Some results about rank reversal instances in the PROMETHEE methods

Céline Verly and Yves De Smet*

CoDE-SMG,
Engineering Faculty,
Université libre de Bruxelles,
Boulevard du Triomphe CP 210-01, 1050 Bruxelles, Belgium
E-mail: ceverly@ulb.ac.be
E-mail: yves.de.smet@ulb.ac.be
*Corresponding author

Abstract: The multicriteria methods based on pairwise comparisons suffer from possible rank reversal occurrences when the set of alternatives is modified. We study this distinctive feature in the scope of the PROMETHEE I and II methods. First, empirical tests are conducted on the basis of artificial datasets in order to quantify the likelihood of rank reversal instances. Then conditions to avoid this phenomenon are provided. Finally, a comparison with a procedure based on a distillation process is performed.

Keywords: multicriteria decision aid; rank reversal; PROMETHEE methods.

Reference to this paper should be made as follows: Verly, C. and De Smet, Y. (2013) 'Some results about rank reversal instances in the PROMETHEE methods', *Int. J. Multicriteria Decision Making*, Vol. 3, No. 4, pp.325–345.

Biographical notes: Céline Verly is a Teaching Assistant at the Engineering Faculty of the Université libre de Bruxelles. She is both a member of the Computer and Decision Engineering Laboratory and of the SMG Unit. She holds a degree in Mathematics in 2007 and a Master in Actuarial Sciences in 2009. Her research interests are focused on the rank reversal problem in multicriteria decision aid methods.

Yves De Smet is an Assistant Professor at the Engineering Faculty of the Université libre de Bruxelles. He is both the head of the Computer and Decision Engineering Laboratory and of the SMG Unit. He holds a degree in Mathematics in 1998 and a PhD in Applied Sciences in 2005. His research interests are focused on multicriteria decision aid and multi-objective optimisation. He has published 40 papers in international journals and conference proceedings. Besides his academic activities he has been involved in different industrial projects. Since 2010, he has been a co-founder of the decision sights spin-off.

This paper is a revised and expanded version of a paper entitled 'Some considerations about rank reversal occurrences in the PROMETHEE II method' presented at 71st Meeting of the European Working Group 'Multiple Criteria Decision Aiding (MCDA)', Torino, Italy, 25–27 March 2010.

1 Introduction

Multicriteria decision aid methods based on pairwise comparisons may suffer from the well-known *rank reversal* problem. Different authors have already stressed this distinctive feature. For instance, we refer the interested reader to Belton and Gear (1983), Saaty and Vargas (1984), Saaty (1987), Harker and Vargas (1987), Barzilai and Golany (1994), Triantaphyllou (2001), Finan and Hurley (2002), Liberatore and Nydick (2004), Wang and Elhag (2006) and Wijnmalen and Wedley (2009a, 2009b) for discussions related to the analytic hierarchy process (AHP), to Wang and Triantaphyllou (2008) for those related to ELECTRE methods and to De Keyser and Peeters (1996) and Mareschal et al. (2008) for those related to PROMETHEE methods.

First of all, we have to emphasise that the *rank reversal* problem is not clearly defined. Indeed, different interpretations can be found in the literature. For example, some authors (Finan and Hurley, 2002; Liberatore and Nydick, 2004; Wijnmalen and Wedley, 2009b) take an interest in rank reversal occurrences in AHP when a non-discriminating criterion is removed from the problem. Triantaphyllou (2001) studies this issue while considering transitivity properties. Wang and Triantaphyllou (2008) address similar questions for ELETRE methods. Additionally, they study rank reversal instances when a alternative a_k is replaced by a dominated one.

Belton and Gear (1983) are among the first to have highlighted the problem we are interesting in, e.g., the rank reversal phenomenon when a given alternative is deleted or added. The authors focus on rank reversal effects occurring when a copy of an alternative is added. These investigations have led to an important debate about the legitimacy of rank reversal. See, for instance, Wang and Elhag (2006), Harker and Vargas (1987) and Saaty (1987).

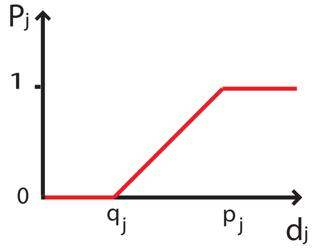
In this paper, we address this problem in the context of the PROMETHEE methods. Section 2 is devoted to a short reminder of the basic notions underlying the PROMETHEE I and II rankings. Then, we study the problem of rank reversal occurring in the PROMETHEE II ranking in Section 3. Firstly, we address the issue of deleting a non-discriminating criterion. Secondly, we focus on the deletion or the addition of a copy of an alternative. Thirdly, like in Wang and Elhag (2006), we study rank reversal instances caused by the addition or the deletion of a randomly chosen alternative. These results constitute an extension of the works initiated in Mareschal et al. (2008). Section 4 is devoted to the study of rank reversal occurrences in the PROMETHEE I partial ranking. Finally, we compare the latter results with partial rankings obtained by a procedure based on a distillation process.

2 PROMETHEE methods

This section is devoted to a short overview of the PROMETHEE I and II ranking methods. A detailed description of these approaches can be found in Brans and Mareschal (2005).

Let $A = \{a_1, a_2, .., a_n\}$ be a set of n alternatives and $F = \{f_1, f_2, .., f_m\}$ be a set of m criteria. We assume that the criteria are real-valued and have to be maximised. For every couple of alternatives (a_i, a_j) and for every criterion f_k, we first compute the difference $d_k(a_i, a_j) = f_k(a_i) - f_k(a_j)$. Then, this difference is transformed into a preference degree denoted $P_k\big(d_k(a_i, a_j)\big) \in [0, 1]$. P_k is a non-decreasing function and is such that $\forall x \leq 0$ $P_k(x) = 0$. The latter can be linear, as illustrated in Figure 1. We refer the interested reader to Brans and Mareschal (2005) for other classical types of preference functions.

Figure 1 The linear preference function (see online version for colours)



The third step consists to compute a global preference degree as follows:

$$\pi(a_i, a_j) = \sum_{k=1}^{m} w_k \cdot P_k\big(d_k(a_i, a_j)\big) \tag{1}$$

where w_k represents the weight associated to criterion f_k. We assume that $w_k > 0$, $\forall k = 1, \ldots, m$, and $\sum_{k=1}^{m} w_k = 1$.

Given these global preference degrees, we are able to compute positive ($\phi^+(a_i)$), negative ($\phi^-(a_i)$) and net flow scores ($\phi(a)$) in the following way:

$$\phi^+(a_i) = \frac{1}{n-1} \sum_{a_j \in A} \pi(a_i, a_j) \tag{2}$$

$$\phi^-(a_i) = \frac{1}{n-1} \sum_{a_j \in A} \pi(a_j, a_i) \tag{3}$$

$$\phi(a_i) = \phi^+(a_i) - \phi^-(a_i). \tag{4}$$

Given these three scores, two rankings can be established. The PROMETHEE I partial ranking is based on both the positive and the negative flow scores. An alternative a_i is ranked before an alternative a_j if $\phi^+(a_i) \geq \phi^+(a_j)$ and $\phi^-(a_i) < \phi^-(a_j)$ or if $\phi^+(a_i) > \phi^+(a_j)$ and $\phi^-(a_i) \leq \phi^-(a_j)$. The alternative a_i has the same rank as a_j if $\phi^+(a_i) = \phi^+(a_j)$ and $\phi^-(a_i) = \phi^-(a_j)$. Otherwise, the alternative a_i is considered to be incomparable to a_j ($a_i J a_j$).

The PROMETHEE II ranking is based on the net flow score. Naturally, we have that a_i is ranked before a_j if $\phi(a_i) > \phi(a_j)$ and that a_i has the same rank as a_j if $\phi(a_i) = \phi(a_j)$.

Finally, it is obvious to note that:

$$\phi(a_i) = \sum_{k=1}^{m} w_k \cdot \phi_k(a_i). \tag{5}$$

where $\phi_k(a_i) = \frac{1}{n-1} \sum_{a_j \in A} \Big(P_k\big(d_k(a_i, a_j)\big) - P_k\big(d_k(a_j, a_i)\big)\Big)$.

As stressed by De Keyser and Peeters (1996), the PROMETHEE I ranking is subject to the rank reversal problem. Indeed, the authors consider the multicriteria problem represented in Table 1. Computing the PROMETHEE I partial ranking of this problem, will lead to cPb, cPa and bJa (see Table 2). If alternative b is deleted, a rank reversal occurs between a and c since, now, we have aPc (see Table 3). At this level, it is worth noting that this phenomenon is only due to small flow score differences.

Table 1 Example of a rank reversal occurrence in PROMETHEE I

	g_1	g_2	g_3
Min/Max	Max	Max	Max
Type	$2(q = 10)$	$2(q = 100)$	$3(p = 50)$
Weight	1/3	1/3	1/3
a	90	800	401
b	88	905	399
c	99	855	400

Notes: Type 2 refers to preference functions where $q = p$. Type 3 refers to preference functions such that $q = 0$.

Source: De Keyser and Peeters (1996)

Table 2 Positive and negative flow scores computed for the example introduced

	ϕ^+	ϕ^-
a	0.010	0.167
b	0.167	0.177
c	0.170	0.003

Source: De Keyser and Peeters (1996)

Table 3 Positive and negative flows scores computed for the example introduced

	ϕ^+	ϕ^-
a	0.007	0.000
c	0.000	0.007

Source: De Keyser and Peeters (1996)

Following these observations, Mareschal et al. (2008) have investigated the conditions under which rank reversal will not occur in the PROMETHEE II method. Their main result can be summarised as follows: the deletion of a given alternative will never lead to a rank reversal between two alternatives a_i and a_j if $|\phi(a_i) - \phi(a_j)| > \frac{2}{n-1}$. In other words, rank reversal instances only happen when pairs of alternatives have close net flow scores. In the next section, we continue to further investigate rank reversal occurrences in the PROMETHEE II ranking method.

3 Rank reversal in the PROMETHEE II ranking

3.1 *Deletion of a non-discriminating criterion*

Following the idea of Finan and Hurley (2002), we first address the problem of the deletion of a non-discriminating criterion. In such a context, the decision maker is supposed to be indifferent among all the alternatives for that specific criterion. Therefore, it is useless in the decision making process and its deletion should not impact the final ranking. Let us note that in practice such a case would not happen and decision makers would simply not consider this criteria. However, like in Finan and Hurley (2002), we take an interest in this specific phenomenon from a purely methodological point of view. The following result confirms our expectations.

Proposition 3.1: Let us assume that $\exists k \in \{1, .., m\} | f_k(a_i) = f_k(a_j)\ \forall a_i, a_j \in A$. Let ϕ and ϕ' be respectively the PROMETHEE II net flow scores computed on the basis of (A, F) and $(A, F \setminus \{f_k\})$. We have that ϕ and ϕ' induce the same ranking.

Proof: It is obvious that $\phi_k(a_i) = 0\ \forall a_i \in A$. As a consequence,

$$\phi(a_i) = \sum_{l=1}^{m} w_l \cdot \phi_l(a_i) \tag{6}$$

$$= \sum_{l=1, l\neq k}^{m} w_l \cdot \phi_l(a_i) \tag{7}$$

$$= W^k \cdot \sum_{l=1, l\neq k}^{m} \frac{w_l}{W^k} \cdot \phi_l(a_i) \tag{8}$$

$$= W^k \cdot \sum_{l=1, l\neq k}^{m} w'_l \cdot \phi_l(a_i) \tag{9}$$

$$= W^k \cdot \phi'(a_i) \tag{10}$$

where $W^k = \sum_{l=1, l\neq k}^{m} w_l$ and $w'_l = \frac{w_l}{W^k}$ is the normalised weight of criterion l after the deletion of criterion k. Given this result, we have that $\phi(a_i) > \phi(a_j) \Leftrightarrow \phi'(a_i) > \phi'(a_j)$ and $\phi(a_i) = \phi(a_j) \Leftrightarrow \phi'(a_i) = \phi'(a_j)$. □

3.2 Deletion or addition of a copy of an existing alternative

As already stressed, the rank reversal phenomenon can be studied with respect to the addition or the deletion of a copy of an existing alternative (see for instance Belton and Gear (1983)).

The following example shows that a similar effect may happen when using the PROMETHEE II method.

Example 3.1: If we apply the PROMETHEE II method to the multicriteria problem represented in Table 4, we obtain $a \succ b \succ c$ (see Table 5).

Now, if we add an alternative d, which is an exact copy of c, we observe a rank reversal between the alternatives a and b (see Table 6). One more time, let us stress that this effect is only due to small net flow score differences. Moreover, this result is compatible with the bound developed by Mareschal et al. (2008).

This example clearly illustrates the fact that the PROMETHEE II ranking may suffer from rank reversal instances when a copy of an alternative is added or deleted. A natural question is to investigate potential rank reversal occurrences when any given alternative is deleted from A. This will be the scope of the next section.

Table 4 Example of a multicriteria problem leading to a rank reversal instance when a copy of an alternative is added

	g_1	g_2
Min/Max	Max	Max
p	0.75	0.75
q	0.25	0.25
Weight	0.5	0.5
a	0.5	0.86
b	0.22	0.9
c	0.4	0.2

Table 5 Net flow scores for the problem presented in Table 4

$\phi(a)$	0.22
$\phi(b)$	0.21
$\phi(c)$	–0.43

Table 6 Net flow scores for the problem presented in Table 4 when a copy of c, denoted d, is added

$\phi(a)$	0.283
$\phi(b)$	0.290
$\phi(c)$	–0.287
$\phi(d)$	–0.287

3.3 Deletion of an alternative

Mareschal et al. (2008) are the first who have studied this problem in the context of the PROMETHEE II method. A main result of their contribution is to show that this specific kind of rank reversal between two alternatives a_i and a_j will never happen if $|\phi(a_i) - \phi(a_j)| > \frac{2}{n-1}$. In other words, rank reversal may only happen between two alternatives if the net flow score difference is relatively small (as illustrated in Example 3.1).

In this section, we extend this result. A first question is to study the number of times this bound is reached when we consider the two first alternatives of the PROMETHEE II ranking (denoted respectively a_1 and a_2).

Thousand matrices have been artificially generated using MATLAB. Evaluations are drawn using a uniform probability distribution on $[0, 1]$. More formally, we assume $f_k(a_i) \sim U_{[0,1]}$. Weights are also drawn from a uniform distribution on $[0, 1]$ and are then normalised. Finally, we will only use linear preference functions. We consider $p_k \sim U_{[0,1]}$ and $q_k = \alpha_k \cdot p_k$ where $\alpha_k \sim U_{[0,1]}$. As a consequence, we always have $q_k \leq p_k$.

Table 7 lists the number of times that the bound is greater than $\frac{2}{n-1}$ among $1,000$ iterations when the number of alternatives n varies from 5 to 100 and the number of criteria m varies from 2 to 8. For instance, if we consider ten alternatives and four criteria, the net flow score difference between a_1 and a_2 is lower than the bound $2/9$ in 857 cases. In other words, in only 143 simulations we were sure that rank reversal will not happen according to the results of Mareschal et al. (2008).

The second question is to study the number of rank reversals that effectively happened between the first and the second alternatives when we delete a randomly chosen alternative. Table 8 indicates these frequencies when we consider the same thousand matrices as for the construction of Table 7. For instance, only 37 rank reversals happened for the case of ten alternatives and four criteria (while 857 rank reversals could have eventually happened).

In order to deepen our analysis, we illustrate in Figure 2 the net flow score difference before the withdrawal of an alternative ($\phi(a_1) - \phi(a_2)$) and after its deletion ($\tilde{\phi}(a_1) - \tilde{\phi}(a_2)$) when we consider ten alternatives and four criteria. The red points represent the cases of rank reversal occurrences between a_1 and a_2. The vertical line represents the boundary $2/9$. Clearly, we observe that rank reversal occur when net flow score differences are much lower than the considered bound. Figure 3 offers a complementary information; it represents an histogram of observed net flow score differences for 100 cases of rank reversal between a_1 and a_2.

These observations have led us to refine the bound proposed by Mareschal et al. (2008). To keep notations simple, we will denote by a and b any pair of alternatives from A.

Table 7 Number of times that the net flow score difference between the two first alternatives is greater than $\frac{2}{n-1}$

n\m	*2*	*3*	*4*	*5*	*6*	*7*	*8*
5	128	77	53	38	18	18	20
10	188	152	143	136	107	92	77
15	233	232	228	201	186	176	155
20	237	256	275	293	260	248	230
25	289	335	332	347	294	280	288
30	322	359	350	368	387	354	347
35	329	377	398	393	409	406	378
40	371	409	420	455	435	444	399
45	364	453	472	448	484	467	437
50	383	466	518	506	476	499	464
100	503	582	654	680	668	650	640

Table 8 Number of rank reversal occurrences between the two first alternatives that are really observed among 1,000 simulations (when a random alternative is deleted)

n\m	*2*	*3*	*4*	*5*	*6*	*7*	*8*
5	35	60	71	50	58	65	76
10	31	39	37	48	47	42	47
15	20	25	35	36	37	33	32
20	17	20	27	30	26	37	28
25	18	20	24	20	23	34	34
30	19	22	23	18	28	19	18
35	14	21	15	16	22	20	17
40	12	24	17	11	24	14	15
45	15	26	12	16	15	16	15
50	7	16	19	20	12	16	23
100	2	7	6	14	5	8	16

Figure 2 $\tilde{\phi}(a_1) - \tilde{\phi}(a_2)$ in function of $\phi(a_1) - \phi(a_2)$ (see online version for colours)

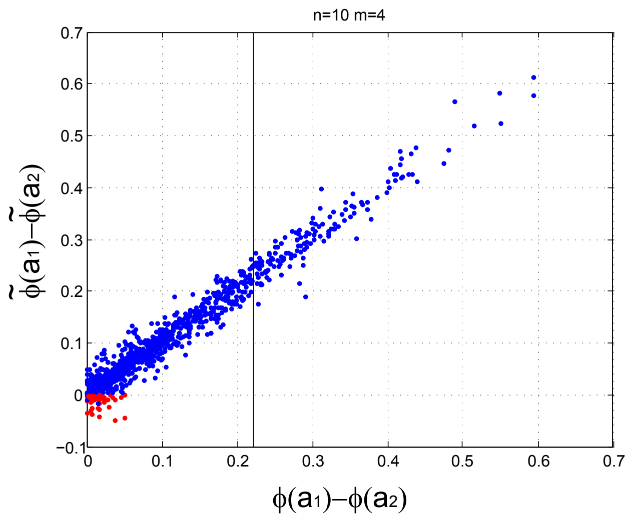
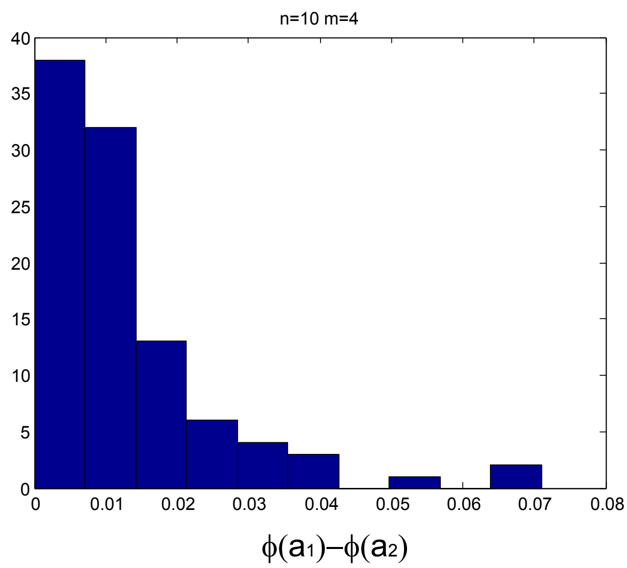


Figure 3 Histogram of $\phi(a_1) - \phi(a_2)$ values for 100 problems where rank reversal between the first and the second alternatives was observed (see online version for colours)



Proposition 3.2: Let us assume that $\phi(a) > \phi(b)$. Rank reversal between two alternatives a and b may only occur if:

$$\phi(a) - \phi(b) < \frac{2}{n-1} \cdot W_{ab}$$

where $W_{ab} = \sum_{k=1, g_k(a) \geq g_k(b)}^{m} w_k$.

Intuitively, W_{ab} represents the weight importance of the coalition of criteria for which a is at least as good as b.

Proof: Let y be the alternative that is randomly chosen to be deleted ($y \neq a$ and $y \neq b$). It is obvious that we have:

$$\tilde{\phi}(a) = \frac{1}{n-2}((n-1)\phi(a) - \pi(a,y) + \pi(y,a)). \tag{11}$$

Thus, we obtain:

$$\tilde{\phi}(a) - \tilde{\phi}(b) = \tfrac{1}{n-2}\big((n-1)(\phi(a) - \phi(b)) - \pi(a,y) + \pi(y,a) + \pi(b,y) - \pi(y,b)\big). \tag{12}$$

If there is no rank reversal, we should have $\tilde{\phi}(a) - \tilde{\phi}(b) > 0$. Then, by using the definition of the global preference degree (1), we would like to have:

$$\phi(a) - \phi(b) + \tfrac{1}{n-1}\textstyle\sum_{k=1}^{m} w_k \cdot (-\pi_k(a,y) + \pi_k(y,a) + \pi_k(b,y) - \pi_k(y,b)) > 0 \tag{13}$$

In Table 9, we discuss the values of $-\pi_k(a,y) + \pi_k(y,a) + \pi_k(b,y) - \pi_k(y,b) = X$ (for any $k \in \{1, 2, \ldots, m\}$) with respect to the different possible unicriterion rankings of the three alternatives a, b and y.

Table 9 Discussion about the values of $-\pi_k(a,y) + \pi_k(y,a) + \pi_k(b,y) - \pi_k(y,b)$ in function of the different permutations of the alternatives

	$\pi_k(a,y)$	$\pi_k(y,a)$	$\pi_k(b,y)$	$\pi_k(y,b)$	X
aby	≤ 1	$= 0$	≥ 0	$= 0$	≥ -1
ayb	≤ 1	$= 0$	$= 0$	≤ 1	≥ -2
bay	$\leq \pi_k(b,y)$	$= 0$	≥ 0	$= 0$	≥ 0
bya	$= 0$	≥ 0	≥ 0	$= 0$	≥ 0
yab	$= 0$	≥ 0	$= 0$	≤ 1	≥ -1
yba	$= 0$	≥ 0	$= 0$	$\leq \pi_k(y,a)$	≥ 0

These values lead us to conclude that:

$$\frac{1}{n-1}\sum_{k=1}^{m} w_k(-\pi_k(a,y) + \pi_k(y,a) + \pi_k(b,y) - \pi_k(y,b) \geq -\frac{2}{n-1}W_{ab}. \tag{14}$$

Consequently, no rank reversal will happen if

$$\phi(a) - \phi(b) > \frac{2}{n-1}W_{ab}.$$

□

3.4 Tests based on real data

In order to complete the tests presented in the previous section, we focus here on the data used to compute the famous Shanghai ranking in 2009 (these are freely available on http://www.arwu.org/ARWU2009.jsp). Our aim is to investigate if conclusions drawn from artificial datasets hold when real data is used. Of course, we do not discuss the validity of the Shanghai ranking itself which is still subject to a number of debates (see for example, Liu and Cheng, 2005; Van Raan, 2005; Kivinen and Hedman, 2008; Buela-Casal et al., 2007). In what follows, we will only consider the hundred first universities of the ranking.

Six criteria are taken into account to build this ranking. The following description comes from http://www.arwu.org/ARWU2009.jsp.

- Alumni: the total number of the alumni of an institution winning nobel prizes and fields medals. Alumni are defined as those who obtain bachelor, master's or doctoral degrees from the institution (weight: 10%).
- Award: the total number of the staff of an institution winning nobel prizes in physics, chemistry, medicine, economics and fields medals in mathematics. Staff is defined as those who work at an institution at the time of they win the prize (weight: 20%).
- Hici: the number of highly cited researchers in 21 subject categories. These individuals are the most highly cited within each category (weight: 20%).
- N&S: the number of papers published in nature and science between 2004 and 2008 (20%).
- PUB: total number of papers indexed in science citation index-expanded and social science citation index in 2008. Only publications of 'article' and 'proceedings paper' types are considered (20%).
- PCP: the weighted scores of the above five indicators divided by the number of full-time equivalent academic staff (10%).

The PROMETHEE II method is applied to these data using the same weights as in the initial ranking and linear preference functions. The preference and the indifference thresholds of each criterion k ($k = 1, \ldots, 6$) are respectively the first and the third quartiles of the difference sets $\{|g_k(a_i) - g_k(a_j)| \forall a_i, a_j \in A | i > j\}$

Tests have been conducted according to the following procedure:

1. The PROMETHEE II ranking is computed.
2. Universities are ranked according to the net flow score. We obtain the vector: $(a_1, a_2, \ldots, a_{100})$.
3. For $i = 1 : 100$.
   a. alternative i ($i = 1, 2 \ldots, 100$) is deleted. The vector $(a'_1, a'_2, \ldots, a'_{99})$ is computed such that:

$$a'_j = \begin{cases} a_j \text{ if } j \in \{1, 2, \ldots, i-1\} \\ a_{j+1} \text{ if } j \in \{i+1, i+2, \ldots, 100\} \end{cases} .$$

b the PROMETHEE II ranking of the set $\{a'_1, a'_2, \ldots, a'_{99}\}$ is computed.

Figure 4 represents how the rank of an alternative a'_l may vary when another alternative is deleted. It clearly shows that the rank reversal problem has a limited effect on the Shanghai ranking when the PROMETHEE II method is used. Indeed, we see for example that no rank reversal happens between the nine first alternatives. Furthermore, we may conclude that rank reversals do not exceed two places excepted for the worst case illustrated in Figure 5 (the alternative a'_{83} can take the place 86 and the alternative a'_{86} can take the place 83).

In Figure 6, we compute the histogram of net flow score differences when rank reversals happened. We can observe that these values are lower than 0.0025 (while the theoretical boundary $\frac{2}{n-1}$ is equal to 0.0202).

4 Rank reversal in PROMETHEE I partial ranking

As shown by De Keyser and Peeters (1996), rank reversal instances may also happen in the PROMETHEE I ranking. Nevertheless, it is worth noting that the problem is a bit more complex in a partial ranking than in a complete one. Indeed, a strict preference between two alternatives can be reversed or being transformed into an incomparability or an indifference relation (or vice versa). In what follows all these situations will be referred to as rank reversals.

Figure 4 Shanghai ranking: variation of the rank of every alternative when another alternative is deleted (see online version for colours)

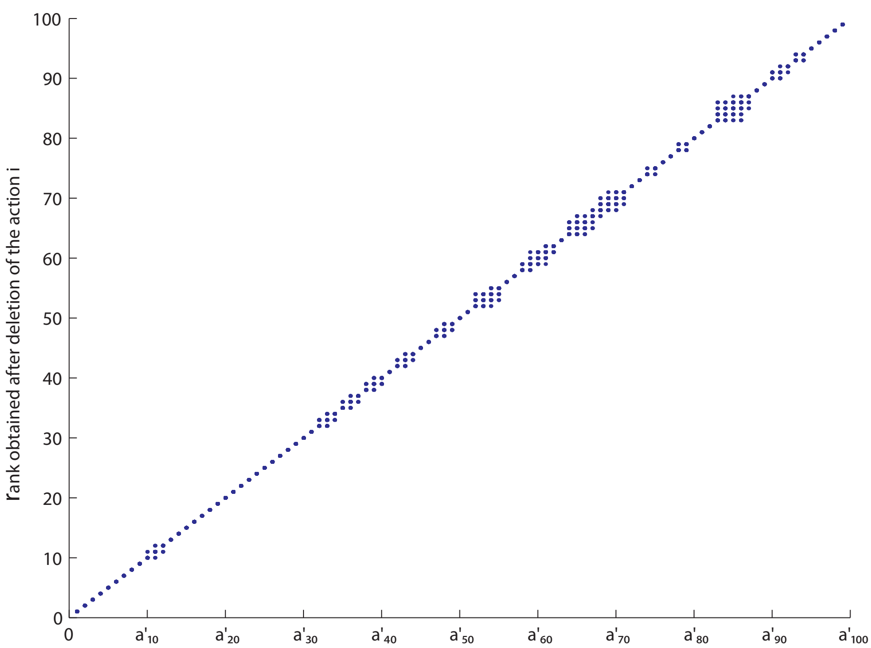


Figure 5 Zoom of Figure 4 on the alternatives $a'_{80}, \ldots, a'_{90}$ (see online version for colours)

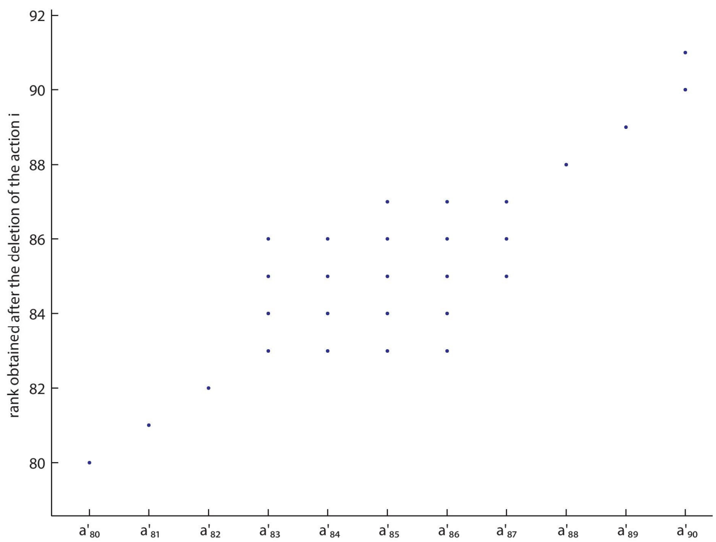
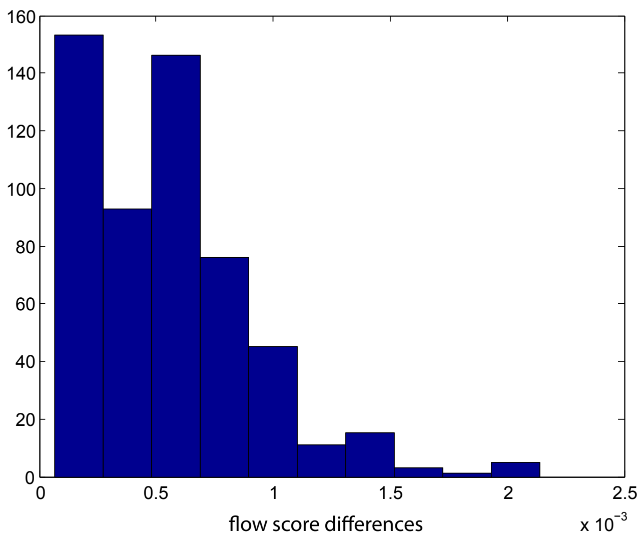


Figure 6 Histogram of the net flow score differences when rank reversal occurs (see online version for colours)



Note: A total number of 548 rank reversal occurrences were observed.

Firstly, let us note that the deletion of a non-discriminating criterion will not lead to rank reversal in the PROMETHEE I ranking. The proof is similar to the demonstration of Proposition 3.1.

Secondly, our principal aim was to highlight a bound (like the one available for PROMETHEE II) that allows us to identify situations for which no rank reversal might occur in the PROMETHEE I ranking. This bound is a direct corollary of Proposition 4.1.

Let $\tilde{\phi}^+(a)$ (respectively $\tilde{\phi}^-(a)$) be the positive (respectively the negative) flow score after the deletion of an alternative denoted y ($y \neq a$). It is obvious that:

$$\phi^+(a) = \frac{n-2}{n-1}\tilde{\phi}^+(a) + \frac{1}{n-1}\pi(a,y) \tag{15}$$

$$\phi^-(a) = \frac{n-2}{n-1}\tilde{\phi}^-(a) + \frac{1}{n-1}\pi(y,a) \tag{16}$$

Proposition 4.1: $\forall\ a,b \in A$:

$$|\big(\phi^+(a) - \phi^+(b)\big) - \big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big)| \leq \frac{1}{n-1}$$

$$|\big(\phi^-(a) - \phi^-(b)\big) - \big(\tilde{\phi}^-(a) - \tilde{\phi}^-(b)\big)| \leq \frac{1}{n-1}.$$

Proof: First, let us remark that we will only prove the first inequality (since the demonstration of the second one is similar). We distinguish three cases: $\tilde{\phi}^+(a) - \tilde{\phi}^+(b) = 0$, $\tilde{\phi}^+(a) - \tilde{\phi}^+(b) > 0$ or $\tilde{\phi}^+(a) - \tilde{\phi}^+(b) < 0$.

The first case is obvious. In what follows we will focus on the second case: $\tilde{\phi}^+(a) - \tilde{\phi}^+(b) > 0$ (the demonstration of the third case is similar).

As a direct consequence of relation 15, we have:

$$\phi^+(a) - \phi^+(b) = \frac{n-2}{n-1}\big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big) + \frac{1}{n-1}\big(\pi(a,y) - \pi(b,y)\big) \tag{17}$$

Since $\big(\pi(a,y) - \pi(b,y)\big) \leq 1$, we have:

$$\big(\phi^+(a) - \phi^+(b)\big) - \frac{n-2}{n-1}\big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big) \leq \frac{1}{n-1} \tag{18}$$

Due to our initial assumption, we have:

$$\big(\phi^+(a) - \phi^+(b)\big) - \big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big) \leq \frac{1}{n-1} \tag{19}$$

Secondly, let us show that:

$$\big(\phi^+(a) - \phi^+(b)\big) - \big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big) \geq -\frac{1}{n-1}. \tag{20}$$

We can write:

$$\begin{aligned}
&\big(\phi^+(a) - \phi^+(b)\big) - \big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big) \\
&= \big(\phi^+(a) - \phi^+(b)\big) - \frac{n-2}{n-1}\big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big) \\
&\quad - \frac{1}{n-1}\big(\tilde{\phi}^+(a) - \tilde{\phi}^+(b)\big).
\end{aligned} \tag{21}$$

By using (17), we find

$$\begin{aligned}&\big(\phi^+(a)-\phi^+(b)\big)-\big(\tilde{\phi}^+(a)-\tilde{\phi}^+(b)\big)\\&=\frac{1}{n-1}\big(\pi(a,y)-\pi(b,y)\big)-\frac{1}{n-1}\big(\tilde{\phi}^+(a)-\tilde{\phi}^+(b)\big).\end{aligned}\tag{22}$$

Consequently, we still have to show:

$$\big(\tilde{\phi}^+(a)-\tilde{\phi}^+(b)\big)-\big(\pi(a,y)-\pi(b,y)\big)\leq 1.\tag{23}$$

We have:

$$\begin{aligned}&\big(\tilde{\phi}^+(a)-\tilde{\phi}^+(b)\big)-\big(\pi(a,y)-\pi(b,y)\big)\\&=\frac{1}{n-2}\sum_{x\in A,x\neq y}\big(\pi(a,x)-\pi(b,x)\big)\qquad(24)\\&\quad-\big(\pi(a,y)-\pi(b,y)\big)\\&=\sum_{k=1}^{m}w_k\cdot\Big(\frac{1}{n-2}\sum_{x\in A,x\neq y}\big(\pi_k(a,x)-\pi_k(b,x)\big)\qquad(25)\\&\quad-\big(\pi_k(a,y)+\pi_k(b,y)\big)\Big).\end{aligned}$$

By contradiction, let us suppose that:

$$\sum_{k=1}^{m}w_k\Big(\frac{1}{n-2}\sum_{x\in A,x\neq y}\big(\pi_k(a,x)-\pi_k(b,x)\big)-\big(\pi_k(a,y)-\pi_k(b,y)\big)\Big)>1.\tag{26}$$

Since $\forall k\in\{1,2,\dots,m\}$ $w_k>0$ and $\sum_{k=1}^{m}w_k=1$, there exists at least one $l\in\{1,2,\dots,m\}$ such that

$$\frac{1}{n-2}\sum_{x\in A,x\neq y}\big(\pi_l(a,x)-\pi_l(b,x)\big)-\big(\pi_l(a,y))-\pi_l(b,y)\big)>1.\tag{27}$$

In fact, $\frac{1}{n-2}\sum_{x\in A,x\neq y}\big(\pi_l(a,x)-\pi_l(b,x)\big)=\tilde{\phi}_l^+(a)-\tilde{\phi}_l^+(b)$. We know that $\tilde{\phi}_l^+(a)-\tilde{\phi}_l^+(b)\in[-1,1]$. Moreover, $\pi_l(a,y)-\pi_l(b,y)\in[-1,1]$. If (27) is true, we should to have $\tilde{\phi}_l^+(a)-\tilde{\phi}_l^+(b)>0$ and $\pi_l(a,y)-\pi_l(b,y)<0$.

Firstly, $\pi_l(a,y)-\pi_l(b,y)<0$ implies that $g_l(b)>g_l(a)$. We split all the other alternatives in a partition of three sets:

- $B=\{c\in A\setminus\{y\}|g_l(c)\geq g_l(b)\}$
- $C=\{c\in A\setminus\{y\}|g_l(b)>g_l(c)\geq g_l(a)\}$
- $D=\{c\in A\setminus\{y\}|g_l(a)>g_l(c)\}$.

We can write:

$$\tilde{\phi}_l^+(a) - \tilde{\phi}_l^+(b) = \frac{1}{n-2}\Big(\underbrace{\sum_{c \in B} \big(\pi_l(a,c) - \pi_l(b,c)\big)}_{\alpha} + \underbrace{\sum_{c \in C} \big(\pi_l(a,c) - \pi_l(b,c)\big)}_{\beta} + \underbrace{\sum_{c \in D} \big(\pi_l(a,c) - \pi_l(b,c)\big)}_{\gamma} \Big). \tag{28}$$

As a consequence, we have:

- $\alpha = 0$,
- $\beta = -\sum_{c \in C} \pi_l(b,c) \leq 0$,
- $\gamma = \sum_{c \in D} \big(\pi_l(a,c) - \pi_l(b,c)\big) \leq 0$ (since $\pi_l(b,c) \geq \pi_l(a,c)$).

That implies

$$\tilde{\phi}_l^+(a) - \tilde{\phi}_l^+(b) = \frac{1}{n-2}\Big(\sum_{c \in C} -\pi_l(b,c) + \sum_{c \in D} \big(\pi_l(a,c) - \pi_l(b,c)\big) \Big) \leq 0. \tag{29}$$

In summary, if $\pi_l(a,y) - \pi_l(b,y) < 0$, it is impossible to have $\tilde{\phi}_l^+(a) - \tilde{\phi}_l^+(b) > 0$.

In a similar way, it is easy to show that if $\tilde{\phi}_l^+(a) - \tilde{\phi}_l^+(b) > 0$ then $\pi_l(a,y) - \pi_l(b,y) > 0$.

Corollary 4.1: Rank reversal between two alternatives a and b can not happen in the PROMETHEE I ranking

$$|\phi^+(a) - \phi^+(b)| \geq \frac{1}{n-1} \text{ and } |\phi^-(a) - \phi^-(b)| \geq \frac{1}{n-1}.$$

Figure 7 illustrates this result. We proceed in the same way than explained in Section 3.3. (1,000 multicriteria matrices were randomly generated). The two lines in Figure 7 illustrate the fact that $|\phi^+(a) - \phi^+(b)|$ can not vary more than $\frac{1}{n-1}$ when we delete a random alternative. The area in red represents the points for which rank reversals happened. Figure 8 is a zoom of these areas. They correspond well to a change of sign for $\phi^+(a) - \phi^+(b)$ compared with $\tilde{\phi}^+(a) - \tilde{\phi}^+(b)$. Some points in those areas are blue. They correspond to alternatives a and b incomparable before the deletion and still incomparable after the deletion. In this specific case, the two differences $\phi^+(a) - \phi^+(b)$ and $\phi^-(a) - \phi^-(b)$ have simultaneously changed their sign so that incomparability is well preserved.

In a second step, we realise a comparison of rank reversal instances happening with PROMETHEE I and the one happening when we use a procedure based on a distillation process. This kind of procedure is at the core of ELECTRE ranking methods (Figueira et al., 2005).

To do so, we define the concordance index $c(a,b)$ between two alternatives a and b as follows:

$$c(a,b) = 1 - \pi(b,a).$$

Figure 7 Plot of the couples $(\phi^+(a) - \phi^+(b), \tilde{\phi}^+(a) - \tilde{\phi}^+(b))$ and $(\phi^-(a) - \phi^-(b), \tilde{\phi}^-(a) - \tilde{\phi}^-(b))$ when $n = 10$ and $m = 4$ (see online version for colours)

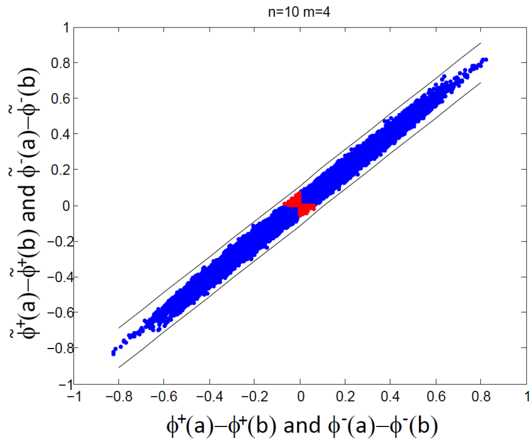
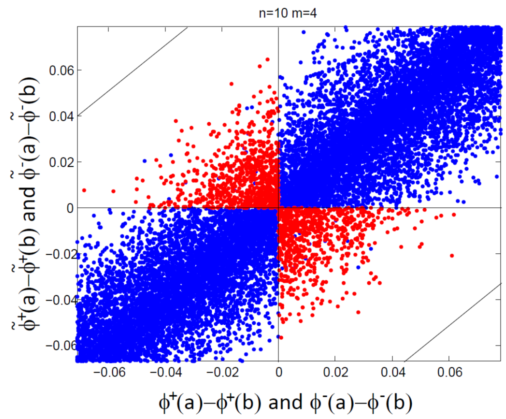


Figure 8 Zoom of Figure 7 (see online version for colours)



Afterwards, we define a strongly outranking relation s_F and a weakly outranking relation s_f in the following way.

$$as_F b \Leftrightarrow \begin{cases} c(a,b) & \geq 0.75 \\ c(a,b) & > c(b,a) \end{cases}$$

$$as_f b \Leftrightarrow \begin{cases} c(a,b) & \geq 0.6 \\ c(a,b) & > c(b,a). \end{cases}$$

Finally, we apply the same distillation process used in ELECTRE II to obtain a partial ranking (Figueira et al., 2005).

In order to illustrate this procedure, let us consider the example presented in Table 10. Applying the distillation procedure will lead to the ranking represented in Figure 9.

Table 10 Example of a multicriteria decision problem leading to a rank reversal instance when applying the PROMETHEE I ranking but not when applying a distillation procedure

	g_1	g_2	g_3	g_4
Min/Max	Max	Max	Max	Max
p	0.5870	0.2077	0.3012	0.4709
q	0.1353	0.1754	0.0587	0.1064
w	0.3143	0.2932	0.1841	0.2084
a	0.1835	0.9294	0.3063	0.6443
b	0.3685	0.7757	0.5085	0.3786
c	0.6256	0.4868	0.5108	0.8116
d	0.7802	0.4359	0.8176	0.5328
e	0.0811	0.4468	0.7948	0.3507

Figure 9 Distillation procedure applied to the example presented in Table 10

$$\mathrm{d} \longrightarrow \mathrm{b} \longrightarrow \mathrm{c} \longrightarrow \mathrm{a} \longrightarrow \mathrm{e}$$

If we delete the alternative a, we obtain the ranking represented in Figure 10. No rank reversal occur.

Figure 10 Distillation procedure applied to the example presented in Table 10 when the alternative a is deleted

$$\mathrm{d} \longrightarrow \mathrm{b} \longrightarrow \mathrm{c} \longrightarrow \mathrm{e}$$

Now, if we compute the PROMETHEE I ranking, we obtain the partial ranking represented in Figure 11.

Figure 11 PROMETHEE I applied to the example presented in Table 10

$$\begin{tikzcd} & \mathrm{c} \arrow[dr] & & \\ \mathrm{d} \arrow[ur] \arrow[dr] & & \mathrm{a} \arrow[r] & \mathrm{e} \\ & \mathrm{b} \arrow[ur] & & \end{tikzcd}$$

When a is deleted, one may observe that the relation between d and b changes in the PROMETHEE I ranking (see Figures 11 and 12).

It is obvious that the two methods often give different results. Sometimes rank reversal can occur in a method but not in the other. To compare the rank reversal phenomenon in the two approaches, we define the similarity rate as the number of common relations before and after the deletion of an alternative divided by the total number of relations after the deletion. Obviously if the similarity rate is equal to 1, there is no rank reversal.

Figure 12 PROMETHEE I applied to the example presented in Table 10 when alternative a is deleted

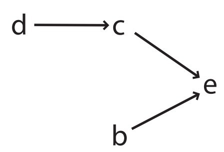
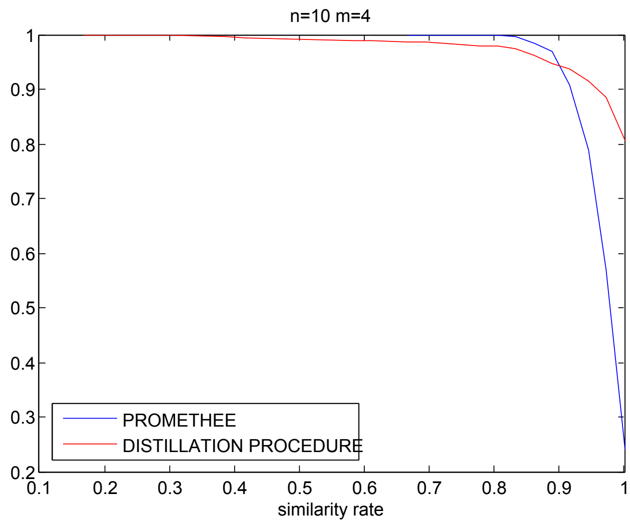


Figure 13 illustrates the complementary cumulative distribution of the similarity rate. We build this distribution by simulating 1000 multicriteria problems as explained in Section 3.3

Figure 13 Complementary cumulative distribution of the similarity rate (see online version for colours)



The distillation procedure does not suffer from rank reversal in about 80% of the cases while the PROMETHEE I ranking suffer from it in about 75%.

Tables 11 and 12 summarise the different kinds of rank reversal instances which occur in the two procedures. We denote J the incomparability relation and I the indifference relation. About the strict preference, denoted P, we split it into two sets: P^+ and P^- such that $P^+ \cup P^- = P$ and $P^+ \cap P^- = \emptyset$. This partition is useful to point up the inversion in a preference (aPb becomes bPa).

The first line of these tables represents how the incomparability is modified after the deletion of an alternative. For example, we see that for the distillation procedure 34(= 13 + 21) incomparabilities become preferences. This number is equal to 735(= 363 + 372) in the case of PROMETHEE I. In both cases, no incomparability becomes indifference.

The second and the third lines deal with strict preferences. We see that 33(= 17 + 16) preferences become incomparabilities in the case of the distillation procedure while this number is equal to 793(= 409 + 384) for PROMETHEE I. We notice that in the first case 59(= 31 + 28) preferences are reversed and in the second case only 24(= 12 + 12) preferences are reversed. In the distillation procedure, we have also 94(= 44 + 50) preferences which become indifferences.

The last line is dedicated to indifference relations. We see that for PROMETHEE I no change occurs at this level. This is logical because our data are randomly generated and we use a linear function to define the preference degree. Therefore, the probability of having an indifference relation tends to 0. In the distillation case, we see that 616(= 326 + 290) indifferences become preferences and 1 indifference becomes an incomparibility.

In conclusion, a procedure based on a distillation process seems to be more stable than PROMETHEE I. This result is not surprising since the PROMETHEE I partial ranking is based on ordinal information deduced from the crisp comparison between real number.

In other words, a will be ranked before b if, for instance, $\phi^+(a) > \phi^+(b)$ and $\phi^-(a) < \phi^-(b)$ no matter the difference between $\phi^+(a)$ and $\phi^+(b)$ (or between $\phi^-(a)$ and $\phi^-(b)$). Of course, results presented in Proposition 4.1 and Corollary 4.1 plead for a softer interpretation of these scores.

Table 11 Different types of rank reversal instances in the distillation procedure

		After deletion			
		J	P^+	P^-	I
Before deletion	J	0	13	21	0
	P^+	17	0	31	44
	P^-	16	28	0	50
	I	1	326	290	0

Table 12 Different types of rank reversal instances in PROMETHEE I

		After deletion			
		J	P^+	P^-	I
Before deletion	J	0	363	372	0
	P^+	409	0	12	0
	P^-	384	12	0	0
	I	0	0	0	0

5 Conclusions

Multicriteria methods like AHP, ELECTRE and PROMETHEE suffer from rank reversal problems. Although this distinctive feature has already been studied by several authors, it is not yet fully formalised and still generates a lot of debates.

In this paper, we have focused our attention on rank reversal instances in the PROMETHEE I and II methods. In both kinds of rankings, we have shown that rank reversal may only occur if flow scores differences are relatively small. This has been justified on the basis of empirical simulations and two theoretical bounds have been provided. We have also proven that the deletion of a non-discriminating criterion will never induce a rank reversal. Finally, a comparison with a procedure based on a distillation process has been realised.

The main conclusion of this paper is that rank reversal may effectively occur in the PROMETHEE I and II ranking. Nevertheless, it is possible to identify situations that will not lead to these kinds of phenomena. The results shown in this contribution clearly prove that rank reversal strongly depends on flow score differences. If we consider the deletion of one alternative, only close alternatives can eventually be swapped. As a consequence, we plead, for a softer interpretation of flow score comparison to deduce ordinal information.

References

Barzilai, J. and Golany, B. (1994) 'AHP rank reversal, normalization and aggregation rules', *INFOR*, Vol. 32, No. 2, pp.57–63.

Belton, V. and Gear, T. (1983) 'On a short-comming of Saaty's method of analytic hierarchies', *Omega*, Vol. 11, No. 3, pp.228–230.

Brans, J.P. and Mareschal, B. (2005) 'PROMETHEE methods', in Figueira, J., Greco, S. and Ehrgott, M. (Eds.): *Multiple Criteria Decision Analysis: State of the Art Surveys*, pp.163–195, Springer.

Buela-Casal, G., Gutiérrez-Martinez, O., Bermudez-Sanchez, M. and Vadillo-Munoz, O. (2007) 'Comparative study of international academic rankings of universities', *Scientometrics*, Vol. 71, No. 3, pp.349–365.

De Keyser, W. and Peeters, P. (1996) 'A note on the use of PROMETHHE multicriteria methods', *European Journal of Operational Research*, Vol. 89, No. 3, pp.457–461.

Figueira, J., Mousseau, V. and Roy, B. (2005) 'ELECTRE methods', in Figueira, J., Greco, S. and Ehrgott, M. (Eds.): *Multiple Criteria Decision Analysis: State of the Art Surveys*, pp.133–162, Springer.

Finan, J.S. and Hurley, W.J. (2002) 'The analytic hierarchy process: can was criteria be ignored', *Computer & Operations Research*, Vol. 29, No. 8, pp.1025–1030.

Harker, P.T. and Vargas, L.G. (1987) 'The theory of ratio scale estimation: Saaty's analytic hierarchy process', *Management Science*, Vol. 33, No. 11, pp.1383–1403.

Kivinen, O. and Hedman J. (2008) 'World-wide university rankings: a Scandinavian approach', *Scientometrics*, Vol. 74, No. 3, pp.391–408.

Liberatore, M.J. and Nydick, R.L. (2004) 'Wash criteria and the analytic hierarchy process', *Computer & Operations Research*, Vol. 31, No. 6, pp.889–892.

Liu, N.C. and Cheng, Y. (2005) 'The academic ranking of world universities', *Higher Education in Europe*, Vol. 30, No. 2, pp.127–136.

Mareschal, B., De Smet, Y. and Nemery, P. (2008) 'Rank reversal in the PROMETHEE II method: Some new results', *Proceedings of de IEEE 2008 International Conference on Industrial Engineering and Engineering Management*, Singapore, pp.959–963.

Saaty, T.L. (1987) 'Decision making, new information, ranking and structure', *Mathematical Modelling*, Vol. 8, No. 8, pp.125–132.

Saaty, T.L. and Vargas, L.G. (1984) 'The legitimacy of rank reversal', *Omega*, Vol. 12, No. 5, pp.513–516.

Triantaphyllou, E. (2001) 'Two new cases of rank reversal when the AHP and some of its additive variants are used that do not occur with the multiplicative AHP', *Journal of Multi-Criteria Decision Analysis*, Vol. 10, No. 1, pp.11–25.

Van Raan, A.F.J. (2005) 'Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods', *Scientometrics*, Vol. 62, No. 1, pp.133–143.

Wang, Y-M. and Elhag, T.M.S. (2006) 'An approach to avoiding rank reversal in AHP', *Decision Support Systems*, Vol. 42, No. 3, pp.1474–1480.

Wang, X. and Triantaphyllou, E. (2008) 'Ranking irregularities when evaluating alternatives by using some ELECTRE methods', *Omega*, Vol. 36, No. 1, pp.45–63.

Wijnmalen, D.J.D. and Wedley, W.C. (2009a) 'Correcting illegitimate rank reversals: proper adjustment of criteria weights prevent alleged AHP intransitivity', *Journal of Multi-Criteria Decision Analysis*, Vol. 15, Nos. 5–6, pp.135–141.

Wijnmalen, D.J.D. and Wedley, W.C. (2009b) 'Non-discriminating criteria in the AHP: removal and rank reversal', *Journal of Multi-Criteria Decision Analysis*, Vol. 15, Nos. 5–6, pp.143–149.