael Nadal	Andre Agassi	Rafael Nadal	Rafael Nadal
Roger Federer	Novak Djokovic	Roger Federer	Pete Sampras	Roger Federer	Novak Djokovic	Rafael Nadal	Novak Djokovic	Novak Djokovic
Tomas Berdych	J.M. del Potro	Novak Djokovic	Novak Djokovic	Rafael Nadal	Andy Murray	Andy Murray	Andy Murray	Andy Murray
Robin Soderling	Andre Agassi	Andre Agassi	Rafael Nadal	Novak Djokovic	Stefan Edberg	Novak Djokovic	Andre Agassi	Andre Agassi
Patrick Rafter	Jim Courier	P. Kohlschreiber	J.M. del Potro	Andre Agassi	Roger Federer	Stefan Edberg	Stefan Edberg	Stefan Edberg
Todd Martin	Roger Federer	Stan Wawrinka	Patrick Rafter	Todd Martin	Tim Henman	Jim Courier	Robin Soderling	Roger Federer
Nicolas Almagro	Radek Stepanek	Jim Courier	Tomas Berdych	Patrick Rafter	Andre Agassi	Daniil Medvedev	Jim Courier	Jim Courier
Andy Murray	Patrick Rafter	J.M. del Potro	Jim Courier	Stan Wawrinka	M. Washington	Tomas Berdych	M. Washington	Tim Henman
J.M. del Potro	Richard Gasquet	Richard Gasquet	Andre Agassi	J.M. del Potro	Richard Gasquet	Todd Martin	J.M. del Potro	M. Washington
Tim Henman	Todd Martin	Nicolas Almagro	Robin Soderling	Stefan Edberg	James Blake	J.M. del Potro	James Blake	J.M. del Potro

C1: Career 1st Serve Points Won On All Surfaces From All Countries.
C2: Career 1st Serve On All Surfaces From All Countries.
C3: Career 2nd Serve Points Won On All Surfaces From All Countries.
C4: Career Service Games Won On All Surfaces From All Countries.
C5: Career Break Points Saved On All Surfaces From All Countries.
C6: Career 1st Serve Return Points Won On All Surfaces From All Countries.
C7: Career 2nd Serve Return Points Won On All Surfaces From All Countries.
C8: Career Break Points Converted On All Surfaces From All Countries.
C9: Career Return Games Won On All Surfaces From All Countries.

Table 6
Ranking of sets.

1	C5: Career Break Points Saved On All Surfaces From All Countries
2	C4: Career Service Games Won On All Surfaces From All Countries
3	C9: Career Return Games Won On All Surfaces From All Countries
4	C3: Career 2nd Serve Points Won On All Surfaces From All Countries
5	C6: Career 1st Serve Return Points Won On All Surfaces From All Countries
6	C2: Career 1st Serve On All Surfaces From All Countries
7	C7: Career 2nd Serve Return Points Won On All Surfaces From All Countries
8	C8: Career Break Points Converted On All Surfaces From All Countries
9	C1: Career 1st Serve Points Won On All Surfaces From All Countries

Table 7
Ranking given as input.

GS	$:= (\{1,2,3,4,5,6,7\}\|\{8,17,27,31,40,41,42,43\})$
BI	$:= (\{7\},\{3,2\}\|\{31,41,4,5,1\}\|\{8\}\|\{27,43\}\|\{42\}\|\{40\}\|\{6\}\|\{17\})$
IE	$:= (\{7\},\{31,41,4,5,1,3,2\}\|\{8\}\|\{6,17\},\{27,40,42,43\})$
PR	$:= (\{7\},\{31,41,4,5,1,2\}\|\{3\}\|\{27\}\|\{8\}\|\{42,6,43\}\|\{40\}\|\{17\})$
PC	$:= (\{7\},\{3,2\}\|\{31,41,4,5,1\}\|\{8\}\|\{6\}\|\{17,27,40,43\}\|\{42\})$

Using our bilevel Model $(BIP)^3$, the following consensus ranking is obtained:

$\pi(BIP^3) := (7|2|3|5|1|4|31|41|8|6|27|43|42|40|17)$

The distance of $\pi(BIP^3)$ to rankings (BI, IE, PR, PC) and (GS, BI, IE, PR, PC) computed as the Generalized Kendall Tau distance is 92 and 144, respectively. Thus, our consensus ranking is slightly better. Moreover, the ranking $\pi(BIP^3)$ satisfies 382 preferences of 401 preferences ($\sum_{i,j} m_{ij} = 401$) while ranking $\pi(Fagin)$ satisfies 381. Finally, ranking $\pi(BIP^3)$ provides a ranking of sets at a distance of 10 from the initial rankings (GS, BI, IE, PR, PC), while $\pi(Fagin)$ is at a distance of 11.

6. Conclusions

In this paper, we have introduced a new methodology for obtaining a ranking of items/sets from a ranking of sets/inputs. Since the new methodology applied to the resulting ranking of items/sets, deals with another ranking of sets/items, we propose an optimization bilievel model that provides ranking of items/sets that, when solving the inverse problem, gives a ranking close to the input ranking.

Since problems of ranking items and sets are combinatorial, the combination of both problems leads to a combinatorial bilevel optimization model, with 0–1 variables in first and second level problems. The bilevel optimization problem is formulated as a single level problem with a huge number of constraints. We propose a cutting-plane approach in an attempt to resolve this issue.

Different examples throughout the work illustrate the advantages of the new models. In addition, two experiments show the applicability of the proposed models. The experiments consider both partial and total rankings.

In general, with the procedures in this paper each time we have a double entry table and a linear order of its columns (rows) we can obtain a linear order of its rows (columns). Depending on whether the linear order is for items or for sets, we will select one procedure or the other.

It remains for future research to implement techniques that allow dealing with very large instances and the extension of these models to more general situations.

CRediT authorship contribution statement

Martine Labbé: Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Mercedes Landete:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Juan F. Monge:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Software, Validation, Investigation, Visualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Juan Francisco Monge reports financial support was provided by Ministerio de Ciencia e Innovación. Spain. Juan Francisco Monge and Mercedes Landete reports financial support was provided by Government of the Valencian Community, Spain.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by grants PID2019-105952GB-I00 and PID2021-122344NB-I00 funded by MCIN/AEI/10.13039/501100011-033 and "ERDF A way of making Europe", and PROMETEO/2021/063 funded by the government of the Valencian Community, Spain.

References

[1] Charon I, Hudry O. A survey on the linear ordering problem for weighted or unweighted tournament. 4OR 2007;5:5–60.

[2] Charon I, Hudry O. An updated survey on the linear ordering problem for weighted or unweighted tournament. Ann Oper Res 2010;175:107–58.

[3] Kendall M. A new measure of rank correlation. Biometrika 1938;30:81–9.

[4] Kemeny J. Mathematics without numbers. Daedalus 1959;88:577–91.

[5] Aledo JA, Gómez JA, Molina D. Using extension sets to aggregate partial rankings in a flexible setting. Appl Math Comput 2016;290:208–23.

[6] Martí R, Reinelt G. The lineal ordering problem. In: Exact and heuristic methods in combinational optimization. Springer; 2011.

[7] García-Nové EM, Alcaraz J, Landete M, Puerto J, Monge JF. Rank aggregation in cyclic sequences. Optim Lett 2017;11:667–78.

[8] Barberà S, Bossert W, Pattanaik PK. Ranking sets of objects. In: Barberà S, Hammond PJ, Seidl C, editors. Hand book of utility theory. Springer; 2004.

[9] Darmann A, Klamler C. Using the borda rule for ranking sets of objects. Soc Choice Welf 2019;53:399–414.

[10] Aparicio J, Landete M, J.F. Monge. A linear ordering problem of sets. Ann Oper Res 2020;288(2020):45–64.

[11] Alcaraz J, García-Nové EM, Landete M, Monge JF. The linear ordering problem with clusters: a new partial ranking. TOP 2020;28:646–71.

[12] Alcaraz J, Landete M, Monge JF. Rank aggregation: Models and algorithm. In: Salhi, Boylan, editors. The palgrave handbook of operations research. 2022.

[13] Fagin R, Kumar R, Mahdian M, Sivakumar D. Comparing and aggregating rankings with ties. In: Proceedings of PODS. 2004, p. 47–55.

[14] Brancotte B, Yang B, Blin G, Cohen-Boulakia S, Denise A, Hamel S. Rank aggregation with ties: Experiments and analysis. Proc VLDB Endow 2015;8(11):1202–13.

[15] Monge JF. The Concordance coefficient: An alternative to the Kruskal–Wallis test. 2019, arXiv:1912.12880v2.

[16] Gaar E, Lee J, Ljubić I, Sinnl M, Taninmiş K. SOCP-based disjunctive cuts for a class of integer nonlinear bilevel programs. In: Aardal K, Sanità L, editors. Integer programming and combinatorial optimization. IPCO 2022, Lecture notes in Computer Science, vol. 13265, Springer; 2022.

[17] Landete M, Monge JF, Ruiz JL. Robust DEA efficiency scores: A probabilistic/combinatorial approach. Expert Syst Appl 2017;86:145–54.

[18] Cohen-Boulakia S, Denise A, Hamel S. Using medians to generate consensus rankings for biological data. In: Scientific and statistical database management. vol. 6809 of LNCS, Springer; 2001, p. 73–90.