Ann Oper Res (2008) 163: 63–88
DOI 10.1007/s10479-008-0353-y

NP-hardness results for the aggregation of linear orders into median orders

Olivier Hudry

Published online: 29 May 2008

Abstract Given a collection Π of individual preferences defined on a same finite set of candidates, we consider the problem of aggregating them into a collective preference minimizing the number of disagreements with respect to Π and verifying some structural properties. We study the complexity of this problem when the individual preferences belong to any set containing linear orders and when the collective preference must verify different properties, for instance transitivity. We show that the considered aggregation problems are NP-hard for different types of collective preferences (including linear orders, acyclic relations, complete preorders, interval orders, semiorders, quasi-orders or weak orders), if the number of individual preferences is sufficiently large.

Keywords Complexity · Partially ordered relations · Median relations · Aggregation of preferences

1 Introduction

The problem that we deal with in this paper can be stated as follows: given a collection (called a *profile*) $\Pi = (R_1, R_2, \dots, R_m)$ of m binary relations R_i $(1 \le i \le m)$ defined on the same finite set X, find a binary relation R^* defined on X verifying certain properties like transitivity and summarizing Π as accurately as possible. This problem occurs in different fields, for instance in the social sciences, in electrical engineering, in agronomy or in mathematics (for references, see for example Hubert 1976; Barthélemy and Monjardet 1981, 1988; Barthélemy et al. 1986, 1989, 1995; Jünger 1985; Reinelt 1985; Guénoche et al. 1994; Charon and Hudry 2007). For example, in voting theory, X can be considered as a set of candidates, Π as a profile of individual preferences expressed by voters and R^* as the collective preference that we look for. Though the problem occurs in different fields, as said above, we shall keep this illustration from the aggregation of preferences in the following.

O. Hudry (✉)
École Nationale Supérieure des Télécommunications, 46, rue Barrault, 75634 Paris cedex 13, France
e-mail: Olivier.Hudry@enst.fr

The aim of this paper is to study the complexity of finding R^* (for the theory of complexity, see for instance Garey and Johnson 1979 or Barthélemy et al. 1996). We consider different types of ordered relations for the individual preferences of Π as well as for R^*, but we assume that the individual preferences of Π all belong to a set containing linear orders. We show then that, for the considered structures for R^*, the computation of R^* is usually an NP-hard problem if we assume that the number m of individual preferences is large enough.

Similar problems have already been studied, for instance the aggregation of a profile of linear orders into a linear order by Orlin (1981), Bartholdi III et al. (1989), and Dwork et al. (2001), the aggregation of a profile of binary relations into some partially ordered relations by Wakabayashi (1986, 1998), the aggregation of a profile of some partially ordered relations into some partially ordered relations by Hudry (1989, 2004), the approximation of a complete asymmetric relation (a tournament) by a linear order by Ailon et al. (2005), Alon (2006), Conitzer (2006), and Charbit et al. (2007), the approximation of a symmetric relation by an equivalence relation by Krivanek and Moravek (1986), the aggregation of a profile of symmetric relations or of equivalence relations into an equivalence relation by Barthélemy and Leclerc (1995) (see also Hudry 2008 for the complexity of other voting procedures).

The paper is organized as follows. Section 2 recalls the definitions of the ordered relations that we take into account. In Sect. 3, we show how the aggregation problems can be formulated in graph theoretical terms. Then we state our complexity results upon these aggregation problems in Sect. 4. The conclusions take place in Sect. 5 and summarize the main results obtained in Sect. 4.

2 The ordered relations

Given a finite set X, a binary relation R defined on X is a subset of $X \times X = \{(x, y) : x \in X \text{ and } y \in X\}$. We note n the number of elements of X and we suppose that n is large enough (as assumed usually in the theory of complexity). We note xRy instead of $(x, y) \in R$ and $x\overline{R}y$ instead of $(x, y) \notin R$. The following properties that a binary relation R can satisfy are basic:

- reflexivity: $\forall x \in X, xRx$;
- irreflexivity: $\forall x \in X, x\overline{R}x$;
- antisymmetry: $\forall (x, y) \in X^2$, (xRy and $x \neq y$) $\Rightarrow y\overline{R}x$;
- asymmetry: $\forall (x, y) \in X^2, xRy \Rightarrow y\overline{R}x$;
- transitivity: $\forall (x, y, z) \in X^3$, ($xRy$ and yRz) $\Rightarrow xRz$;
- completeness: $\forall (x, y) \in X^2$ with $x \neq y$, xRy or (inclusively) yRx.

From any binary relation R, we may define an asymmetric relation R^a (called the *asymmetric part* of R) by: $xR^ay \Leftrightarrow (xRy$ and $y\overline{R}x)$.

By combining the above properties, we may define different types of binary relations (see for instance Barthélemy and Monjardet 1981; Caspard et al. 2007; Fishburn 1985 or Monjardet 1978). As a binary relation R defined on X is the same as the directed graph $G = (X, R)$ (i.e. (x, y) is an arc of G if and only if we have xRy), we illustrate these types with graph theoretic examples. Usually, it is possible to define reflexive or irreflexive versions of the following structures; as reflexivity and irreflexivity do not matter for the complexity results (see Lemma 1), we give only one version below among these two possibilities.

- A *partial order* is an asymmetric and transitive binary relation (see Fig. 1); $\mathcal{O}$ will denote the set of the partial orders defined on X.

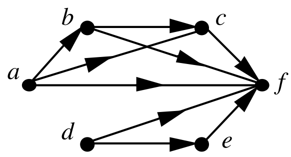


Fig. 1 A partial order

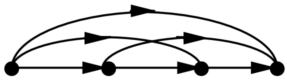


Fig. 2 A linear order. The partial order of Fig. 1 is not a linear order, for instance because the vertices a and d are not compared

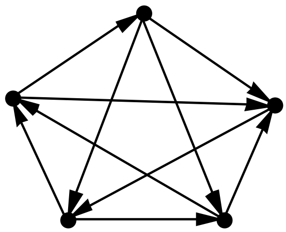


Fig. 3 A tournament

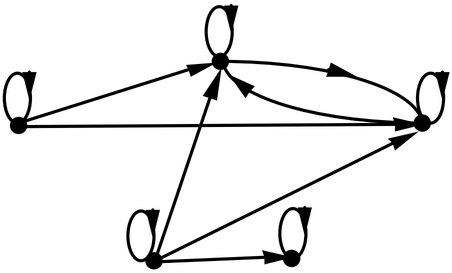


Fig. 4 A preorder

- A *linear order* is a complete partial order (see Fig. 2); $\mathcal{L}$ will denote the set of the linear orders defined on X.
- A *tournament* is a complete and asymmetric binary relation (see Fig. 3); $\mathcal{T}$ will denote the set of the tournaments defined on X; notice that a transitive tournament is a linear order and conversely.
- A *preorder* is a reflexive and transitive binary relation (see Fig. 4); $\mathcal{P}$ will denote the set of the preorders defined on X.
- A *complete preorder* is a reflexive, transitive and complete binary relation (see Fig. 5); $\mathcal{C}$ will denote the set of the complete preorders defined on X.
- A *weak order* (also sometimes called *strict weak order*) is the asymmetric part of a complete preorder (see Fig. 6); $\mathcal{W}$ will denote the set of the weak orders defined on X. Notice that a weak order is a partial order (the converse is not necessarily true: it is easy to see that the partial order of Fig. 1 is not a weak order, for instance because of vertices a, b, d).
- An *interval order* is a partial order R satisfying: $\forall(x, y, z, t) \in X^4$, $(xRy$ and $zRt) \Rightarrow (xRt$ or (inclusively) $zRy)$; $\mathcal{I}$ will denote the set of the interval orders defined on X. The previous relation means that the graph associated with a partial order cannot contain two arcs (x, y) and (z, t) without containing also at least one of the two arcs (x, t) or (z, y)

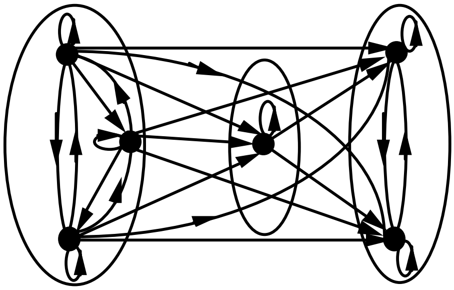


Fig. 5 A complete preorder

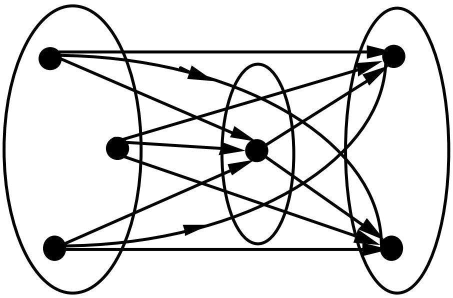


Fig. 6 A weak order (namely, the asymmetric part of the complete preorder of Fig. 5)

(see Fig. 7); for instance, the partial order of Fig. 1 is not an interval order because of the vertices b, c, d, e: we have bRc and dRe but we have not bRe nor dRc. The name "interval order" comes from the fact that such an order may be associated with intervals defined on the set of real numbers and spread over the real axis (see Fig. 7). The relation xRy then means that the interval associated with x is utterly on the left of the interval associated with y; if the two intervals overlap, then we have simultaneously $x\overline{R}y$ and $y\overline{R}x$. The implication (xRy and zRt) $\Rightarrow$ (xRt or (inclusively) zRy) means that, if the interval associated with x (resp. z) is completely on the left of the interval associated with y (resp. t), the interval associated with x must be on the left of the interval associated with t or the interval associated with z must be on the left of the interval associated with y. Notice that the lengths of the intervals are not necessarily the same for all the intervals.

We may derive another set from $\mathcal{I}$: the set $\mathcal{J}$ of relations such that the asymmetric part is an interval order, sometimes called *interval relation* (see Fig. 8). In other words, $\mathcal{J}$ is defined by: $J \in \mathcal{J} \Leftrightarrow J^a \in \mathcal{I}$.

- A *semiorder* is an interval order R satisfying: $\forall (x, y, z, t) \in X^4$, ($xRy$ and yRz) $\Rightarrow$ (xRt or (inclusively) tRz) (see Fig. 9); $\mathcal{S}$ will denote the set of the semiorders defined on X. For instance, the interval order of Fig. 7 is not a semiorder because of the vertices b, c, d, e: we have cRe and eRd but we have not cRb nor bRd. As an interval order, a semiorder may be represented by intervals. In this case, all the intervals have the same length. Then the implication (xRy and yRz) $\Rightarrow$ (xRt or (inclusively) tRz) means that, if the interval associated with x (resp. y) is completely on the left of the interval associated with y (resp. z), then the interval associated with any extra t cannot overlap the interval associated with x and the one associated with z simultaneously: the length of the interval associated with t is too small for that.
- A *quasi-order* is a reflexive and complete relation of which the asymmetric part is a semiorder (see Fig. 10); $\mathcal{Q}$ will denote the set of the quasi-orders defined on X.

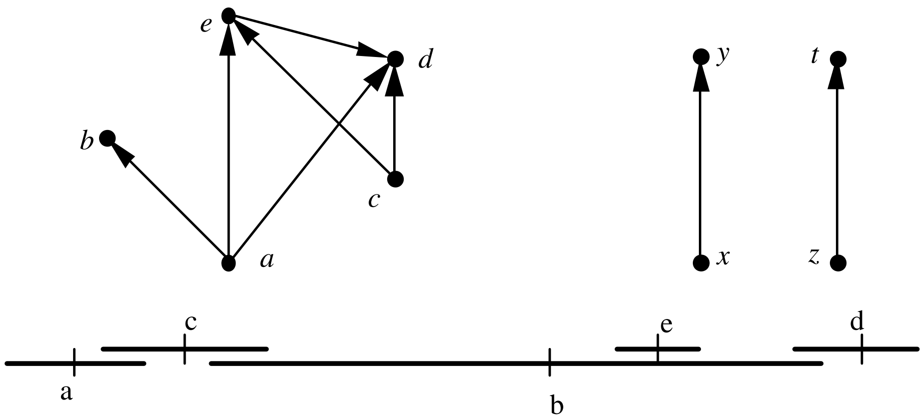
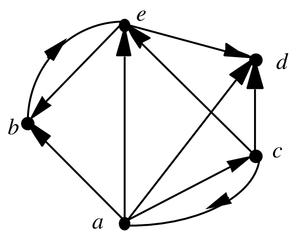


Fig. 7 An interval order (*left*), a possible representation as intervals (*bottom*) and the forbidden configuration (*right*)

Fig. 8 A relation belonging to $\mathcal{J}$ (its asymmetric part is an interval order; more precisely, its asymmetric part is the interval order of Fig. 7)



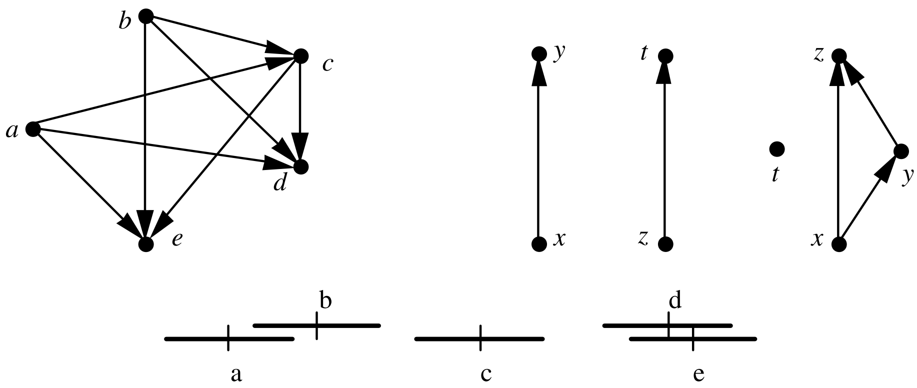


Fig. 9 A semiorder (*left*), a possible representation as intervals (*bottom*), and the forbidden two configurations—including the one of interval orders—(*right*)

- An *acyclic relation* (we should rather say *circuitless*, but *acyclic* is the usual term) is a relation R of which the associated graph G is without any circuit (i.e., any directed cycle) (see Fig. 11). In other words, if, for an appropriate integer k with $1 \le k \le n$ and an appropriate numbering $i_1, i_2, \dots, i_k$ of some vertices x_{i_j} of G, we have $x_{i_j} R x_{i_{j+1}}$ for $1 \le j \le k-1$, then we must have $x_{i_k} \overline{R} x_{i_1}$; $\mathcal{A}$ will denote the set of acyclic relations defined on X.

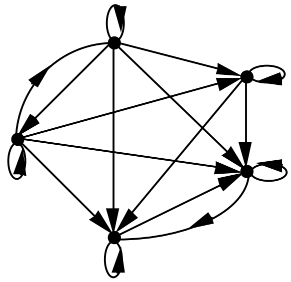


Fig. 10 A quasi-order. Its asymmetric part is the semiorder of Fig. 9

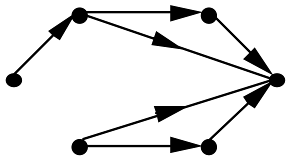


Fig. 11 An acyclic relation. Its transitive closure is the partial order of Fig. 1

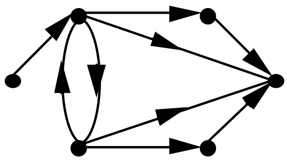


Fig. 12 A relation belonging to $\mathcal{B}$ (its symmetric part is acyclic; more precisely, its symmetric part is the acyclic relation of Fig. 11)

We may derive another set from $\mathcal{A}$: the set $\mathcal{B}$ of relations such that the asymmetric part is acyclic (see Fig. 12). In other words, $\mathcal{B}$ is defined by: $B \in \mathcal{B} \Leftrightarrow B^a \in \mathcal{A}$.

Checking that a given relation (or a given graph) fulfills the requirements of these structures can be done in polynomial time with respect to n. From this remark, it will follow that the decision problems associated with the problems considered below all belong to NP.

As said above, it is possible to get other structures by adding or by removing reflexivity or irreflexivity from the above definition (and by changing asymmetry by antisymmetry or conversely when necessary). In fact, the distinction between reflexive and irreflexive relations is not relevant for this study, as we shall see below (Lemma 1): the complexity results will remain the same. Thus, in the following, we do not take reflexivity or irreflexivity into account (for instance, we will consider that a linear order is also a preorder).

These types include the most studied and used partially ordered relations. We will also consider generic binary relations, without any particular property. The set of the binary relations will be noted $\mathcal{R}$. We may notice several inclusions between these sets, especially the following ones: $\forall \mathcal{Z} \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, W\}$, $\mathcal{L} \subseteq \mathcal{Z}$; in other words, a linear order can be considered as a special case of any one of the other types.

3 Formulations of the aggregation problems

In order to get an optimization problem to deal with, it is necessary to explicit what we mean when we say that R^* must summarize Π "as accurately as possible". To do so, we consider the symmetric difference distance δ: given two binary relations R and S defined on the same set X, we have

$$\delta(R, S) = \left|\left\{(x, y) \in X^2 : [xRy \text{ and } x\overline{S}y] \text{ or } [x\overline{R}y \text{ and } xSy]\right\}\right|.$$

This quantity $\delta(R,S)$ measures the number of disagreements between R and S. Though it is possible to consider another distance, δ is used widely and is appropriate for many applications. Barthélemy (1979) shows that δ satisfies a number of naturally desirable properties, and Barthélemy and Monjardet (1981) recall that $\delta(R,S)$ is the Hamming distance between the characteristic vectors of R and S and point out the links between δ and the L_1 metric or the square of the Euclidean distance between these vectors (see also Bogart 1973, 1975 and Monjardet 1979, 1990).

Then, for a profile $\Pi = (R_1, R_2, \dots, R_m)$ of m relations, we can define the *remoteness* $\Delta(\Pi, R)$ (Barthélemy and Monjardet 1981) between a relation R and the profile Π by:

$$\Delta(\Pi, R) = \sum_{i=1}^{m} \delta(R, R_i).$$

The remoteness $\Delta(\Pi, R)$ measures the total number of disagreements between Π and R.

Our aggregation problem can be seen now as a combinatorial optimization problem: given a profile Π, determine a binary relation R^* minimizing Δ over one of the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{W}$. Such a relation R^* will be called a *median relation* of Π (Barthélemy and Monjardet 1981). According to the number m of relations of the profile and to the properties assumed for the relations belonging to Π or required from the median relation, we get many combinatorial optimization problems. They are too numerous to state all of them explicitly; so we note them as follows:

Problems $P_m(\mathcal{Y}, \mathcal{Z})$ For $\mathcal{Y}$ belonging to $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{W}\}$ and $\mathcal{Z}$ belonging also to $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{W}\}$, for a positive integer m, $P_m(\mathcal{Y}, \mathcal{Z})$ denotes the following problem: given a finite set X of n elements, given a profile Π of m binary relations all belonging to $\mathcal{Y}$, find a relation R^* belonging to $\mathcal{Z}$ with the following property:

$$\Delta(\Pi, R^*) = \mathrm{Min}\{\Delta(\Pi, R) \text{ for } R \in \mathcal{Z}\}.$$

We are going to study the complexity of the problems $P_m(\mathcal{L}, \mathcal{Z})$ for different sets $\mathcal{Z}$ more particularly. In order to simplify the notations, $P_m(\mathcal{Z})$ will denote the problem $P_m(\mathcal{L}, \mathcal{Z})$. We will see that the parity of m will play a role in the study of the problems $P_m(\mathcal{Z})$. Anyway, it will be easy to see from the following computations that, if $P_m(\mathcal{Z})$ is NP-hard for some m and some set $\mathcal{Z}$, then $P_{m+2}(\mathcal{Z})$ will also be NP-hard (see Lemma 3 below).

We do not explicit the statements of the decision problems associated with the problems $P_m(\mathcal{Y}, \mathcal{Z})$, because they are obvious. Similarly, it is obvious to show that these decision problems belong to NP. Thus, we deal with the NP-hardness of $P_m(\mathcal{Y}, \mathcal{Z})$, but we could have dealt with the NP-completeness of the decision problems associated with $P_m(\mathcal{Y}, \mathcal{Z})$ in a similar way.

To study the complexity of $P_m(\mathcal{Y}, \mathcal{Z})$, we develop the expression of $\Delta(\Pi, R)$, for any profile $\Pi = (R_1, R_2, \dots, R_m)$ of m binary relations R_i $(1 \le i \le m)$ all belonging to $\mathcal{Y}$. For this, consider the characteristic vectors $r^i = (r^i_{xy})_{(x,y)\in X^2}$ of the relations R_i $(1 \le i \le m)$ defined by $r^i_{xy} = 1$ if $x R_i y$ and $r^i_{xy} = 0$ otherwise, and similarly the characteristic vector $r = (r_{xy})_{(x,y)\in X^2}$ of any binary relation R. Then, it is easy to get a linear expression of $\Delta(\Pi, R)$:

$$\delta(R, R_i) = \sum_{(x,y)\in X^2} |r_{xy} - r^i_{xy}| = \sum_{(x,y)\in X^2} |r_{xy} - r^i_{xy}|^2 = \sum_{(x,y)\in X^2} [r_{xy}(1 - 2r^i_{xy}) + r^i_{xy}]$$

hence

$$\Delta(\Pi, R) = \sum_{i=1}^{m} \sum_{(x,y)\in X^2} |r_{xy} - r^i_{xy}|$$

and, after simplifications:

$$\Delta(\Pi, R) = C - \sum_{(x,y)\in X^2} m_{xy} \cdot r_{xy}$$

with $C = \sum_{i=1}^{m} \sum_{(x,y)\in X^2} r^i_{xy}$ and $m_{xy} = \sum_{i=1}^{m}(2r^i_{xy} - 1) = 2\sum_{i=1}^{m} r^i_{xy} - m$.

Notice that the quantities m_{xy} can be non-positive or non-negative, and that they all have the same parity (the one of m). If we consider a profile of m linear orders R_i $(1 \le i \le m)$, then, for any i with $1 \le i \le m$, we have $r^i_{xy} + r^i_{yx} = 1$ for any x and y with $x \ne y$; hence it comes, still for a profile of linear orders and still for $x \ne y$, the relation $m_{xy} = -m_{yx}$. From this expression of $\Delta(\Pi, R)$, it would be easy to get a 0-1 linear programming formulation of the problems $P_m(\mathcal{Y}, \mathcal{Z})$ by adding the 0-1 linear constraints associated with each type of median relation (but it will not be the way that we are going to follow in the sequel). For example, the transitivity of R can be written: $\forall(x, y, z) \in X^3$, $r_{xy} + r_{yz} - r_{xz} \le 1$ (see for instance Wakabayashi 1986 or Hudry 1989 for details). Such a 0-1 linear programming formulation was applied as soon as 1960 (Tucker 1960; see also Younger 1963, de Cani 1969, Arditti 1984, and more generally Barthélemy and Monjardet 1981 or Charon and Hudry 2007 for references).

Before going further, the following lemma shows that reflexivity or irreflexivity of the median relation do not change the complexity of the problems $P_m(\mathcal{Y}, \mathcal{Z})$.

Lemma 1 *For any set $\mathcal{Z}$ of median relations, let $\mathcal{Z}_r$ (resp. $\mathcal{Z}_i$) be the set of median relations got from the elements of $\mathcal{Z}$ by imposing the reflexivity (resp. irreflexivity) property. Then, for any set $\mathcal{Y}$ and any integer m, $P_m(\mathcal{Y}, \mathcal{Z})$, $P_m(\mathcal{Y}, \mathcal{Z}_r)$, and $P_m(\mathcal{Y}, \mathcal{Z}_i)$ have the same complexity.*

Proof To show this result, consider any profile Π of m relations belonging to $\mathcal{Y}$ and any relation Z belonging to $\mathcal{Z}$. Let Z_r(resp. Z_i) be the reflexive (resp. irreflexive) relation got from Z by imposing the reflexivity (resp. irreflexivity) property. Then it is easy to state the following relations:

$$\Delta(\Pi, Z_r) = \Delta(\Pi, Z) - \sum_{x:(x,x)\notin Z} m_{xx} \quad \text{and} \quad \Delta(\Pi, Z_i) = \Delta(\Pi, Z) + \sum_{x:(x,x)\in Z} m_{xx}.$$

Hence the result, since the computation of $\sum_{x:(x,x)\notin Z} m_{xx}$ and of $\sum_{x:(x,x)\in Z} m_{xx}$ can trivially be done in polynomial time with respect to the size of the considered instance. □

Because of Lemma 1, we shall not pay attention from now on to reflexivity or irreflexivity: all the complexity results remain the same if we add or remove reflexivity or irreflexivity.

In the following, we will not consider the previous 0-1 linear programming formulation to study the complexity of the problems $P_m(\mathcal{Y}, \mathcal{Z})$, but a graph theoretic representation. Indeed, we may associate a complete, symmetric, weighted, directed graph $G = (X, U_X)$ to any profile Π: the vertex set of G is X and G owns all the arcs (i.e. directed edges) that a simple graph can own; in other words, we have: $U_X = X \times X - \{(x, x) \text{ for } x \in X\}$

(remember that reflexivity does not matter now on). The arcs (x, y) of G (with $x \in X$, $y \in X$ and $x \neq y$) are weighted by m_{xy}. Then minimizing $\Delta(\Pi, Z)$ for Z belonging to one of the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{W}$ is exactly the same as extracting a partial graph $H = (X, Z)$ from G in order to maximize $\sum_{(x,y)\in Z} m_{xy}$ while the kept arcs describe the structure that Z must respect (H must belong to $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{W}$, where these sets are seen as sets of graphs).

As we are going to deal with profiles of linear orders, the following question arises: which weighted graphs $G = (X, U_X)$ can be associated to a profile Π of linear orders? Debord (1987) provided a characterisation of weighted graphs that represent profiles of linear orders, at least when the number of linear orders is large enough. The number $N(G)$ of linear orders required in his construction is about $\sum_{m_{xy}>0} m_{xy}$ when this quantity is not equal to 0 (as it will be the case below).

Theorem 2 *The graph $G = (X, U_X)$ weighted by the (non-positive or non-negative) integers m_{xy} represents a profile Π of m linear orders if the following conditions are fulfilled*:

1. *all the weights m_{xy} have the same parity*;
2. *m has the same parity as the weights m_{xy}*;
3. $\forall (x, y) \in U_X$, $m_{xy} = -m_{yx}$;
4. $m \geq N(G)$, *where $N(G)$ is the number of linear orders in Debord's construction*.

It is sometimes possible to find a profile with fewer linear orders associated with G, but minimizing the number of linear orders required for the construction of such a profile seems quite difficult in general. Anyway, it is possible to construct a profile of $N(G)$ linear orders with $N(G) = \theta(\sum_{m_{xy}>0} m_{xy})$ (when this sum is not equal to 0). From the particular value $N(G)$, it is easy to build a profile with as many linear orders as wished; this is specified by the following lemma:

Lemma 3 *Let $G = (X, U_X)$ be a graph representing a profile of m linear orders. Then G represents also a profile of $m + 2$ linear orders.*

Proof Let Π be a profile of m relations associated with G. Then consider the profile got from Π by adding any linear order and its reverse order. It is not difficult, from the computations developed before Lemma 1 above, to show that G represents also this new profile. □

It follows also from this construction that if, for some set $\mathcal{Z}$ and some integer m, $P_m(\mathcal{Z})$ is NP-hard, then $P_{m+2}(\mathcal{Z})$ is NP-hard as well, and so is $P_\mu(\mathcal{Z})$ for any integer μ greater than or equal to m and with the same parity as m. In particular, the NP-hardness of $P_m(\mathcal{Z})$ for m greater than or equal to $N(G)$ and with the same parity as $N(G)$ will follow from the one of $P_{N(G)}(\mathcal{Z})$.

Moreover, in the sequel, the weights will be upper-bounded by a polynomial in n; then the construction of Π can be done in polynomial time with respect to the size of G. Indeed, in this case, $N(G)$ is also upper-bounded by a polynomial in n; thus, as any binary relation R defined on X can be described by $O(n^2)$ bits (for instance through the adjacency matrix associated with the graph representing R), it is possible to code $\Pi = (R_1, R_2, \ldots, R_{N(G)})$ with $O(n^2 \cdot N(G))$ bits, what we may upper-bound by a polynomial in n, while the size of G is at least about n^2 (at least 1 bit for the weight of each arc of G); hence the result.

From this polynomial link between the problems $P_m(\mathcal{Z})$ and their graph theoretic representations, it appears that we may study the complexity of the problems $P_m(\mathcal{Z})$ with the help of weighted graphs. It is what we do below. More precisely, we are going to study the following decision problems, stated as graph theoretic problems:

Problems $Q_0(\mathcal{Z})$ for $\mathcal{Z} \in \{\mathcal{A}, \mathcal{C}, \mathcal{I}, \mathcal{L}, \mathcal{S}\}$

Instance: a graph $G = (X, U_X)$ weighted by (non-positive or non-negative) even integers m_{xy} with $m_{xy} = -m_{yx}$ for $x \neq y$; an integer K;
Question: does there exist a partial graph (X, U) of G belonging to $\mathcal{Z}$ with $\sum_{(x,y)\in U} m_{xy} \geq K$?

Problems $Q_1(\mathcal{Z})$ for $\mathcal{Z} \in \{\mathcal{A}, \mathcal{C}, \mathcal{I}, \mathcal{L}, \mathcal{S}\}$

Instance: a graph $G = (X, U_X)$ weighted by (positive or negative) odd integers m_{xy} with $m_{xy} = -m_{yx}$ for $x \neq y$; an integer K;
Question: does there exist a partial graph (X, U) of G belonging to $\mathcal{Z}$ with $\sum_{(x,y)\in U} m_{xy} \geq K$?

Notice that, by Theorem 2, the graphs in the instances of $Q_0(\mathcal{Z})$ or of $Q_1(\mathcal{Z})$ can be polynomialy associated with a profile of linear orders, for an appropriate number of linear orders.

In the following, we will apply the following (obvious) lemma:

Lemma 4

(a) *Any partial graph of a graph without circuit is itself without circuit.*
(b) *Any graph without circuit can be extended into a linear order by adding appropriate arcs.*

4 The complexity results

In this section, we shall study the NP-completeness of the problems $Q_0(\mathcal{Z})$ and $Q_1(\mathcal{Z})$. The NP-hardness of the problems $P_m(\mathcal{Z})$ for the considered sets $\mathcal{Z}$ will follow. As the belonging of the problems $Q_0(\mathcal{Z})$ and $Q_1(\mathcal{Z})$ to NP is obvious for $\mathcal{Z} \in \{\mathcal{A}, \mathcal{C}, \mathcal{I}, \mathcal{L}, \mathcal{S}\}$, we shall omit to verify this belonging in the sequel.

The complexity of $P_4(\mathcal{L})$ has been established by Dwork et al. (2001). From this, the complexity of $P_m(\mathcal{L})$ follows immediately for any even integer $m \geq 4$, as stated by Theorem 5.

Theorem 5 *Let m be any even integer with $m \geq 4$. Then $P_m(\mathcal{L})$ is NP-hard.*

Proof The NP-hardness of $P_4(\mathcal{L})$ is shown by Dwork et al. (2001). The NP-hardness of $P_m(\mathcal{L})$ for $m \geq 4$ follows from the one of $P_4(\mathcal{L})$ by adding $(m-4)/2$ times a linear order L and $(m-4)/2$ times the reversed order of L to the profile of four linear orders associated with any instance of $P_4(\mathcal{L})$, as described above, after Lemma 3. □

As a corollary of Theorem 5, we get obviously the NP-completeness of $Q_0(\mathcal{L})$, even if the absolute values of the weights of the instances of $Q_0(\mathcal{L})$ are upper bounded by any constant greater than or equal to 4.

Corollary 6 *$Q_0(\mathcal{L})$ is NP-complete, and remains so if all the (even) weights are upper-bounded by any (even) constant m with $m \geq 4$.*

Proof It is shown in Sect. 3 how to build, in polynomial time, an instance of $Q_0(\mathcal{L})$ from any instance of $P_m(\mathcal{L})$ for m even. Moreover, in this case, the weights of the instance of $Q_0(\mathcal{L})$ are upper-bounded by m. Hence the statement of Corollary 6, from Theorem 5. □

Remark 7

1. Notice that, as shown by Monjardet (1973), when all the linear orders of the profile agree to consider a candidate x as better than another candidate y, then any median linear order must also consider x as better than y (this is the *unanimity property*). From this remark, it is easy to show that $P_2(\mathcal{L})$ is polynomial (see Charon and Hudry 2007 for details).
2. Anyway, it is possible to show that $Q_0(\mathcal{L})$ remains NP-complete when the absolute values of the weights of the instances of $Q_0(\mathcal{L})$ are upper-bounded by 2 (or even equal to 2; see Remark 11). This qualitative difference with respect to $P_2(\mathcal{L})$ illustrates the fact that, in spite of their likeness, problems $P_m(\mathcal{L})$ and $Q_0(\mathcal{L})$, and more generally problems $P_m(\mathcal{Z})$ and $Q_0(\mathcal{Z})$ for $\mathcal{Z}$ belonging to $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{W}\}$, do not necessarily behave similarly, even if the absolute values of the weights of the instances of $Q_0(\mathcal{L})$ are upper bounded by m. In other words, dealing with graphs instead of profiles may introduce a bias if we do not take care.

Theorem 8 *Let m be any even integer with $m \geq 4$. Then $P_m(\mathcal{A})$ is NP-hard.*

Proof Let Π be the profile of any instance of $P_m(\mathcal{L})$. Let $M_{\mathcal{L}}(\Pi)$ (respectively $M_{\mathcal{A}}(\Pi)$) be the value of a minimum solution of Π with respect to $P_m(\mathcal{L})$ (respectively $P_m(\mathcal{A})$). We are going to establish a link between $M_{\mathcal{L}}(\Pi)$ and $M_{\mathcal{A}}(\Pi)$.

For this, let $G = (X, U_X)$ be the graph associated with Π, and assume that G is weighted by (non-positive or non-negative) even integers m_{xy} with $m_{xy} = -m_{yx}$ for $x \neq y$.

Let (X, L) be a linear order which is an optimal solution for G when we want to extract from G a linear order with a maximum weight. We may write L as the union of two subsets L^+ and L^- of U_X, with $L^+ = \{(x, y) \in L \text{ with } m_{xy} > 0\}$ and $L^- = \{(x, y) \in L \text{ with } m_{xy} \leq 0\}$. Let μ denote the sum of all the positive weights of G: $\mu = \sum_{m_{xy}>0} m_{xy}$; notice that μ depends only on G. Because of the relation $m_{xy} = -m_{yx}$ and because (X, L) is a linear order, we have: $\mu = \sum_{(x,y)\in L^+} m_{xy} - \sum_{(x,y)\in L^-} m_{xy}$. Moreover, $\sum_{(x,y)\in L} m_{xy} = \sum_{(x,y)\in L^+} m_{xy} + \sum_{(x,y)\in L^-} m_{xy}$. Hence the relation $\sum_{(x,y)\in L} m_{xy} = 2\sum_{(x,y)\in L^+} m_{xy} - \mu$. Notice that (X, L^+) is an acyclic graph. So, from any solution (X, L) optimal with respect to the search of a linear order, we may draw a solution (X, L^+) with respect to the search of an acyclic relation with $\sum_{(x,y)\in L^+} m_{xy} = (\sum_{(x,y)\in L} m_{xy} + \mu)/2$. Remember the relations (see above): $\Delta(\Pi, L) = C - \sum_{(x,y)\in L} m_{xy}$ and $\Delta(\Pi, L^+) = C - \sum_{(x,y)\in L^+} m_{xy}$. Hence the equalities $\Delta(\Pi, L^+) = (C - \mu + \Delta(\Pi, L))/2 = (C - \mu + M_{\mathcal{L}}(\Pi))/2$, and then, by definition of a minimum, the inequality $M_{\mathcal{A}}(\Pi) \leq (C - \mu + M_{\mathcal{L}}(\Pi))/2$.

Conversely, let (X, A) be an acyclic graph which is an optimal solution for G when we want to extract from G an acyclic relation with a maximum weight. According to Lemma 4, we may extend A into a linear order L with $A \subseteq L$. Set $L^- = L - A$. Quite obviously, all the arcs which belong to L^- have a non-positive weight (otherwise, as a linear order is acyclic, it would have been possible to add the positive arcs to A, a contradiction with the optimality of A). We may write $\sum_{(x,y)\in L} m_{xy} = \sum_{(x,y)\in A} m_{xy} + \sum_{(x,y)\in L^-} m_{xy}$. As above, we have also the relation $\mu = \sum_{(x,y)\in A} m_{xy} - \sum_{(x,y)\in L^-} m_{xy}$. So: $\sum_{(x,y)\in L} m_{xy} = 2\sum_{(x,y)\in A} m_{xy} - \mu$. From any solution (X, A) optimal with respect to the search of an acyclic relation, we may draw a solution (X, L) with respect to the search of a linear order with $\sum_{(x,y)\in L} m_{xy} = 2\sum_{(x,y)\in A} m_{xy} - \mu$. Hence, after some computations similar to the previous ones: $M_{\mathcal{L}}(\Pi) \leq 2M_{\mathcal{A}}(\Pi) - C + \mu$.

We conclude from this and from the previous inequality: $M_{\mathcal{L}}(\Pi) = 2M_{\mathcal{A}}(\Pi) - C + \mu$. So, computing $M_{\mathcal{L}}(\Pi)$ and computing $M_{\mathcal{A}}(\Pi)$ have the same complexity. Hence the statement of Theorem 8, because of the NP-hardness of $P_m(\mathcal{L})$ for any even m with $m \geq 4$. □

Notice that, from the polynomiality of $P_2(\mathcal{L})$, it is possible to show that $P_2(\mathcal{A})$ is polynomial.

To study the complexity of $P_m(\mathcal{L})$ when m is odd, we use a recent result, dealing with the so-called *Feedback Arcset Problem* (FAS) when applied to tournaments. This problem can be stated as follows, for general oriented graphs.

Problem FAS

Instance: a directed, asymmetric graph $H = (Y, W)$; an integer h;
Question: does there exist $W' \subset W$ with $|W'| \leq h$ and such that removing the elements of W' from H leaves a graph $(Y, W - W')$ without any circuit?

This problem has been known to be NP-complete for a long time (Karp 1972; see also Garey and Johnson 1979). Its status for tournaments remained open until recently. Indeed, Alon (2006), Conitzer (2006), and Charbit et al. (2007) show independently that FAS is also NP-complete even if the graph H of the considered instance is assumed to be a tournament. We call FAST this variant of FAS (it is also known as Slater's problem—see Slater 1961—as well as under other names—see for instance Charon and Hudry 2007 for a survey on Slater's problem):

Theorem 9 *The following problem is NP-complete.*

Problem FAST

Instance: a tournament $T = (Y, W)$; an integer h;
Question: does there exist $W' \subset W$ with $|W'| \leq h$ and such that removing the elements of W' from T leaves a graph $(Y, W - W')$ without any circuit?

From the NP-completeness of FAST, it is easy to derive the one of $Q_1(\mathcal{L})$:

Theorem 10 *Problem $Q_1(\mathcal{L})$ is NP-complete, even if all the weights involved in the instances of $Q_1(\mathcal{L})$ belong to* $\{-1, 1\}$.

Proof Let $T = (Y, W)$ and h be any instance of FAST, where T is assumed to be a tournament and h an integer. We transform this instance of FAST into an instance $G = (X, U_X)$ and K of $Q_1(\mathcal{L})$ as follows. First, we set $X = Y$. For all the arcs (x, y) of U_X, we choose the weight m_{xy} as equal to 1 if (x, y) belongs to W, to -1 otherwise. Notice that, as T is a tournament, we have $m_{xy} = -m_{yx}$ for any arc (x, y): so this transformation defines an instance of $Q_1(\mathcal{L})$. Then we set $K = n(n-1)/2 - 2h$.

This transformation is obviously polynomial with respect to the size of the instance (T, h) of FAST.

Moreover, it keeps the answer.

Indeed, assume that (T, h) admits the answer "yes". Then there exists $W' \subset W$ with $|W'| \leq h$ and such that the graph $(Y, W - W')$ is without any circuit. By Lemma 4b, we may complete $(Y, W - W')$ into a linear order (X, L), with $W - W' \subseteq L$. Notice that it may happen that L contains some elements of W'. But, in any case, there are at most $|W'|$

elements of L which do not belong to W, and thus which have a weight equal to -1. The other arcs of L have a weight equal to 1. So, the weight of (X, L) is at least equal to $|W - W'| - |W'|$, i.e. at least equal to $|W| - 2|W'|$, which is greater than or equal to K, since W owns $n(n-1)/2$ elements while $|W'|$ is less than or equal to h.

Conversely, assume that the built instance (G, K) of $Q_1(\mathcal{L})$ admits the answer "yes": there exists a linear order (X, L) extracted from G with a weight greater than or equal to K. Then define W' as the set of the arcs belonging to W but not to $L : W' = W - L$. As $W - W'$ is included into L which is without circuit, then, by Lemma 4a, $(Y, W - W')$ is without circuit. Moreover, the arcs (x, y) of L which do not belong to W are such that (y, x) belongs to W' (because of the completeness of a tournament and of a linear order). The weights of these arcs are equal to -1, while the other arcs of L, which belong to W, have a weight equal to 1. So, the weight of L, which is greater than or equal to K, is also equal to $|W| - 2|W'|$. This involves that $|W'|$ is less than or equal to h. So, the instance (T, h) of FAST admits the answer "yes".

This completes the proof. □

Remark 11 From the NP-completeness of $Q_1(\mathcal{L})$ when all the weights are equal to -1 or to 1, it is easy to draw the NP-completeness of $Q_0(\mathcal{L})$ when all the weights are equal to -2 or to 2: it is sufficient to double all the weights and the integer K of any instance of $Q_1(\mathcal{L})$ to obtain polynomialy an instance of $Q_0(\mathcal{L})$ admitting the same answer. Anyway, $P_2(\mathcal{L})$ is polynomial (see Remark 7).

The transformation described in Remark 11 can be done more generally from any problem $Q_1(\mathcal{Z})$ into the problem $Q_0(\mathcal{Z})$, for any set $\mathcal{Z}$. Anyway, we will not apply such a transformation below. On the contrary, we will first establish the NP-completeness of some problems $Q_0(\mathcal{Z})$ because we can exploit the possibility of some weights equal to 0. Then it is easy, by changing the range of the weights to establish the NP-completeness of the homologous problems $Q_1(\mathcal{Z})$. This is specified by the next lemma.

Lemma 12 *Let $\mathcal{Z}$ be a set of relations. If $Q_0(\mathcal{Z})$ is NP-complete, then $Q_1(\mathcal{Z})$ is also NP-complete, even if the range of the weights involved in $Q_1(\mathcal{Z})$ is upper-bounded by n^2 times the range of the weights involved in $Q_0(\mathcal{Z})$.*

Proof Let (G_0, K_0) be any instance of $Q_0(\mathcal{Z})$. We define an instance (G_1, K_1) of $Q_0(\mathcal{Z})$ as follows. First, we set $G_1 = G_0$, but the weights of G_0 and of G_1 will not be the same. Let m^0_{xy} (respectively m^1_{xy}) denote the weight of an arc (x, y) of G_0 (respectively of G_1). If m^0_{xy} is not equal to 0, we set $m^1_{xy} = 2n^2 m^0_{xy} + 1$ if m^0_{xy} is positive, or $m^1_{xy} = 2n^2\ m^0_{xy} - 1$ otherwise. For each pair of arcs (x, y) and (y, x) with $m^0_{xy} = m^0_{yx} = 0$, we set ($m^1_{xy} = 1$ and $m^1_{yx} = -1$) or ($m^1_{xy} = -1$ and $m^1_{yx} = 1$) arbitrarily. Notice that the weights m^1_{xy} of G_1 satisfy the requirements of Problem $Q_1(\mathcal{Z})$. Notice also the following relation, for any arc (x, y): $m^1_{xy} = 2n^2 m^0_{xy} \pm 1$. Last, we set $K_1 = 2n^2 K_0 - n(n-1)$.

This transformation is obviously polynomial.

Let us show that the answer is kept. Indeed, if (G_0, K_0) admits the answer "yes", there exists a partial graph H of G_0 satisfying the properties of $\mathcal{Z}$ and with a weight greater than or equal to K_0: $\sum_{(x,y)\in H} m^0_{xy} \geq K_0$. If we consider H as a partial graph of G_1, as H contains at most $n(n-1)$ arcs and because of the relation $m^1_{xy} \geq 2n^2 m^0_{xy} - 1$ for any arc (x, y), its weight with respect to G_1 is at least K_1: $\sum_{(x,y)\in H} m^1_{xy} \geq \sum_{(x,y)\in H}(2n^2 m^0_{xy} - 1) \geq 2n^2 K_0 - n(n-1) = K_1$. So the answer admitted by (G_1, K_1) is also "yes".

Conversely, if (G_1, K_1) admits the answer "yes", there exists a partial graph H of G_1 satisfying the properties of $\mathcal{Z}$ and with a weight greater than or equal to K_1 : $\sum_{(x,y)\in H} m^1_{xy} \geq K_1$. If we consider H as a partial graph of G_0, let us show that its weight with respect to G_0 is at least K_0. Assume the contrary. It would involve the following inequality: $\sum_{(x,y)\in H} m^0_{xy} \leq K_0 - 1$. As H contains at most $n(n-1)$ arcs and because of the relation $m^1_{xy} \leq 2n^2 m^0_{xy} + 1$ for any arc (x, y), we would have: $\sum_{(x,y)\in H} m^1_{xy} \leq 2(K_0 - 1)n^2 + n(n-1) = K_1 - 2n$, a contradiction with $\sum_{(x,y)\in H} m^1_{xy} \geq K_1$. So (G_0, K_0) must admit the answer "yes", and this completes the proof. □

For the next results relative to the complexities of the problems $P_m(\mathcal{Z})$, for appropriate sets $\mathcal{Z}$, we keep only the range of the values of m (expressed with the help of the usual function θ) for which the problems $P_m(\mathcal{Z})$ are NP-hard, even if more accurate results may be derived from the theorems dealing with the complexities of the problems $Q_0(\mathcal{Z})$ or $Q_1(\mathcal{Z})$.

For $P_m(\mathcal{L})$ with odd values of m, we may derive the next corollary from Theorem 10:

Corollary 13 *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^2)$ such that, for $m \geq \zeta(n)$ and with m taking odd values, $P_m(\mathcal{L})$ is NP-hard.*

Proof We know from Sect. 3 that we may construct, in polynomial time, a profile Π of $N(G)$ linear orders associated with the graph G of the proof of Theorem 10, with $N(G) = \theta(n^2)$. Set $\zeta(n) = N(G)$. With the help of this construction, it is trivial to polynomialy transform $Q_1(\mathcal{L})$ into the decision problem associated with $P_{\zeta(n)}(\mathcal{L})$. The NP-hardness of $P_{\zeta(n)}(\mathcal{L})$ follows, and yields, by Lemma 3, the one of $P_m(\mathcal{L})$ for $m \geq \zeta(n)$ and with m odd. □

As more generally the transformation from a problem $Q_0(\mathcal{Z})$ or $Q_1(\mathcal{Z})$, for some appropriate set $\mathcal{Z}$, to the corresponding problem $P_m(\mathcal{Z})$ with m even or odd is always the same as in Corollary 13, with a number of linear orders m greater than or equal to $N(G)$, we shall shorten the proofs of similar corollaries in the sequel.

Similar results may be obtained for $P_m(\mathcal{A})$ when m is odd. We first consider $Q_1(\mathcal{A})$.

Theorem 14 *Problem $Q_1(\mathcal{A})$ is NP-complete, even if all the weights involved in the instances of $Q_1(\mathcal{A})$ belong to $\{-1, 1\}$.*

Proof We transform any instance $(T = (Y, W), h)$ of FAST into an instance (G, K) of $Q_1(\mathcal{A})$ as in the proof of Theorem 10, except for K, which we set equal to $n(n-1)/2 - h$. The transformation remains polynomial.

It keeps also the answer.

Indeed, if (T, h) admits the answer "yes", then there exist $W' \subset W$ with $|W'| \leq h$ and such that $(Y, W - W')$ is without any circuit. Then we may consider the same graph $(X, W - W')$ (remember that X is equal to Y) as an induced subgraph without circuit of G. Moreover, the weight of $(X, W - W')$ is equal to $n(n-1)/2 - |W'|$, which is greater than or equal to K since $|W'|$ is less than or equal to h.

Conversely, if (G, K) admits the answer "yes", then there exist $A \subseteq U_X$ with a weight greater than or equal to K such that (X, A) is without circuit. Then, by removing from (X, A) all the arcs (if any) with a weight equal to -1 (such an arc does not belong to W), we get another graph (X, A') with $A' \subseteq W$ and still with a weight greater than or equal to K. Then it is easy to show that $W' = W - A'$ gives the answer "yes" to (T, h). □

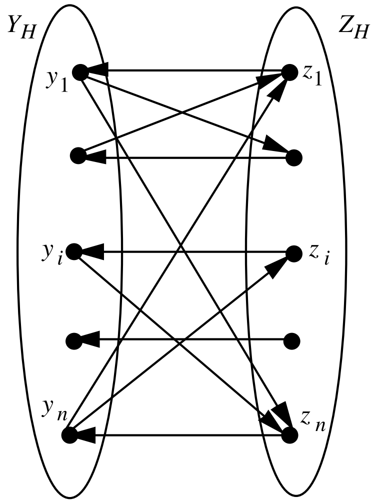


Fig. 13 the graph H of an instance of BFAS

Corollary 15

1. *Problem $Q_0(\mathcal{A})$ is NP-complete, even if all the absolute values of the weights involved in the instances of $Q_0(\mathcal{A})$ are upper bounded by* 2.
2. *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^2)$ such that, for $m \geq \zeta(n)$ and with m taking odd values, $P_m(\mathcal{A})$ is NP-hard.*

Proof

1. It is sufficient to apply the transformation described in Remark 11 to show the NP-completeness of $Q_0(\mathcal{A})$ from the one of $Q_1(\mathcal{A})$.
2. The proof is quite similar to the one of Corollary 13 (see the comment after the proof of Corollary 13). As the positive weights involved in the proof of the NP-completeness of $Q_1(\mathcal{A})$ are equal to 1, we get that $\zeta(n) = \theta(n^2)$ is sufficient to yield the NP-hardness of $P_{\zeta(n)}(\mathcal{A})$. The NP-hardness of $P_m(\mathcal{A})$ for $m \geq \zeta(n)$ follows, from Lemma 3. □

To study the complexity of the problems $Q_0(\mathcal{Z})$ and $Q_1(\mathcal{Z})$ for $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$, we use a variant of FAS, that we call BFAS.

Problem BFAS

Instance: a directed, asymmetric, and bipartite graph $H = (Y_H \cup Z_H, W_1 \cup W_2)$ where $Y_H = \{y_i : 1 \leq i \leq |Y_H|\}$ and $Z_H = \{z_i : 1 \leq i \leq |Z_H| = |Y_H|\}$ give the two classes of H and with $W_1 = \{(z_i, y_i)$ for $1 \leq i \leq |Y_H|\}$ and $W_2 \subset \{(y_i, z_j)$ for $1 \leq i \leq |Y_H|$ and $1 \leq j \leq |Y_H|, i \neq j\}$, and with $|W_1 \cup W_2| = O(|Y_H \cup Z_H|)$; an integer h;
Question: does there exist $W' \subset W_1$ with $|W'| \leq h$ and such that removing the elements of W' from H leaves a graph $(Y_H \cup Z_H,\ (W_1 - W') \cup W_2)$, without any circuit?

Figure 13 shows how such a graph H looks like.

Theorem 16 *Problem BFAS is NP-complete.*

Proof It is easy to show that BFAS belongs to NP (details are left to the reader). To prove that it is NP-complete, we transform into BFAS a variant of the well-known problem called Vertex Cover, variant in which the considered graph is assumed to be planar:

Problem Planar Vertex Cover (PVC)

Instance: an undirected planar graph $G = (X_G, U)$; an integer g;
Question: does there exist $X' \subseteq X_G$ with $|X'| \leq g$ and verifying the following property: $\forall \{x, y\} \in U$, $x \in X'$ or $y \in X'$ (X' is then a *vertex cover* of G of cardinality at most g)?

It is known that VC is NP-complete for general graphs (Karp 1972) and remains so even when restricted to planar graphs (Garey and Johnson 1977). Let (G, g) be any instance of PVC. We define an instance (H, h) of BFAS as follows, with $p = |X_G|$:

- for any vertex $x_i \in X_G$ $(1 \leq i \leq p)$, we create two vertices of H: y_i and z_i $(1 \leq i \leq p)$ and we set $Y_H = \{y_i,\ 1 \leq i \leq p\}$ and $Z_H = \{z_i,\ 1 \leq i \leq p\}$;
- for any edge $\{x_i, x_j\}$ of H, we create two arcs of H: (y_i, z_j) and (y_j, z_i); W_2 will denote the set of these arcs: $W_2 = \{(y_i, z_j), (y_j, z_i)$, for i and j such that $\{x_i, x_j\}$ belongs to $U\}$;
- we complete H by adding all the arcs of the form (z_i, y_i) for $1 \leq i \leq p$; they constitute the set $W_1 : W_1 = \{(z_i, y_i)$ for $1 \leq i \leq p\}$;
- we set $g = h$.

Notice that the number of vertices of H is equal to twice the number of vertices of G, and that the number of edges of H is equal to twice the number of edges of G plus the number of vertices of G. As G is assumed to be planar, the number of edges of G is upper-bounded by three times the number of vertices of G (see Diestel 2005 for instance). Hence the relation $|W_1 \cup W_2| = \mathrm{O}(|Y_H \cup Z_H|)$.

Then we claim that G admits a vertex cover of cardinality at most g if and only if H admits a feedback arc set included into W_1 of cardinality at most h.

Indeed, assume that there exists a vertex cover X' of G with $|X'| \leq g$. Then let W' be defined by: $W' = \{(z_i, y_i)$ for $x_i \in X'\}$. We clearly have $W' \subseteq W_1$ and $|W'| = |X'| \leq g = h$. Moreover, assume that there exists a circuit in the graph $(Y_H \cup Z_H, (W_1 - W') \cup W_2)$. Then this circuit necessarily goes through an arc (z_i, y_i) for some i such that x_i does not belong to X' and then goes through an arc (y_i, z_j) for an appropriate $j\, (1 \leq j \leq p)$; the only way to go on the circuit is to follow the arc (z_j, y_j) (it is the only arc with z_j as its tail), which involves that x_j does not belong to X'. But, as the arc (y_i, z_j) exists in H, $\{x_i, x_j\}$ must be an edge of G, and this edge is not covered by X', a contradiction.

Conversely, assume that (H, h) admits a subset W' of W_1 which is a feedback arc set of cardinality at most h. Define X' as the set of vertices of G associated with the elements of $W' : X' = \{x_i$ for $(z_i, y_i) \in W'\}$. Then obviously: $|X'| \leq g$. Moreover, assume that X' is not a vertex cover of G. It means that there exists an edge $\{x_i, x_j\}$ of G with $x_i \notin X'$ and $x_j \notin X'$. So, similarly, we have in $H : (z_i, y_i) \notin W'$ and $(z_j, y_j) \notin W'$. But in these conditions, the arcs (z_i, y_i), (y_i, z_j), (z_j, y_j), and (y_j, z_i) (these four arcs do exist in H) define a circuit in the graph $(Y_H \cup Z_H, (W_1 - W') \cup W_2)$, and W' is not a feedback arc set, a contradiction.

So the proposed transformation keeps the answer. As it is trivially polynomial with respect to the size of the transformed instance (G, g), BFAS is NP-complete. □

We now study the complexity of the problems $Q_0(\mathcal{Z})$ for $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$.

Theorem 17 *For $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$, $Q_0(\mathcal{Z})$ is NP-complete.*

Proof Let $\mathcal{Z}$ belong to $\{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$. As for the other problems above, $Q_0(\mathcal{Z})$ obviously belongs to NP. We transform BFAS into $Q_0(\mathcal{Z})$. Let $(H = (Y_H \cup Z_H, W_1 \cup W_2), h)$ be any instance of BFAS, with the same notations as above (with $p = |Y_H| = |Z_H|$); in particular, we have $|W_1 \cup W_2| = O(p)$. Let (G, K) be the instance of $Q_0(\mathcal{Z})$ defined by:

- the vertex set of G is $X = Y_H \cup Z_H$ (then we have $|X| = n = 2p$);
- the arc set of G is U_X;
- for any arc (x, y) of G, the weight m_{xy} of (x, y) is equal to: 2 if $(x, y) \in W_1$, -2 if $(y, x) \in W_1$, $4p - 2$ if $(x, y) \in W_2$, $-(4p - 2)$ if $(y, x) \in W_2$, 0 otherwise;
- $K = (4p - 2)|W_2| + 2p - 4h$.

Notice that G is well defined since H is asymmetric, and that all the weights are even. Moreover, the transformation is clearly polynomial with respect to the size of the instance (H, h) of BFAS.

Now, assume that the instance (H, h) of BFAS admits the answer "yes": there exists $W' \subset W_1$ with $|W'| \leq h$ and such that the graph $(X, (W_1 - W') \cup W_2)$ is without any circuit. We prove that the instance (G, K) admits also the answer "yes". If we consider $(X, (W_1 - W') \cup W_2)$ as a subgraph of G, its weight is $2|W_1 - W'| + (4p - 2)|W_2|$. By Lemma 4b, we may complete $(X, (W_1 - W') \cup W_2)$ into a linear order (X, L) (that we shall consider as an interval order if $\mathcal{Z}$ is $\mathcal{I}$, or as a semiorder if $\mathcal{Z}$ is $\mathcal{S}$, or as a complete preorder if $\mathcal{Z}$ is $\mathcal{C}$) by adding appropriate arcs. Among these extra arcs, there are at most $|W'|$ arcs (x, y) such that (y, x) belongs to W_1, i.e. at most $|W'|$ arcs with a weight equal to -2, while the other extra arcs all belong to $U_X - (W_1 \cup W_2)$ and have a weight equal to 0. More precisely, the weights of the arcs of L belonging to W_1 are equal to 2; there are at least $|W_1 - W'|$ such arcs. The weights of the $|W_2|$ arcs of L belonging to W_2 are equal to $4p - 2$. The weights of the arcs (x, y) of L such that (y, x) belongs to W_1 are equal to -2; there are at most $|W'|$ such arcs. The other arcs of L have a weight equal to 0. Hence, since $|W_1| = p$ and $W' \subseteq W_1$:

$$\sum_{(x,y)\in L} m_{xy} \geq 2|W_1 - W'| + (4p - 2)|W_2| - 2|W'| = 2p + (4p - 2)|W_2| - 4|W'|$$

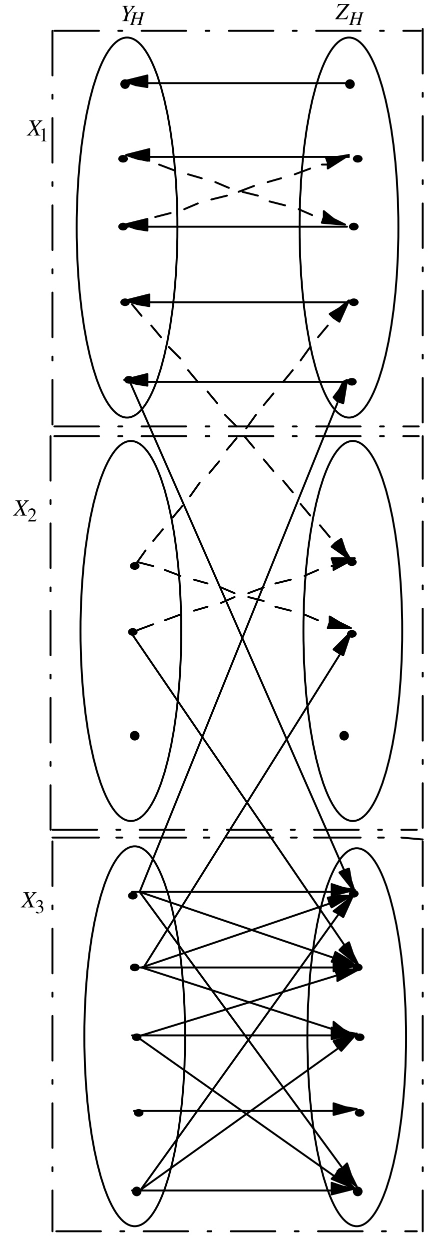
and, since $|W'| \leq h$: $\sum_{(x,y)\in L} m_{xy} \geq 2p + (4p - 2)|W_2| - 4h = K$. So (G, K) admits also the answer "yes".

Conversely, assume that (G, K) admits the answer "yes". We consider two main subcases: $\mathcal{Z} \in \{\mathcal{I}, \mathcal{S}\}$ or $\mathcal{Z} = \mathcal{C}$.

4.1 1st subcase: $\mathcal{Z} \in \{\mathcal{I}, \mathcal{S}\}$

Since a semiorder is an interval order, there exists $I \subseteq U_X$ with $\sum_{(x,y)\in I} m_{xy} \geq K$ and such that (X, I) is an interval order. We want to show that then the instance (H, h) of BFAS also admits the answer "yes". Notice that if h is greater than or equal to p, the answer of (H, h) is trivially "yes"; so, assume that we have $h \leq p - 1$. Let us show that I contains all the arcs of W_2. Assume the contrary. The arcs of I with a non-negative weight would be at most the p elements of W_1 (with a weight equal to 2) and at most $|W_2| - 1$ arcs of W_2 (with a weight equal to $4p - 2$). So we would get: $\sum_{(x,y)\in I} m_{xy} \leq 2p + (4p - 2)(|W_2| - 1)$. On the other hand, we are supposed to have: $\sum_{(x,y)\in I} m_{xy} \geq K = (4p - 2)|W_2| + 2p - 4h$. From these two inequalities, we draw $4h \geq 4p - 2$, which is incompatible with $h \leq p - 1$. Hence: $W_2 \subseteq I$ and, because of the antisymmetry of I, there is no arc in I of the form (z_i, y_j) with $i \neq j$. We prove now that we may construct a linear order L with $\sum_{(x,y)\in L} m_{xy} \geq K$ from I. For this, set $J = I - \{(x, y) \in I \text{ with } m_{xy} = 0\}$, and gather the vertices of X into the following three sets:

Fig. 14 The graph induced by J for $Q_0(\mathcal{Z})$ when $\mathcal{Z}$ is equal to $\mathcal{I}$ or $\mathcal{S}$



- $X_1 = \{y_k \in Y_H, z_k \in Z_H$ such that $(z_k, y_k) \in J\}$
- $X_2 = \{y_k \in Y_H, z_k \in Z_H$ such that $(y_k, z_k) \notin J$ and $(z_k, y_k) \notin J\}$
- $X_3 = \{y_k \in Y_H, z_k \in Z_H$ such that $(y_k, z_k) \in J\}$.

The situation is illustrated by Fig. 14. We are going to show that the dashed arcs of Fig. 14 do not exist in fact. Notice that, as a subset of I which contains no circuit, J contains no circuit.

The dashed arcs with their two extremities inside X_1 cannot exist, otherwise there would exist a circuit in J. Now consider an arc (y_j, z_i) with $y_j \in Y_H$, $z_i \in Z_H$ (thus $i \neq j$) and with an extremity inside X_1 and the other inside X_2. As one extremity belongs to X_1, the arc (z_i, y_i) or the arc (z_j, y_j) exists in I, and thus in J since its weight is not equal to 0. Also, by construction of G, (y_i, z_j) is an arc of G, and thus of I ($W_2 \subseteq I$). Assume without lack of generality that (z_i, y_i) belongs to I (and thus to J); then y_i and z_i belong to X_1, while y_j and z_j belong to X_2. In this case, as I is transitive, the arcs (y_j, z_i), (z_i, y_i), (y_i, z_j) involve the existence of the arc (y_j, z_j), a contradiction with the belonging of y_j and z_j to X_2. Last, the dashed arcs with their two extremities inside X_2 cannot exist in I. Indeed, assume that such a pair of arcs (y_j, z_i) and (y_i, z_j) exist in I, and thus in J, with $y_i \in Y_H \cap X_2$, $y_j \in Y_H \cap X_2$, $z_i \in Z_H \cap X_2$, and $z_j \in Z_H \cap X_2$ ($i \neq j$) (notice that if one of these two arcs exists in J, the other one must exist too). As y_i, y_j, z_i, and z_j belong to X_2, the arcs (y_i, z_i) and (y_j, z_j) do not belong to J, and thus do not belong to I. But then the arcs (y_j, z_i) and (y_i, z_j) of I do not respect the definition of an interval order. So the look of J is as the one shown by Fig. 14 without the dashed arcs.

Now, set $J' = J \cup \{(z, y) \in W_1$ for $y \in X_2$ and $z \in X_2\}$ (with respect to Fig. 14, we add all the horizontal arcs from the right to the left with their two extremities in X_2). As the vertices of X_2 are linked only with vertices of X_3, it is easy to see that J' is still without any circuit. As the weights of these arcs are positive, then: $\sum_{(x,y)\in J'} m_{xy} \geq \sum_{(x,y)\in J} m_{xy} = \sum_{(x,y)\in I} m_{xy} \geq K$. As J' is without circuit, by Lemma 4b, we may extend J' into a linear order L. As we have $W_2 \subseteq I$ and as, for any index k with $1 \leq k \leq n$, y_i and z_i are now linked by an arc belonging to J', all the arcs that we add in order to define L from J' have a weight equal to 0. Hence $\sum_{(x,y)\in L} m_{xy} = \sum_{(x,y)\in J'} m_{xy} \geq K$. Now consider the set $W' = (W_1 \cup W_2) - (W_1 \cup W_2) \cap L$: W' is the set of the arcs of H which have been removed from G in order to get the linear order (X, L); by Lemma 4a, the graph induced by $(W_1 \cup W_2) - W'$, i.e. by $(W_1 \cup W_2) \cap L$, is acyclic because (X, L) is circuitless. Moreover, we have:

$$|W'| = |(W_1 \cup W_2) - (W_1 \cup W_2) \cap L| = |(W_1 \cup W_2)| - |(W_1 \cup W_2) \cap L|.$$

Suppose that we have $|W'| > h$. There are $|(W_1 \cup W_2)| - W'|$ arcs which belong simultaneously to $W_1 \cup W_2$ and to L. As the weights of the other arcs are non-positive, we get:

$$\sum_{(x,y)\in L} m_{xy} \leq 2|(W_1 \cup W_2) - W'| = 2|W_1 \cup W_2| - 2|W'| < 2|W_1 \cup W_2| - 2h = K,$$

a contradiction. It involves that W' satisfies all the conditions and the answer admitted by (H, h) is "yes". This completes the proof for the subcase $\mathcal{Z} \in \{\mathcal{I}, \mathcal{S}\}$.

4.2 2nd subcase: $\mathcal{Z} = \mathcal{C}$

As for the previous case, we are going to prove that, if the answer admitted by the instance (G, K) is "yes", then we can build a linear order which gives this answer "yes". Then the conclusion will be the same as above.

So, assume that there exists a subset C of U_X such that (X, C) is a complete preorder with $\sum_{(x,y)\in C} m_{xy} \geq K$. As above, this inequality involves that C contains all the elements of W_2 and no arc (x, y) such that (y, x) would belong to W_2 (details are left to the reader). Let D be the set made of the arcs of C with a non-zero weight: $D = C - \{(x, y) \in C$ with $m_{xy} = 0\}$. Moreover, gather the vertices of X into the following three sets:

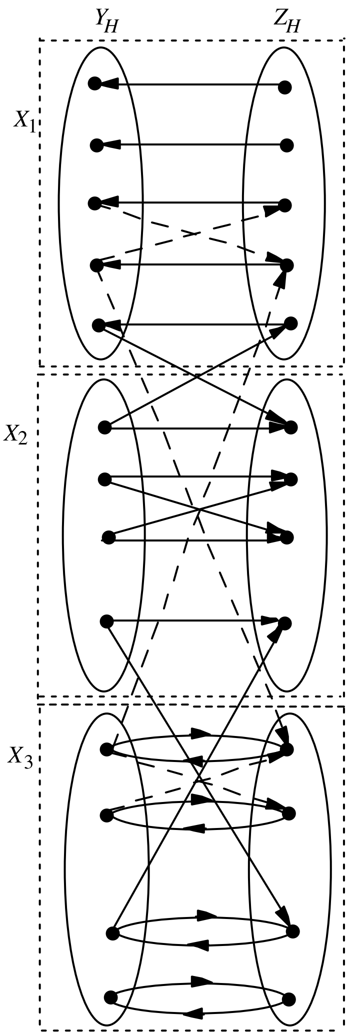


Fig. 15 The graph induced by D for $Q_0(\mathcal{C})$

- $X_1 = \{y_k \in Y_H, z_k \in Z_H$ such that $(z_k, y_k) \in D$ and $(y_k, z_k) \notin D\}$
- $X_2 = \{y_k \in Y_H, z_k \in Z_H$ such that $(y_k, z_k) \in D$ and $(z_k, y_k) \notin D\}$
- $X_3 = \{y_k \in Y_H, z_k \in Z_H$ such that $(y_k, z_k) \in D$ and $(z_k, y_k) \in D\}$.

The look of the graph induced by D is given by Fig. 15. We are going to show that the dashed arcs of Fig. 15 do not exist in D.

Indeed, let (y_j, z_i) be such an arc of D with $z_i \in Z_H$ and $y_j \in Y_H$. Then its weight is not equal to 0, and it is the same for (y_i, z_j), which thus belongs to D (and to C). If we assume that the four vertices y_i, y_j, z_i, and z_j belong to $X_1 \cup X_3$, then the arcs (z_i, y_i) and (z_j, y_j) belong to D, and so to C, and, by transitivity of C, the arcs (z_j, y_i) and (z_i, y_j) also belong to C, and then to D, since the weights of these arcs are not equal to 0 (they are the opposites of the weights of the arcs (y_i, z_j) and (y_j, z_i), which are not equal 0, since these two arcs belong to D), what is impossible, as shown above. So, the look of the graph induced by D is the one depicted by Fig. 15 without the dashed arcs.

The next step consists in showing that we may extract a set D' of arcs from D such that D' is without circuit while its weight $\sum_{(x,y)\in D'} m_{xy}$ is still greater than or equal to K. For this, let D' be defined by $D' = D - \{(y_i, z_i)$ for $y_i \in X_3$, $z_i \in X_3\}$ (in other words, with respect to Fig. 15, we remove the—almost—horizontal arcs inside X_3 and directed from the left to the right). As the removed arcs have a negative weight, we get: $\sum_{(x,y)\in D'} m_{xy} \geq \sum_{(x,y)\in D} m_{xy} \geq K$. Moreover, D' is without circuit. Indeed, consider any circuit in D; it is also a circuit in C, which is transitive. Such a circuit must contain an arc of the form (z_i, y_i) with $y_i \in Y_H$ and $z_i \in Z_H$ (since it is the only way to go from Z_H to Y_H in the graph induced by D). Because of the transitivity of C applied to the considered circuit, (y_i, z_i) must also be an arc of C and, because its weight is not equal to 0, of D. So y_i and z_i must belong to X_3. Hence the removal from D of the arcs (y_i, z_i) with $y_i \in X_3$, $z_i \in X_3$ leaves a graph (induced by D') without any circuit.

We may now conclude. As D' is without any circuit and by Lemma 4b, we may complete it into a linear order L by adding appropriate arcs. As D' already contains W_2 and, for $1 \leq i \leq p$, exactly one of the two arcs (y_i, z_i) or (z_i, y_i), the extra arcs have a weight equal to 0. So, we get: $\sum_{(x,y)\in L} m_{xy} = \sum_{(x,y)\in D'} m_{xy} \geq K$. Then it is sufficient to apply the same argument as in the first subcase to show the existence of a subset W' of $W_1 \cup W_2$ which gives the answer "yes" to the instance (H, h) of BFAS. This completes the proof for the subcase $\mathcal{Z} = \mathcal{C}$. □

Corollary 18 *For $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$, there exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^2)$ such that, for $m \geq \zeta(n)$ and with m taking even values, $P_m(\mathcal{Z})$ is NP-hard.*

Proof As above, we may associate a profile of linear orders to the graph constructed in the proof of Theorem 17. With the same notation as in this proof, the required number of linear orders is at most $\sum_{m_{xy}>0} m_{xy}$. There are at most $\mathrm{O}(n)$ arcs whose weight is not equal to 0. When the weight is not equal to 0, the weight is at most $\mathrm{O}(n)$. Hence the relation: $\sum_{m_{xy}>0} m_{xy} = \mathrm{O}(n^2)$. So, there exists a function $\zeta(n) = \theta(n^2)$ such that $P_{\zeta(n)}(\mathcal{Z})$ is NP-hard. The statement of Corollary 18 follows, by applying Lemma 3. □

Remark 19 If we transpose the proof of Theorem 17 in terms of preferences, we build a profile of linear orders such that there exists an optimal interval order, or an optimal semi-order, or an optimal complete preorder which is in fact a linear order. An interesting open problem would be to characterize the profiles of linear orders for which such a situation always occurs. Notice that it is not necessarily the case when the median relation must just be a partial order, as shown by the example of Fig. 16.

For this graph, it is easy to show that all the arcs with a weight equal to 8 constitute a partial order which is not a linear order (it is not even an interval order); its weight is equal to 48. Consider now any linear order L extracted from this graph. If L does not contain the 6 arcs with a weight equal to 8, its weight is less than or equal to $5 \times 8 + 3 \times 2 = 46$, what is less than 48. If L contains the 6 arcs with a weight equal to 8 and an arc (α, β) with a weight equal to 2 (one "horizontal" arc from the right to the left with respect to Fig. 16), then, because of its transitivity, L must contain the two arcs with a weight equal to -2 which are not (β, α) (what involves that L cannot contain the 6 arcs with a weight equal to 8 and more than one arc with a weight equal to 2), and so its total weight cannot exceed $6 \times 8 + 1 \times 2 + 2 \times (-2) = 46$. If L contains the 6 arcs with a weight equal to 8 and none of the arcs with a weight equal to 2, then, because of the completeness of a linear order, L must contain the 3 arcs with a weight equal to -2, and its total weight cannot exceed $48 - 6 = 42$.

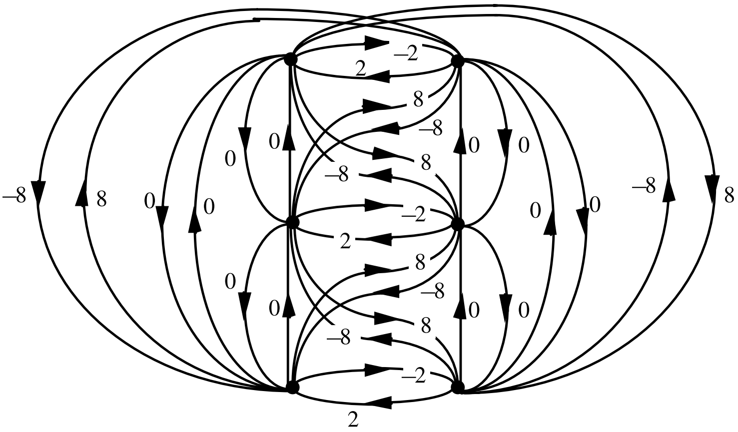


Fig. 16 A graph representing a profile of linear orders and for which the optimal partial order is not a linear order

In all the cases, the total weight of L is less than 48, and therefore no linear order can be an optimal partial order of the graph of Fig. 16.

The next theorem provides similar results when m is odd:

Theorem 20 *For $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$, the problems $Q_1(\mathcal{Z})$ are NP-complete. For $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$, there exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^4)$ such that, for $m \geq \zeta(n)$ and with m taking odd values, $P_m(\mathcal{Z})$ is NP-hard.*

Proof The NP-completeness of $Q_1(\mathcal{Z})$ for $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$ comes from the application of Lemma 12 and from the NP-completeness of $Q_0(\mathcal{Z})$ for the same sets (Theorem 17).

In the transformation involved in this application of Lemma 12, the weights are, broadly speaking, multiplied by $2n^2$. So, now, the sums of the weights of the graphs involved in the instances of $Q_1(\mathcal{Z})$ are upper-bounded by a function $\zeta(n)$ taking odd values and with $\zeta(n) = \theta(n^4)$ (instead of $\theta(n^2)$ in Corollary 18). Hence the NP-hardness of $P_{\zeta(n)}(\mathcal{Z})$ for $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{S}\}$, and then of $P_m(\mathcal{Z})$ for $m \geq \zeta(n)$ and m odd by Lemma 3. □

To study the complexity of the problems $P_m(\mathcal{B})$, $P_m(\mathcal{J})$, $P_m(\mathcal{Q})$ and $P_m(\mathcal{W})$, we first prove a lemma. In order to state it, we introduce a new notation. For any set $\mathcal{Z}$ of binary relations defined by some properties, we define $\mathcal{Z}^a$ as the set of preferences which are the asymmetric part of a preference belonging to $\mathcal{Z}$. In particular, we have $\mathcal{B}^a = \mathcal{A}$, $\mathcal{C}^a = \mathcal{W}$, $\mathcal{J}^a = \mathcal{I}$, $\mathcal{Q}^a = \mathcal{S}$, and $\mathcal{P}^a = \mathcal{O}$.

Lemma 21 *For any set $\mathcal{Z}$ and any integer m, $P_m(\mathcal{Z})$ and $P_m(\mathcal{Z}^a)$ have the same complexity.*

Proof The result comes from the fact that we have $\Delta(\Pi, Z) = \Delta(\Pi, Z^a)$, for any profile Π of linear orders and any element Z of $\mathcal{Z}$. □

Corollary 22

1. *For $m \geq 4$ and with m taking even values, $P_m(\mathcal{B})$ is NP-hard.*
2. *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^2)$ such that, for $m \geq \zeta(n)$ and with m taking odd values, $P_m(\mathcal{B})$ is NP-hard.*

3. *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^2)$ such that, for $m \geq \zeta(n)$ and with m taking even values, $P_m(\mathcal{J})$, $P_m(\mathcal{Q})$ and $P_m(\mathcal{W})$ are NP-hard.*
4. *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^4)$ such that, for $m \geq \zeta(n)$ and with m taking odd values, $P_m(\mathcal{J})$, $P_m(\mathcal{Q})$ and $P_m(\mathcal{W})$ are NP-hard.*

Proof These results come as consequences of Theorem 8, Corollary 15, Corollary 18 and Theorem 20, and from the application of Lemma 21 to $\mathcal{Z} = \mathcal{B}$, $\mathcal{Z} = \mathcal{J}$, $\mathcal{Z} = \mathcal{Q}$ and to $\mathcal{Z} = \mathcal{C}$. □

The last result of this section deals with any set $\mathcal{Y}$ containing $\mathcal{L}$.

Theorem 23

1. *For any even integer $m \geq 4$, for any set $\mathcal{Y}$ with $\mathcal{L} \subseteq \mathcal{Y}$, for $\mathcal{Z} \in \{\mathcal{A}, \mathcal{B}, \mathcal{L}\}$, $P_m(\mathcal{Y}, \mathcal{Z})$ is NP-hard.*
2. *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^2)$ such that, for $m \geq \zeta(n)$ and with m taking odd values, for any set $\mathcal{Y}$ with $\mathcal{L} \subseteq \mathcal{Y}$, for $\mathcal{Z} \in \{\mathcal{A}, \mathcal{B}, \mathcal{L}\}$, $P_m(\mathcal{Y}, \mathcal{Z})$ is NP-hard.*
3. *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^2)$ such that, for $m \geq \zeta(n)$ and with m taking even values, for any set $\mathcal{Y}$ with $\mathcal{L} \subseteq \mathcal{Y}$, for $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{Q}, \mathcal{S}, \mathcal{W}\}$, $P_m(\mathcal{Y}, \mathcal{Z})$ is NP-hard.*
4. *There exists a function $\zeta(n)$ with $\zeta(n) = \theta(n^4)$ such that, for $m \geq \zeta(n)$ and with m taking odd values, for any set $\mathcal{Y}$ with $\mathcal{L} \subseteq \mathcal{Y}$, for $\mathcal{Z} \in \{\mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{Q}, \mathcal{S}, \mathcal{W}\}$, $P_m(\mathcal{Y}, \mathcal{Z})$ is NP-hard.*

Proof The previous results give the statement of Theorem 23 for $\mathcal{Y} = \mathcal{L}$. For $\mathcal{L} \subset \mathcal{Y}$, it is sufficient to consider any instance of the NP-hard problem $P_m(\mathcal{Z})$ (that is, $P_m(\mathcal{L}, \mathcal{Z})$) as an instance of $P_m(\mathcal{Y}, \mathcal{Z})$. This transformation (the identity !) is obviously polynomial and keeps the answer. Hence the result. □

In particular, we may apply Theorem 23 when $\mathcal{Y}$ is any one of the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}$, or $\mathcal{W}$, but also to "mixed" profiles belonging to any union of two or more sets $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{I}, \mathcal{J}, \mathcal{L}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}$, or $\mathcal{W}$, for instance to profiles which may contain tournaments, preorders, and interval orders simultaneously.

5 Conclusion

The previous section was devoted to NP-hard problems. There are also some problems $P_m(\mathcal{Y}, \mathcal{Z})$ which are polynomial. It is trivially the case for $P_m(\mathcal{Y}, \mathcal{R})$ and for $P_m(\mathcal{Y}, \mathcal{T})$, for any set $\mathcal{Y}$ and any integer m. Indeed, if we consider the associated graph theoretic problems (similar to the problems Q_0 and Q_1 defined in Sect. 3) but with respect to a profile of relations all belonging to $\mathcal{Y}$ (then the relation $m_{xy} = -m_{yx}$ is no longer necessarily true), it is easy to see that an optimal solution consists in keeping all the arcs of G with a positive weight for $\mathcal{Z} = \mathcal{R}$, or in keeping, for each pair of arcs (x, y) and (y, x), the arc with the greatest weight for $\mathcal{Z} = \mathcal{T}$ (see for instance Hudry et al. 2006). Another polynomial case, quite interesting, is the one of the so-called *unimodular orders*; in this case, the aggregation of unimodular orders into a unimodular order is polynomial (see Black 1948).

Anyway, it seems that properties like transitivity or the lack of circuits usually lead to NP-hard problems. The previous complexity results illustrate this trend. We may summarize them by the table of Fig. 17. In this table, the letter "P" means that $P_m(\mathcal{Y}, \mathcal{Z})$ is (trivially) polynomial, while "NPH" means that the considered problem $P_m(\mathcal{Y}, \mathcal{Z})$ is NP-hard. In this case, we indicate the range of a lower bound of the number m of relations inside the profile

Median relation ($\mathcal{Z}$)	$\Pi \in \mathcal{Y}^m$ with $\mathcal{L} \subseteq \mathcal{Y}$, m odd with $m \geq \zeta(n)$	$\Pi \in \mathcal{Y}^m$ with $\mathcal{L} \subseteq \mathcal{Y}$, m even with $m \geq \zeta(n)$
acyclic relation ($\mathcal{A}$)	NPH, $\zeta(n) = \theta(n^2)$	NPH, $\zeta(n) = 4$
asymmetric part: acyclic ($\mathcal{B}$)	NPH, $\zeta(n) = \theta(n^2)$	NPH, $\zeta(n) = 4$
complete preorder ($\mathcal{C}$)	NPH, $\zeta(n) = \theta(n^4)$	NPH, $\zeta(n) = \theta(n^2)$
interval order ($\mathcal{I}$)	NPH, $\zeta(n) = \theta(n^4)$	NPH, $\zeta(n) = \theta(n^2)$
asymmetric part: interval order ($\mathcal{J}$)	NPH, $\zeta(n) = \theta(n^4)$	NPH, $\zeta(n) = \theta(n^2)$
linear order ($\mathcal{L}$)	NPH, $\zeta(n) = \theta(n^2)$	NPH, $\zeta(n) = 4$
partial order ($\mathcal{O}$)	?	?
preorder ($\mathcal{P}$)	?	?
quasi-order ($\mathcal{Q}$)	NPH, $\zeta(n) = \theta(n^4)$	NPH, $\zeta(n) = \theta(n^2)$
binary relation ($\mathcal{R}$)	P	P
semiorders ($\mathcal{S}$)	NPH, $\zeta(n) = \theta(n^4)$	NPH, $\zeta(n) = \theta(n^2)$
tournament ($\mathcal{T}$)	P	P
weak order ($\mathcal{W}$)	NPH, $\zeta(n) = \theta(n^4)$	NPH, $\zeta(n) = \theta(n^2)$

Fig. 17 Complexity of the problems $P_m(\mathcal{Y}, \mathcal{Z})$

which ensures that $P_m(\mathcal{Y}, \mathcal{Z})$ is NP-hard; for instance, "NPH, $\zeta(n) = \theta(n^2)$" in the row associated with $\mathcal{A}$ and with the column "m odd with $m \geq \zeta(n)$" means that there exists a constant λ such that, for $m \geq \lambda n^2$ with m odd, the problem $P_m(\mathcal{Y}, \mathcal{A})$ is NP-hard.

To my knowledge, when not trivial, the complexity for lower values of m remains unknown. In particular, it would be interesting to know whether some of the problems $P_m(\mathcal{Y}, \mathcal{Z})$ remain NP-hard if m is a given constant. The result by Dwork et al. (2001) shows that it is the case for $P_m(\mathcal{Y}, \mathcal{L})$ for any set $\mathcal{Y}$ containing $\mathcal{L}$ and when m is an even constant greater than or equal to 4 (and 4 is the lowest possible value if $\mathcal{Y}$ is equal to $\mathcal{L}$). It is also the case, still for m even and greater than or equal to 4, for $\mathcal{Z} = \mathcal{A}$ or $\mathcal{Z} = \mathcal{B}$. It would be interesting to decide whether it is still the case for fixed values of m with m odd or for other sets $\mathcal{Z}$.

As a general result, remember that all the results displayed in the table of Fig. 17 remain the same if we add the reflexivity or the irreflexivity to the considered types of relations.

In this table, we see that the lower bounds of m are not the same when m is odd or when it is even: in this case, the ratio between these bounds, for a given structure $\mathcal{Z}$, is n^2. As noticed above, this comes from the fact that, for m even, we may apply 0 as a weight of the arcs of the graph representing the profile (which means that there is a tie between the two candidates represented by the extremities of this arc); when m is odd, we replace 0 by 1 or -1 and, broadly speaking, we multiply the other values by $2n^2$ to compensate. Hence the ratio n^2 between these two cases.

Last, we may see that some cases are not solved, marked by "?" in Fig. 17: it is when the median relation must be a partial order or a preorder. The structure of partial order is more constrained than the structure of binary relation, but less than the one of linear order or of interval order for instance. With this respect, it is a kind of intermediary structure, and maybe the complexity for finding a median partial order is also between the polynomiality of the search of a median binary relation and the NP-hardness of the search of a median linear order. Nevertheless, I conjecture that the search of a median partial order or of a median preorder is also NP-hard (it is the case for profiles of binary relations, as shown by Wakabayashi 1986, when we have $m \geq 2$ for a median partial order and $m \geq \zeta(n)$ with $\zeta(n) = \theta(n^6)$ for a median preorder). Because of Lemma 21, notice that it is sufficient to prove that one is NP-hard in order to prove that both are. But such a proof remains to be found.

Acknowledgements I would like to thank Alexis Tsoukiàs for his great patience, and the anonymous referees for their valuable comments. They helped me very much in improving this paper, for its presentation as well as for its content.

References

Ailon, N., Charikar, M., & Newman, A. (2005). Aggregating inconsistent information: ranking and clustering. In *Proceedings of the 37th annual ACM symposium on Theory of computing (STOC)* (pp. 684–693).

Alon, N. (2006). Ranking tournaments. *SIAM Journal on Discrete Mathematics*, *20*(1), 137–142. doi:10.1137/050623905

Arditti, D. (1984). Un nouvel algorithme de recherche d'un ordre induit par des comparaisons par paires. In E. Diday et al. (Eds.), *Data analysis and informatics III* (pp. 323–343). Amsterdam: North-Holland.

Barthélemy, J.-P. (1979). Caractérisations axiomatiques de la distance de la différence symétrique entre des relations binaires. *Mathématiques et Sciences Humaines*, *67*, 85–113.

Barthélemy, J.-P., & Leclerc, B. (1995). The median procedure for partitions. In I. J. Cox, P. Hansen, & B. Julesz (Eds.), *DIMACS series in discrete mathematics and theoretical computer science: Vol. 19. Partitioning data sets* (pp. 3–34). Providence: Am. Math. Soc.

Barthélemy, J.-P., & Monjardet, B. (1981). The median procedure in cluster analysis and social choice theory. *Mathematical Social Sciences*, *1*, 235–267.

Barthélemy, J.-P., & Monjardet, B. (1988). The median procedure in data analysis: new results and open problems. In H. H. Bock (Ed.), *Classification and related methods of data analysis*. Amsterdam: North-Holland.

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

Barthélemy, J.-P., Flament, C., & Monjardet, B. (1989). Ordered sets and social sciences. In I. Rival (Ed.), *Ordered sets* (pp. 721–758). Dordrecht: Reidel.

Barthélemy, J.-P., Hudry, O., Isaak, G., Roberts, F. S., & Tesman, B. (1995). The reversing number of a digraph. *Discrete Applied Mathematics*, *60*, 39–76.

Barthélemy, J.-P., Cohen, G., & Lobstein, A. (1996). *Algorithmic complexity and communication problems*. London: UCL Press.

Bartholdi, III J. J., Tovey, C. A., & Trick, M. A. (1989). Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, *6*, 157–165.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, *56*, 23–34.

Bogart, K. P. (1973). Preference structures I: distance between transitive preference relations. *Journal of Mathematical Sociology*, *3*, 49–67.

Bogart, K. P. (1975). Preference structures II: distances between asymmetric relations. *SIAM Journal of Applied Mathematics*, *29*(2), 254–262.

Caspard, N., Leclerc, B., & Monjardet, B. (2007). *Ensembles ordonnés finis: concepts, résultats et usages*. Berlin: Springer.

Charbit, P., Thomasse, S., & Yeo, A. (2007). The minimum feedback arc set problem is NP-hard for tournaments. *Combinatorics, Probability and Computing*, *16*(1), 1–4.

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

Conitzer, V. (2006). Computing Slater rankings using similarities among candidates. In *Proceedings of the 21st national conference on artificial intelligence* (*AAAI-06*) (pp. 613–619), Boston, USA.

de Cani, J. S. (1969). Maximum likelihood paired comparison ranking by linear programming. *Biometrika*, *3*, 537–545.

Debord, B. (1987). Caractérisation des matrices de préférences nettes et méthodes d'agrégation associées. *Mathématiques et Sciences Humaines*, *97*, 5–17.

Diestel, R. (2005). *Graph theory* (3rd edn.). Berlin: Springer.

Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the Web. *Proceedings of the 10th international conference on World Wide Web* (*WWW10*) (pp. 613–622).

Fishburn, P. C. (1985). *Interval orders and interval graphs, a study of partially ordered sets*. New York: Wiley.

Garey, M. R., & Johnson, D. S. (1977). The rectilinear Steiner tree problem is NP-complete. *SIAM Journal on Applied Mathematics*, *32*, 826–834.

Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability, a guide to the theory of NP-completeness*. New York: Freeman.

Guénoche, A., Vandeputte-Riboud, B., & Denis, J.-B. (1994). Selecting varieties using a series of trials and a combinatorial ordering method. *Agronomie*, *14*, 363–375.
Hubert, L. (1976). Seriation using asymmetric proximity measures. *British Journal of Mathematical & Statistical Psychology*, *29*, 32–52.
Hudry, O. (1989). *Recherche d'ordres médians: complexité, algorithmique et problèmes combinatoires*. Thèse de l'ENST, Paris.
Hudry, O. (2004). Computation of median orders: complexity results. In *Annales du LAMSADE: Vol. 3. Proceedings of the workshop on computer science and decision theory* (pp. 179–214), DIMACS.
Hudry, O. (2008, to appear). Complexity of voting procedures. In R. Meyers (Ed.), *Encyclopedia of complexity and systems science*. Berlin: Springer.
Hudry, O., Leclerc, B., Monjardet, B., & Barthélemy, J.-P. (2006). Médianes métriques et laticielles. In D. Dubois, M. Pirlot, & D. Bouyssou (Eds.), *Concepts et méthodes pour l'aide à la décision* (pp. 271–316). Paris: Hermès (translation to appear).
Jünger, M. (1985). *Polyhedral combinatorics and the acyclic subdigraph problem*. Berlin: Heldermann.
Karp, R. M. (1972). Reducibility among combinatorial problems. In R. E. Miller & J. W. Tatcher (Eds.), *Complexity of computer computations* (pp. 85–103). New York: Plenum.
Krivanek, M., & Moravek, J. (1986). NP-hard problems in hierarchical-tree clustering. *Acta Informatica*, *23*, 311–323.
Monjardet, B. (1973). Tournois et ordres médians pour une opinion. *Mathématiques et Sciences Humaines*, *43*, 55–73.
Monjardet, B. (1978). Axiomatiques et propriétés des quasi-ordres. *Mathématiques et Sciences Humaines*, *63*, 51–82.
Monjardet, B. (1979). Relations à éloignement minimum de relations binaires, note bibliographique. *Mathématiques et Sciences Humaines*, *67*, 115–122.
Monjardet, B. (1990). Sur diverses formes de la "règle de Condorcet" d'agrégation des préférences. *Mathématiques, Informatique et Sciences Humaines*, *111*, 61–71.
Orlin, J. B. (1981). Unpublished manuscript.
Reinelt, G. (1985). *The linear ordering problem: algorithms and applications*. *Research and exposition in mathematics*, *Vol. 8*. Berlin: Heldermann.
Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika*, *48*, 303–312.
Tucker, A. W. (1960). On directed graphs and integer programs. In *1960 symposium on combinatorial problems*, Princeton University, USA.
Wakabayashi, Y. (1986). *Aggregation of binary relations: algorithmic and polyhedral investigations*. PhD thesis, Augsbourg.
Wakabayashi, Y. (1998). The complexity of computing medians of relations. *Resenhas*, *3*(3), 323–349.
Younger, D. H. (1963). Minimum feedback arc sets for a directed graph. *IEEE Transactions of the Group in Circuit Theory*, *10*(2), 238–245.