Journal of Computer and System Sciences 105 (2019) 130–157

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

k-Majority digraphs and the hardness of voting with a constant number of voters

Georg Bachmeier [a], Felix Brandt [b], Christian Geist [b], Paul Harrenstein [c], Keyvan Kardel [b], Dominik Peters [c,*], Hans Georg Seedig [b]

[a] ETH Zürich, Switzerland
[b] TU München, Germany
[c] University of Oxford, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Article history:
Received 17 June 2017
Received in revised form 2 April 2019
Accepted 19 April 2019
Available online 9 May 2019

Keywords:
Computational social choice
Voting
Tournaments
NP-hardness

ABSTRACT

Many hardness results in computational social choice use the fact that every digraph may be induced as the pairwise majority relation of some preference profile. The standard construction requires a number of voters that is almost linear in the number of alternatives and it is unclear whether hardness holds when the number of voters is bounded. In this paper, we systematically study majority digraphs inducible by a constant number of voters. First, we characterize digraphs inducible by two or three voters, and give sufficient conditions for more voters. Second, we use SAT solvers to compute the minimum number of voters required to induce digraphs given by generated and real-world preference profiles. Finally, using our sufficient conditions, we show that several voting rules remain hard to evaluate for small constant numbers of voters. Kemeny's rule remains hard for 7 voters; previous methods could only prove this for constant even numbers of voters.

1. Introduction

A significant part of *computational social choice* is concerned with the computational complexity of voting problems. For most of the voting rules proposed in the social choice literature, it has been studied how hard it is to determine winners, to identify beneficial strategic manipulations, or to influence the outcome by bribing, partitioning, adding, or deleting voters (see, e.g., [12,50,25]). In many cases, the corresponding problems turned out to be NP-hard. Depending on the nature of the problem, this can be interpreted as bad news—as in the case of winner determination—or good news—as in the case of manipulation, bribery, and control.

In the standard voting setting, voters report their preferences as a ranking of a set of alternatives. From this information, we can compute the *pairwise majority relation*, which can be drawn as a directed graph which indicates, for each pair a, b of alternatives, whether a majority of voters prefers a to b, or prefers b to a. Many voting rules are based on the majority relation (or a weighted version of this relation), which establishes a fruitful connection between voting theory and graph theory. Perhaps the most fundamental result in this context is *McGarvey's theorem*, which states that *every* asymmetric directed graph may be induced as the pairwise majority relation of some preference profile [46]. McGarvey's theorem is the basis of most hardness results concerning majoritarian voting rules, since it allows reductions to construct a graph rather

* Corresponding author.
E-mail address: dominik.peters@cs.ox.ac.uk (D. Peters).

https://doi.org/10.1016/j.jcss.2019.04.005

than a complicated preference profile. McGarvey's original construction requires $n(n-1)$ voters, where n is the number of alternatives. This number has subsequently been improved by Stearns [52] and Erdős and Moser [24], who have eventually shown that the number of required voters is of order $\Theta(n/\log n)$. Since the result by Erdős and Moser [24] gives a lower bound as well, it implies that, for any constant number of voters, there are majority digraphs that cannot be induced by any preference profile. As a consequence, existing hardness proofs about majoritarian voting rules implicitly require that the number of voters is roughly of the same order as the number of alternatives.

In some applications, however, the number of voters is much smaller than the number of alternatives. A typical example is search engine aggregation, where the voters correspond to Internet search engines and the alternatives correspond to the webpages ranked by the search engines (see, e.g., [22]). In these settings, it is unclear whether hardness still holds. Referring to problems associated with Kemeny's [36] rank aggregation rule, which is based on the weighted majority digraph, Hudry [33] writes that "to my knowledge, when not trivial, the complexity for lower values of n remains unknown. In particular, it would be interesting to know whether some of the problems [...] remain NP-hard if n is a given constant". As a notable exception, Hudry [33] mentions a proof by Dwork et al. [22], which shows NP-hardness of Kemeny's rule without needing to appeal to McGarvey's result and only requires 4 voters. The result also holds for every larger *even* constant number of voters. In the *Handbook of Computational Social Choice*, Fischer et al. [28] note that "quite intriguingly, the case for any odd $n \geq 3$ remains open". In a similar vein, Hudry [33] writes that "it would be interesting to decide whether it is still the case for fixed values of n with n odd", and Biedl et al. [4] find that Dwork et al.'s [22] method "does not work for odd numbers of [voters]" and that particularly the case for $n = 3$ remains "wide open".

In this paper, we provide a systematic study of majority digraphs for a constant number of voters resulting in analytical, experimental, and complexity-theoretic insights. Starting from a discussion of bounds on the size of the smallest tournaments that require a certain number of voters (Section 3), we analyze the structure of majority digraphs inducible by a constant number of voters. Obviously, the fewer voters there are, the more restricted is the corresponding class of inducible majority digraphs. For instance, digraphs induced by two voters have to be acyclic (and are subject to some additional restrictions).

Analytically, we completely characterize digraphs inducible by two and three voters, respectively, and provide sufficient conditions for digraphs to be induced by k voters (Section 4). We propose a surprisingly efficient implementation via SAT solving for computing the minimal number of voters that is required to induce a given digraph (Section 5) and *experimentally* evaluate how many voters are required to induce the majority digraphs of real-world and generated preference profiles (Section 6).

In Section 7, we then finally leverage the conditions from Section 4 to investigate whether common, computationally intractable voting rules (the Banks set, the tournament equilibrium set, the minimal extending set, Kemeny's rule, Slater's rule, and ranked pairs) remain intractable when there is only a small constant number of voters. This is achieved by analyzing existing hardness proofs and checking whether the class of majority digraphs used in these constructions can be induced by small constant numbers of voters. Perhaps surprisingly, it turns out that all hardness proofs we studied can be constructed using at most 11 voters, and for many proofs, including one for Kemeny's rule, 7 voters suffice.

The paper concludes with an overview of the achieved hardness results, summarized in Table 6, and a brief outlook on future research in Section 8.

Our work can be viewed as complementary to the work by Conitzer et al. [19] who considered manipulation problems with a constant number of *candidates*. Conitzer et al. [19] determined how many candidates are required such that the problem of coalitional manipulation with weighted voters becomes NP-hard for a number of tractable voting rules including Borda's rule, Copeland's rule, and maximin.

2. Preliminaries

This section introduces the notation and terminology required to state our results.

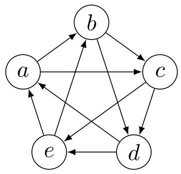
A *directed graph* or *digraph* is a pair (V, E), where V is finite a set of vertices and $E \subseteq V \times V$ is a set of arcs (directed edges). The *size* of a digraph is its number of vertices $|V|$. By $\mathcal{G}$ we denote the class of all digraphs and by $\mathcal{G}_n$ the class all digraphs of size n. The *converse of* E is $\overline{E} = \{(w, v) \colon (v, w) \in E\}$, where the direction of all arcs is reversed. Often it will be useful to effectively disregard orientations by considering $\overleftrightarrow{E} = E \cup \overline{E}$. We say that E_1 and E_2 are *orientation compatible* if $E_1 \cap (\overleftrightarrow{E_1} \cap \overleftrightarrow{E_2}) = E_2 \cap (\overleftrightarrow{E_1} \cap \overleftrightarrow{E_2})$, i.e., if for all $e \in \overleftrightarrow{E_1} \cap \overleftrightarrow{E_2}$, $e \in E_1$ if and only if $e \in E_2$.

The *incomparability graph* $\tilde{G} = (V, \tilde{E})$ associated with a digraph (V, E) is defined such that for all $v, w \in V$,

$$(v, w) \in \tilde{E} \text{ if and only if neither } (v, w) \in E \text{ nor } (w, v) \in E.$$

Obviously, $\overleftrightarrow{\tilde{E}} = \tilde{E}$.

A digraph $G = (V, E)$ is said to be *transitive* if for all $x, y, z \in V$, $(x, y) \in E$ and $(y, z) \in E$ imply $(x, z) \in E$. Moreover, G is *acyclic* if for all $x_1, \dots, x_k \in V$, $(x_1, x_2), (x_2, x_3), \dots, (x_{k-1}, x_k) \in E$ implies $(x_k, x_1) \notin E$. Also G is *asymmetric* if $(v, w) \in E$ implies $(w, v) \notin E$. A *tournament* is an asymmetric digraph (V, E) where E is complete, i.e., if for all distinct $v, w \in V$, either $(v, w) \in E$ or $(w, v) \in E$. We denote the sets of all tournaments by $\mathcal{T}$ and the set of those with n vertices by $\mathcal{T}_n$. Moreover, a digraph (V, E) is *transitively (re)orientable* if there exists a transitive and asymmetric digraph (V, E') with $\overleftrightarrow{E'} = \overleftrightarrow{E}$. E' is also referred to as a *reorientation of* E.



R_1	R_2	R_3
a	d	c
b	e	e
c	a	b
d	b	d
e	c	a

Fig. 1. A majority digraph and a 3-voter preference profile that induces the digraph.

The digraphs in this paper are assumed to be induced by the preferences of a set of voters. Let $N = \{1, \ldots, k\}$ be a set of k voters (or an *electorate* of size k) and V a set of alternatives. The preferences of each voter i are given as *linear orders*, i.e., transitive, complete, and antisymmetric relations R_i over a set of alternatives V. A *preference profile* $R = (R_1, \ldots, R_k)$ associates a preference relation with each voter. Each preference profile gives rise to a *majority relation*, which holds between two alternatives v and w if the number of voters preferring v to w exceeds the number of voters preferring w to v. We say that (V, E) is *the majority digraph* of preference profile R if

$$(v, w) \in E \text{ if and only if } |\{i \in N\colon\ v\ R_i\ w\}| > |\{i \in N\colon\ w\ R_i\ v\}|.$$

Majority digraphs are asymmetric. Moreover, if the number of voters is odd, the majority digraph is complete and thus a tournament.

We say that (V, E) is *k-inducible* if (V, E) is the majority digraph for some preference profile involving k voters. Equivalently, we say that (V, E) is a *k-majority digraph*.[1] As an example, Fig. 1 shows a tournament which is induced by a 3-voter profile, and thus this tournament is a 3-inducible majority digraph.

We also consider *weighted digraphs* (V, w), where V is a set of vertices and $\mathrm{w}\colon V \times V \to \mathbb{Z}$ a weight function assigning weight $\mathrm{w}(v, w)$ to the arc (v, w). With a slight abuse of notation we also refer to weighted digraphs as a pair (V, E), where the weight function is subsumed and it is understood that $E = \{(v, w)\colon \mathrm{w}(v, w) > 0\}$. We say that a weighted digraph (V, w) is induced by R if for all $v, w \in V$, $\mathrm{w}(v, w) = |\{i \in N\colon\ v\ R_i\ w\}| - |\{i \in N\colon\ w\ R_i\ v\}|$. In this case, (V, w) is a *weighted k-majority digraph*.

Given a (weighted) digraph we are interested in the minimal number of voters needed such that the digraph represents the (weighted) majority relation of the voters' preferences. This is captured in the majority dimension of the digraph.[2] Formally, the *majority dimension* of a digraph $G = (V, E)$ or a weighted digraph $G = (V, \mathrm{w})$ is the smallest number of voters in a profile that induces G, i.e.,

$$\dim(G) = \min\{k\colon\ G \text{ is a (weighted) } k\text{-majority digraph}\}.$$

Also, let $k_{\text{maj}}(n)$ denote the minimum electorate size required to induce all digraphs of size n, i.e.,

$$k_{\text{maj}}(n) = \min\{k\colon\ \dim(G) \le k \text{ for all } G \in \mathcal{G}_n\}.$$

If we restrict our attention to tournaments, we will write $k^{\mathcal{T}}_{\text{maj}}(n)$ instead. Note that $k^{\mathcal{T}}_{\text{maj}}(n) \le k_{\text{maj}}(n)$ since $\mathcal{T} \subset \mathcal{G}$.

Conversely, define the *majoritarian expressiveness* of (electorates of size) k to be the maximum integer $n^{\mathcal{T}}(k)$ such that every complete majority relation on up to $n^{\mathcal{T}}(k)$ alternatives is k-inducible. Since the work by Erdős and Moser [24], which we discuss in more detail in Section 3, it is known that $n^{\mathcal{T}}(k)$ is finite for every k. Note that this implies, that the smallest tournament that cannot be induced by k voters is of size $n^{\mathcal{T}}(k) + 1$.

By a *voting rule* we understand a function that maps each preference profile to a non-empty subset of the alternatives. Over the years, a large number of voting rules have been proposed. The ones we will be concerned with in this paper—because they are based on majority digraphs and computationally intractable—are the *Banks set (BA)*, the *tournament equilibrium set (TEQ)*, the *minimal extending set (ME)*, *Slater's rule (SL)*, *Kemeny's rule*, and *ranked pairs (RP)*. The definitions of these rules are given in Section 7.

3. Bounds on the majority dimension and majoritarian expressiveness

This section is concerned with bounds on the majority dimension of digraphs and on the majoritarian expressiveness of electorates of a fixed size.

[1] Alon et al. [2] used the term *k-majority tournament* for tournaments that are induced by a $(2k-1)$-voter profile because every majority consists of at least k voters. We chose to follow the terminology of Kierstead et al. [37] instead.

[2] This complexity measure of digraphs can also be interpreted as a complexity measure of preference profiles. The majority dimension of a given preference profile is then simply defined as the majority dimension of the induced majority digraph.

Table 1
Upper bounds, obtained by counting arguments, on the size $n^{\mathcal{T}}(k)$ of the smallest tournament that is not k-inducible for small odd k.

k	3	5	7	9	11	13	15	17	19	21
$n^{\mathcal{T}}(k)$	18	41	66	93	122	152	183	216	249	282

The first such result is due to McGarvey [46], who showed that every digraph can be induced by some (finite) preference profile. He gave a construction that requires exactly two voters per arc in the digraph. In our notation, this implies that $k_{\text{maj}}(n) \leq n(n-1) < \infty$ for all n.

The work by McGarvey [46] has been followed up by Stearns [52] who showed that $k_{\text{maj}}(n) \leq n+2$, which was later improved by Fiol [27] to $k_{\text{maj}}(n) \leq n - \lfloor \log n \rfloor + 1$. For larger n, Erdős and Moser [24] gave the asymptotically better bound $k_{\text{maj}}(n) \leq c \cdot \frac{n}{\log n}$ for some constant c. Their work nicely complemented an earlier result by Stearns [52] who proved that $k_{\text{maj}}(n) > 0.55 \cdot \frac{n}{\log n}$ for large n using a counting argument. These results together asymptotically capture the growth of $k_{\text{maj}}(n)$.

Theorem 1 *(Stearns [52], Erdös and Moser [24]).* $k_{maj}(n) \in \Theta(\frac{n}{\log n})$.

In the following, we are particularly interested in the majority dimension of tournaments. The following simple observation about the parity of $\dim(G)$ will be useful.

Lemma 1. *The majority dimension* $\dim(G)$ *is odd if* G *is a tournament and even otherwise.*

Proof. Let G be a tournament and assume that $\dim(G) = k$ was even. Then there exists a preference profile R with k voters that induces T. Since k is even, the majority margin must be even for every pair of alternatives and can furthermore never be zero as T is a tournament. Therefore, removing any single voter from R gives a profile R' with just $k-1$ voters that still induces T, a contradiction.

For incomplete digraphs, the statement follows directly from the fact that for all preference profiles R with an odd number of voters k, the majority relation is complete and anti-symmetric (as no majority ties can occur). □

Note that the lower bound on $k_{\text{maj}}(n)$ due to Stearns [52] shows that, for every electorate of size k, there exist *digraphs* that are not k-inducible. Still, the majority dimension of *tournaments*, $k^{\mathcal{T}}_{\text{maj}}(n)$, could be bounded by a constant. In that case, all tournaments could be inducible by some constant-size electorate. The following lemma shows that this is not the case. The argument is similar to the one by Stearns [52].

Lemma 2. *If* $k^{\mathcal{T}}_{maj}(n) = k \geq 3$, *then*

$$\binom{n}{2} \cdot \ln(2) \leq k \cdot \left(\ln(2) + \sum_{i=2}^{n} \ln(i) \right) - \ln(k!). \tag{1}$$

Proof. If every tournament on n vertices can be induced by k voters, then for every $T \in \mathcal{T}_n$, there needs to be at least one k-voter profile that induces T. There are $n!$ possible preference orders over n alternatives, and—when ignoring the identities of voters—the number of k-voter profiles is $\binom{n!+k-1}{k}$. Also, the number of labeled tournaments on n vertices is $2^{\binom{n}{2}}$ implying that

$$2^{\binom{n}{2}} \leq \binom{n!+k-1}{k} \leq \frac{(2(n!))^k}{k!}$$

where the last inequality follows from Fiol's [27] bound stated above. The result follows immediately. □

Using the lemma, we can search for an upper bound on the majoritarian expressiveness $n^{\mathcal{T}}(k)$ for a given k by finding the minimal n such that (1) is violated. Table 1 shows some upper bounds for small k. For example, there exists a tournament of size 42 that is not 5-inducible.[3]

It is clear, however, that these bounds are not tight. For example, the results in the table imply there has to exist a tournament of size 19 that is not 3-inducible. In fact, Shepardson and Tovey [51] found an explicit tournament of size 8 that is not 3-inducible; they exhibit a certain digraph of size 8 such that any tournament that contains it as a subgraph is

[3] A slightly tighter analysis even gives the existence of such a tournament of size 41.

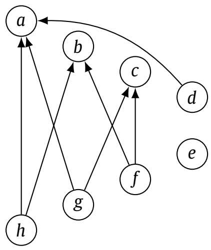


Fig. 2. A tournament on 8 vertices with majority dimension 5. This is a smallest tournament that cannot be induced by three voters. Omitted arcs point downwards.

not 3-inducible. An example of such a tournament is shown in Fig. 2. In Section 6.1, we will argue that every tournament of size less than 8 is 3-inducible.

What about explicit examples of tournaments that are not 5-inducible? An argument of Alon et al. [2] allows one to construct, for any odd k, a concrete tournament that is not k-inducible. The constructed tournaments are *quadratic residue tournaments*. For a prime p, the quadratic residue tournaments $Q_p = (V, E)$ of size p has vertex set $V = (v_1, \dots, v_p)$ and $(v_i, v_j) \in E$ if and only if $(i-j)^{\frac{p-1}{2}} \equiv 1 \mod p$. Alon et al. [2] shows that Q_p is not k-inducible for sufficiently large p.

The argument by Alon et al. [2] runs along the following lines. A *dominating set* of a digraph $G = (V, E)$ is a set $U \subseteq V$ such that for all $v \in V \setminus U$, there exists a $u \in U$ with $(u, v) \in E$. Alon et al. [2] showed that the size of the smallest dominating set of any k-majority digraph for odd k is bounded from above by a function $\mathcal{F}(k)$ with $\mathcal{F}(k) \in O(k \log k)$ and $\mathcal{F}(k) \in \Omega(\frac{k}{\log k})$ with relatively large constants hidden in the Landau notation (80 for the upper bound). This means that if a given tournament T does not have a dominating set of size $\mathcal{F}(k)$, then T is not k-inducible.

We can now construct a tournament that is not k-inducible using the following constructive result by Graham and Spencer [31]. Let $f(x) = p > x^2 2^{2x-2}$, where x is a positive integer and p is the smallest prime congruent to $3 \pmod 4$ satisfying the inequality (the construction works for any such p). Then, the quadratic residue tournament Q_p of size p does not exhibit a dominating set of size x.

Together, this yields, for any odd k, a construction for a tournament on $(f \circ \mathcal{F})(k)$ vertices that is not k-inducible. Unfortunately, $f(x)$ is exponential in x, and the value of $\mathcal{F}(k)$ is known precisely only for $k = 3$ where we have $\mathcal{F}(3) = 3$. To the best of our knowledge, the best currently available bound for $k = 5$ is $\mathcal{F}(5) \le 12$ [26]. Together, we get that for the smallest (or any other) prime p congruent to $3 \pmod 4$ satisfying the inequality

$$p > 12^2 \cdot 2^{2 \cdot 12 - 2} = 603\,979\,776,$$

the tournament Q_p is not 5-inducible. The smallest p satisfying these conditions is $p = 603\,979\,799$. We do not know an explicit tournament smaller than this which is not 5-inducible.

Bounds on $\mathcal{F}(k)$ for larger odd k give significantly worse values: for 7 voters, we only know that $\mathcal{F}(7) \le 44$ [26], which would translate to a quadratic residue tournament with more than 10^{29} vertices.

4. Majority digraphs of few voters

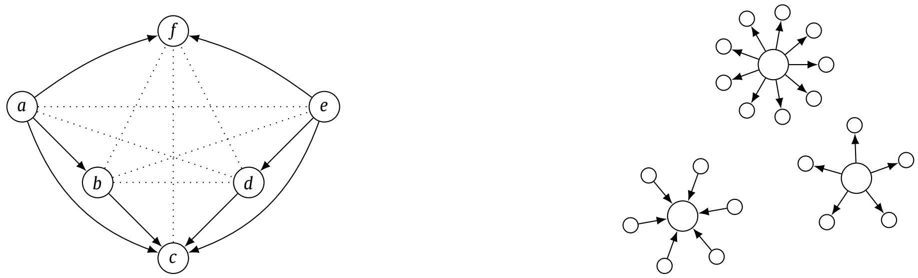
In this section, we analyze the structure of k-inducible digraphs for constant k. Dushnik and Miller [21] characterize 2-majority digraphs. Based on their work, we give a characterization of 3-majority digraphs. In addition, we present sufficient conditions for larger majority dimensions that will be leveraged in Section 7.

4.1. Two and three voters

Given a preference profile R, the *Pareto relation* holds between two alternatives v and w if all voters prefer v over w. Dushnik and Miller [21] specified sufficient and necessary conditions for relations to be induced as the Pareto relation of a 2-voter profile. For two voters, the majority relation and the Pareto relation coincide. Hence, we can rephrase their result for majority digraphs as follows.

Lemma 3 *(Dushnik and Miller [21]). A majority digraph (V, E) is 2-inducible if and only if it is transitive and its incomparability graph $(V, \tilde{E})$ is transitively orientable. Moreover, the weight of every arc is 2.*

See Fig. 3a for an example of a digraph that is not 2-inducible even though it is transitive. If it were 2-inducible, there would have to exist a transitive reorientation E' of $\tilde{E}$. We can assume without loss of generality that $(b, d) \in E'$. Then, $(a, d) \in E'$ as (d, a) would imply $(a, b) \in E'$. Similarly, $(b, d) \in E'$ forces $(b, e) \in E'$, as otherwise $(e, d) \in E'$. Moreover, $(a, e) \in E'$, as $(e, a) \in E'$ would imply $(b, a) \in E'$. Now, $(f, b) \in E'$ would imply $(f, e) \in E'$, and hence $(b, f) \in E'$. It also follows that $(f, d) \in E'$, as otherwise transitivity of E' would imply (a, f). This leaves us to orient the edge $\{f, c\}$. If $(f, c) \in E'$,



(a) This digraph is not 2-inducible. Dotted arcs denote the incomparability graph. (b) Every forest of unidirected stars is 2-inducible.

Fig. 3. Examples of transitive digraphs.

then $(b, f) \in E'$ would imply $(b, c) \in E'$, which cannot be the case. If on the other hand, $(c, f) \in E'$, then $(f, d) \in E'$ and transitivity of E' would imply that $(c, d) \in E'$ as well: a contradiction. We may therefore conclude that the digraph is not 2-inducible.

If, on the other hand, a digraph (V, E) is in fact induced by a 2-voter profile (R_1, R_2), then R_1 and R_2 coincide on E and are opposed on $\tilde{E}$, i.e., $R_1 \cap R_2 = E$. As R_1 and R_2 are both transitive, so is E. If E' is the respective reorientation of $\tilde{E}$, then $R_1 = E \cup E'$ and $R_2 = E \cup \overline{E'}$, or *vice versa*.

A digraph (V, E) is a *unidirected star* if there is some $v^* \in V$ such that either E or $\overline{E}$ equals $\{v^*\} \times (V \setminus \{v^*\})$. Clearly, (V, E) is transitive as there are no $v, w, u \in V$ such that both $(v, w), (w, u) \in E$. Moreover, every transitive relation over the leaves of (V, E) serves as a transitive orientation of $\tilde{E}$. With Lemma 3 this gives us the following result.

Lemma 4. *Every unidirected star is* 2*-inducible.*

Lemma 4 is a special case of Lemma 1 by Erdős and Moser [24], where they consider a larger class of digraphs: Say that (V, E) is a *unidirected bipartite* digraph if there is a partition of V into V_1 and V_2 such that $E = V_1 \times V_2$, i.e., such that every vertex in V_1 has an arc to every vertex in V_2. By a similar argument as above, these graphs are also 2-inducible.

Lemma 5 *(Erdös and Moser [24]). Every unidirected bipartite digraph is* 2*-inducible.*

Erdős and Moser [24] exhibit an explicit inducing profile for such graphs.

Another insight follows from Lemma 3: the union of pairwise disjoint 2-inducible digraphs is itself induced by a 2-voter profile.

Lemma 6. *Let $V_1, \dots, V_k$ be pairwise disjoint and $(V_1, E_1), \dots, (V_k, E_k)$* 2*-inducible majority digraphs. Then, $(V_1 \cup \dots \cup V_k, E_1 \cup \dots \cup E_k)$ is also* 2*-inducible.*

Proof. Let $V = V_1 \cup \dots \cup V_k$ and $E = E_1 \cup \dots \cup E_k$ and consider the digraph (V, E). Since each of $(V_1, E_1), \dots, (V_k, E_k)$ is 2-inducible, by Lemma 3, each of $E_1, \dots, E_k$ is transitive and each of $\tilde{E}_1, \dots, \tilde{E}_k$ is transitively orientable. Let $E'_1, \dots, E'_k$ be the respective transitive reorientations of $\tilde{E}_1, \dots, \tilde{E}_k$. Since $V_1, \dots, V_k$ are pairwise disjoint, $E_1 \cup \dots \cup E_2$ can readily be seen to be transitive as well. Let furthermore $E^* = \bigcup_{1 \le i < j \le k} (V_i \times V_j)$. Observe that $\tilde{E} = \tilde{E}_1 \cup \dots \cup \tilde{E}_k \cup \overleftrightarrow{E^*}$ and that $E'_1 \cup \dots \cup E'_k \cup E^*$ is a transitive reorientation of $\tilde{E}$. The claim then follows by another application of Lemma 3. □

Consequently, every forest of (unidirected) stars such as the one shown in Fig. 3b is 2-inducible. More generally, a disjoint collection of unidirected bipartite digraphs is 2-inducible; Erdős and Moser [24] call these graphs *bilevel graphs*.

Apart from a family of tournaments of order eight that are *not* 3-inducible [51], not much is known about 3-majority digraphs. Clearly, all 3-majority digraphs are tournaments (cf. Lemma 1). We now provide a characterization of these tournaments in graph-theoretic terms. However, there is no obvious computationally efficient way to check our condition.

Lemma 7. *A tournament (V, E) is* 3*-inducible if and only if there are disjoint sets E_1, E_2 with $E = E_1 \cup E_2$ such that E_1 is transitive and E_2 is both acyclic and transitively reorientable. Then, the weight of every arc in E_1 is either* 1 *or* 3 *and that of each arc in E_2 is* 1*.*

Proof. For the if-direction, assume that there are disjoint sets E_1, E_2 with $E = E_1 \cup E_2$ such that E_1 is transitive and E_2 is both acyclic and transitively reorientable. Consider the digraph (V, E_1) and observe that for the corresponding incomparability graph $(V, \tilde{E}_1)$, $\tilde{E}_1 = \overleftrightarrow{E_2}$. It follows that $\tilde{E}_1$ is transitively orientable and, by Lemma 3, that (V, E_1) is induced by a 2-voter profile (R_1, R_2) and that all arcs in E_1 have weight 2. As E_2 is acyclic, there is a (strict) preference relation R_3 with $E_2 \subseteq R_3$. Now consider the majority digraph induced by the preference profile (R_1, R_2, R_3), which apparently coincides with

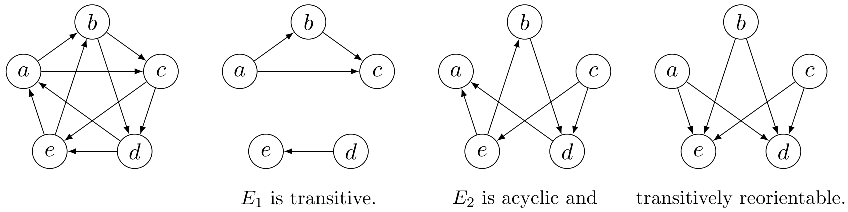


Fig. 4. A 3-inducible majority digraph and its arc set partitioning into E_1 and E_2, satisfying the conditions of Theorem 7. The part E_1 is induced by two voters $a\ R_1\ b\ R_1\ c\ R_1\ d\ R_1\ e$ and $d\ R_2\ e\ R_2\ a\ R_2\ b\ R_2\ c$. The part E_2 is compatible with $c\ R_3\ e\ R_3\ b\ R_3\ d\ R_3\ a$. Together (R_1, R_2, R_3) induce the digraph.

(V, E). E_1 is determined by R_1 and R_2 and each of its arcs obtains weight 1 or 3 depending on whether R_3 agrees with both R_1 and R_2 or not. Moreover, E_2 is determined by R_3, as R_1 and R_2 can be assumed to specify contrary preferences on this part.

For the only-if-direction, assume that (V, E) is the majority digraph induced by the 3-voter profile (R_1, R_2, R_3). Let furthermore (V, E_1) be the majority digraph induced by (R_1, R_2) and $E_2 = R_3 \cap ((V \times V) \setminus \overleftrightarrow{E_1})$. By Lemma 3, (V, E_1) is transitive and $\tilde{E}_1$ is transitively (re)orientable, where $(V, \tilde{E}_1)$ is the incomparability graph of (V, E_1). Since R_3 is transitive (and strict) E_2 is obviously acyclic. Observe furthermore that $\overleftrightarrow{R_3 \cap ((V \times V) \setminus \overleftrightarrow{E_1})} = \overleftrightarrow{\tilde{E}_1}$. It follows that E_2 is transitively reorientable. □

In order to illustrate Theorem 7, again consider the introductory example given in Fig. 1. This digraph is 3-inducible because its arc set can be partitioned into a transitive part and an acyclic and transitively reorientable part (see Fig. 4).

4.2. More than three voters

Extensions of these results provide useful sufficient conditions for a digraph to be induced by a constant larger number of voters. If the arc set of a digraph can be decomposed into pairwise orientation compatible sets that satisfy the conditions of Lemma 3, the digraph is induced by a profile with two voters per set.

Lemma 8. *Let $(V, E_1), \dots, (V, E_k)$ be majority digraphs induced by 2-voter profiles such that $E_1, \dots, E_k$ are pairwise orientation compatible. Then, $(V, E_1 \cup \dots \cup E_k)$ is induced by a $2k$-voter profile.*

Proof. Let for each i with $1 \le i \le k$, (R_1^i, R_2^i) be a 2-voter profile that induces (V, E_i). By Lemma 3, for every $(v, w) \in E_i$ we know that both $v\ R_1^i\ w$ and $v\ R_2^i\ w$ and for every $(v, w) \notin E_i$, $v\ R_1^i\ w$ if and only if $w\ R_2^i\ v$. Now consider the preference profile $(R_1^1, R_2^1, \dots, R_1^k, R_2^k)$ and the majority digraph (V, E) it induces. We argue that $E = E_1 \cup \dots \cup E_k$. First assume that $(v, w) \in E_i$ for some i with $1 \le i \le k$. Then, both $v\ R_1^i\ w$ and $v\ R_2^i\ w$. Since $E_1, \dots, E_k$ are pairwise orientation compatible, $(w, v) \in E_j$ for no j with $1 \le j \le k$, i.e., for all j with $1 \le j \le k$ either $v\ R_1^j\ w$ and $v\ R_2^j\ w$, or $v\ R_1^j\ w$ if and only if $w\ R_2^j\ v$. It follows that a majority prefers v over w and thus $(v, w) \in E$. Now assume that $(v, w) \in E_i$ for no i with $1 \le i \le k$. Then for all i with $1 \le i \le k$ either both $w\ R_1^i\ v$ and $w\ R_2^i\ v$ or $w\ R_1^j\ v$ if and only if $v\ R_2^j\ w$. It is easy to see that v is not majority preferred to w, i.e., $(v, w) \notin E$. □

Next, we show that a similar condition suffices for a digraph to be inducible by a given odd number of voters.[4]

Lemma 9. *Let (V, E) be a tournament and $(V, E_1), \dots, (V, E_k)$ be majority digraphs induced by 2-voter profiles such that $E, E_1, \dots, E_k$ are orientation compatible. Let, moreover, $E_{k+1} \supseteq E \setminus (E_1 \cup \dots \cup E_k)$ be acyclic. Then, (V, E) is induced by a $2k+1$-voter profile.*

Proof. In virtue of Lemma 8 we know that $(V, E_1 \cup \dots \cup E_k)$ is induced by a $2k$-voter profile $(R_1^1, R_2^1, \dots, R_1^k, R_2^k)$. Inspection of the proof also reveals that every arc $(v, w) \in E_1 \cup \dots \cup E_k$ has a positive even weight of at least two. As E_{k+1} is acyclic and asymmetric, there is some (strict) preference relation R^{k+1} with $E_{k+1} \subseteq R^{k+1}$. Moreover, since E_{k+1} corresponds to only one voter and every arc in $E_1 \cup \dots \cup E_k$ has a majority of at least two, E_{k+1} does not have to be orientation compatible with any of $E_1, \dots, E_k$. It can then easily be seen that the majority digraph induced by $(R_1^1, R_2^1, \dots, R_1^k, R_2^k, R^{k+1})$ equals (V, E), $E_1 \cup \dots \cup E_k$ being determined by majorities of at least one in $(R_1^1, R_2^1, \dots, R_1^k, R_2^k, R^{k+1})$ and $E \setminus (E_1 \cup \dots \cup E_k)$ by R^{k+1}, each arc in which has then weight one. □

[4] The if-direction of Lemma 7 can also be obtained as a special case of this lemma.

Table 2
Number of objects involved in the CHECK-k-MAJORITY problem for one, three, and five voters.

	Preference profiles			Tournaments (unlabeled)
	$k=1$	$k=3$	$k=5$	
$n=5$	120	$\sim 1.7 \cdot 10^{6}$	$\sim 2.5 \cdot 10^{10}$	12
$n=10$	$\sim 3.6 \cdot 10^{6}$	$\sim 4.8 \cdot 10^{19}$	$\sim 6.3 \cdot 10^{32}$	$\sim 9.7 \cdot 10^{6}$
$n=25$	$\sim 1.6 \cdot 10^{25}$	$\sim 3.7 \cdot 10^{75}$	$\sim 9.0 \cdot 10^{125}$	$\sim 1.3 \cdot 10^{65}$
$n=50$	$\sim 3.0 \cdot 10^{64}$	$\sim 2.8 \cdot 10^{193}$	$\sim 2.6 \cdot 10^{322}$	$\sim 1.9 \cdot 10^{305}$
$n=100$	$\sim 9.3 \cdot 10^{157}$	$\sim 8.1 \cdot 10^{473}$	$\sim 7.1 \cdot 10^{789}$	$> 10^{1332}$

5. Determining the majority dimension of a digraph

This section addresses the computational problem of computing the majority dimension. To this end, we define the problem of checking whether a given digraph G is k-inducible, i.e., whether G is a k-majority digraph.

CHECK-k-MAJORITY
Input: A digraph G and a positive integer k.
Question: Is G a k-majority digraph?

Recall that for a digraph G, whether $\dim(G)$ is odd or even depends on whether G is complete (i.e., a tournament) or not, according to Lemma 1. While CHECK-2-MAJORITY can be solved in polynomial time [57], the complexity of CHECK-k-MAJORITY remains open for every fixed $k \geq 3$.

In the following, we provide an implementation for solving CHECK-k-MAJORITY. This implementation relies on an encoding of the problem as a Boolean satisfiability (SAT) problem which is then solved by a SAT solver. Although not polynomial time, this technique turns out to be surprisingly efficient and easily outperforms an implementation for CHECK-3-MAJORITY based on the graph-theoretic characterization in Section 4.

5.1. Computing the majority dimension via SAT

The number of objects potentially involved in the CHECK-k-MAJORITY problem are shown in Table 2. It is clear that a naive brute-force algorithm, iterating through all potential inducing preference profiles, will not solve the problem in a satisfactory manner.

Thus, in order to answer CHECK-k-MAJORITY, we translate a given problem instance to propositional logic (on a computer) and use state-of-the-art SAT solvers to find a solution. At a glance, the overall solving steps are shown in Algorithm 1.

Algorithm 1: SAT-CHECK-k-MAJORITY.

```
Input: digraph (V, E), positive integer k
Output: whether (V, E) is a k-majority digraph
/* Encoding of problem in CNF                                  */
File cnfFile;
foreach voter i do
    cnfFile += Encoder.reflexivePreferences(i);
    cnfFile += Encoder.completePreferences(i);
    cnfFile += Encoder.transitivePreferences(i);
    cnfFile += Encoder.antisymmetricPreferences(i);
end
cnfFile += Encoder.majorityImplications((V, E));
if E is not complete then
    cnfFile += Encoder.indifferenceImplications((V, E));
end
/* SAT solving                                                 */
Boolean satOutcome = SATsolver.solve(cnfFile);
return satOutcome;
```

We design the propositional formula so that a satisfying assignment represents a preference profile that induces the given digraph. We encode the preference profile in question using Boolean variables $r_{i,a,b}$, which encode whether $a\ R_i\ b$, i.e., whether voter i ranks alternative a at least as high as alternative b. We then need to impose the following constraints.

1. All k voters have linear orders over the n alternatives as their preferences.
2. (Majority implications.) For each majority arc $(x, y) \in E$ in the digraph, a majority of voters needs to prefer x over y.

3. (Indifference implications.) For each missing arc ($x \not\succ y$ and $y \not\succ x$) in the digraph, *exactly* half the voters need to prefer x over y.[5]

For the first constraint, we encode reflexivity, completeness, transitivity, and anti-symmetry of the relation R_i for all voters i. We only give details of how to translate transitivity to CNF (conjunctive normal form, the established standard input format for SAT solvers); reflexivity, completeness, and anti-symmetry are converted analogously. The translation first restates the transitivity requirement in terms of our Boolean variables, and then rewrites the condition in CNF:

$$
\begin{aligned}
& (\forall i)(\forall x, y, z)\,(x\, R_i\, y \wedge y\, R_i\, z \rightarrow x\, R_i\, z) \\
\equiv\ & (\forall i)(\forall x, y, z)\,\left(r_{i,x,y} \wedge r_{i,y,z} \rightarrow r_{i,x,z}\right) \\
\equiv\ & \bigwedge_i \bigwedge_{x,y,z} \left(\neg\left(r_{i,x,y} \wedge r_{i,y,z}\right) \vee r_{i,x,z}\right) \\
\equiv\ & \bigwedge_i \bigwedge_{x,y,z} \left(\neg r_{i,x,y} \vee \neg r_{i,y,z} \vee r_{i,x,z}\right).
\end{aligned}
$$

The condition in the first line is, at first sight, a higher order axiom. But we can still translate it to propositional logic: since the domains of the quantifiers are finite, the quantifiers can be replaced by finite conjunctions or disjunctions (as also pointed out by Geist and Endriss [30]). The same is true for the other constraints.

Majority and indifference implications can be formalized in a similar fashion. We describe the translation for the majority implications here; the procedure for the indifference implications (needed for incomplete digraphs) is analogous. In the following, we denote the smallest number of voters required for a positive majority margin by $m(k) := \lfloor k \cdot \frac{1}{2} \rfloor + 1$. Note that, due to the anti-symmetry of individual preferences, for $(x, y) \in E$ it suffices that there exist $m(k)$ voters who prefer x to y. In formal terms:

$$
\begin{aligned}
& (\forall x, y)\,((x, y) \in E \rightarrow |\{i\colon x\, R_i\, y\}| > |\{i\colon y\, R_i\, x\}|) \\
\equiv\ & (\forall x, y)\,((x, y) \in E \rightarrow |\{i\colon x\, R_i\, y\}| \geq m(k)) \\
\equiv\ & (\forall x, y)\,((x, y) \in E \rightarrow (\exists M \subseteq \{1, \ldots, k\})|M| = m(k) \wedge (\forall i \in M) x\, R_i\, y) \\
\equiv\ & \bigwedge_{(x,y)\in E}\ \bigvee_{|M|=m(k)}\ \bigwedge_{i \in M} r_{i,x,y}.
\end{aligned}
$$

The resulting formula is not in CNF, so it needs to be converted. In order to avoid an exponential blow-up, we apply the standard technique of variable replacement (also known as Tseitin transformation [55]). Note that conditions like $|M| = m(k)$ can easily be checked during generation of the corresponding CNF formula on a computer.

Overall, after the Tseitin transformation, this encoding leads to a total of $k \cdot n^2 + \binom{k}{m(k)} \cdot n^2 = n^2 \cdot \left(k + \binom{k}{m(k)}\right)$ variables for the case of tournaments and $n^2 \cdot \left(k + \binom{k}{m(k)} + \binom{k}{k/2}\right)$ variables for incomplete digraphs. The number of clauses is equal to $k \cdot (n^3 + n^2) + \frac{n^2 - n}{2} \cdot \left(1 + \binom{k}{m(k)} \cdot m(k)\right)$ for tournaments, and at most $k \cdot (n^3 + n^2) + (n^2 - n) \cdot \left(1 + \binom{k}{k/2} \cdot \frac{k}{2}\right)$ for incomplete digraphs, respectively.

With all axioms formalized in propositional logic, we are now ready to analyze arbitrary digraphs G for their majority dimension $\dim(G)$. Before we do so, however, we describe an optimization technique for tournament graphs, which, for certain instances, significantly speeds up the computation.

5.1.1. Optimization through decomposition of tournaments

Decompositions of tournaments into components have been widely studied (see, e.g., [38,11]). Algorithmically, Brandt et al. [9] show that voting problems can be solved more efficiently on tournaments that admit a decomposition, using recursive algorithms. In this section, we prove that a similar optimization can be carried out for the computation of the majority dimension $\dim(G)$ of a given tournament G.

As an example, consider the tournament shown in Fig. 5. This tournament has two *components*, namely the cycle on $\{a, b, c\}$ and the cycle on $\{d, e, f\}$. All arcs between these two components point in the same direction (from $\{a, b, c\}$ to $\{d, e, f\}$). If we want to construct a profile inducing the tournament, this structure allows us to consider the two components separately: find profiles inducing each component, and then glue the profiles together. We use this strategy to show that the majority dimension of a tournament is equal to the maximum of the majority dimension of its components and the corresponding summary tournament.

In formal terms, a non-empty subset C of V is a *component* of a tournament $G = (V, E)$ if, for all $v \in V \setminus C$, either $(v, w) \in E$ for all $w \in C$ or $(w, v) \in E$ for all $w \in C$. A *decomposition* of G is a set of pairwise disjoint components

[5] This axiom is only required for incomplete digraphs.

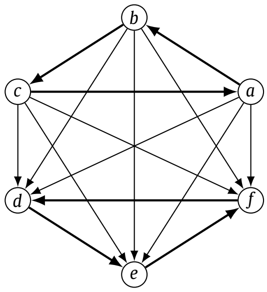


Fig. 5. A tournament that can be decomposed into two components $\{a, b, c\}$ and $\{d, e, f\}$.

$\{C_1, \ldots, C_p\}$ of T such that $V = \bigcup_{j=1}^{p} C_j$. Every tournament admits a decomposition that is minimal in a well-defined sense [38] and that can be computed in linear time [45,9]. Given a particular decomposition $\tilde{C} = \{C_1, \ldots, C_p\}$, the *summary* of (V, E) with respect to $\tilde{C}$ is defined as the tournament $(\tilde{C}, \tilde{E})$ on the individual components rather than the alternatives, i.e.,

$$(C_q, C_r) \in \tilde{E} \quad \text{if and only if} \quad (v, w) \in E \text{ for all } v \in C_q, w \in C_r.$$

The following lemma enables the recursive computation of $\dim(G)$ along the component structure of G.

Lemma 10. *Let $G = (V, E)$ be a tournament, $\tilde{C} = \{C_1, \ldots, C_p\}$ a decomposition of G, $G_j = (C_j, E|_{C_j})$ for all $j \in \{1, \ldots, p\}$, and $\tilde{G} = (\tilde{C}, \tilde{E})$. Then*

$$\dim(G) = \max_j\{\dim(G_j), \dim(\tilde{G})\}.$$

Proof. Let R be a minimal profile inducing G. Then, $R|_{C_j}$ induces G_j for every C_j establishing $\dim(G) \geq \dim(G_j)$. That $\dim(G) \geq \dim(\tilde{G})$ holds is also easy to see by considering a variant of R in which from each component all but one vertex are chosen arbitrarily and removed. The remaining profile then induces $\tilde{G}$. For the other direction, let $z = \max_j\{\dim(G_j), \dim(\tilde{G})\}$. We know by Lemma 1 that $\dim(\tilde{G})$ and every $\dim(G_j)$ is odd as these are all tournaments. Each G_j (and $\tilde{G}$) has a minimal profile R^j (and $\tilde{R}$, respectively). We can add pairs of voters with opposing preferences to each profile without changing its majority relation. This way, we get profiles R'^j (and $\tilde{R}'$) that still induce G_j (and $\tilde{G}$, respectively) but now all have the same number of voters z. Now, create a new profile R from R' in which $R_i'^j$ replaces alternative j as a segment in R_i' for each voter i and every alternative j. It is easy to check that R has z voters and still induces G, i.e., $\dim(G) \leq z$. $\square$

We have implemented this optimization and found that many real-world majority digraphs exhibit non-trivial decompositions, speeding up the computation of Sat-Check-k-Majority.

5.2. Computational efficiency

The characterization of 3-majority digraphs in Section 4 suggests an alternative algorithm based on enumerating 2-partitions of the input tournament, and checking the condition of Lemma 7. The corresponding algorithm 2-Partition-Check-3-Majority is given in Algorithm 2. Besides enumerating all 2-partitions of the majority arcs, the only non-trivial part is to check whether a relation has a transitive reorientation. This can be done efficiently using an algorithm by Pnueli et al. [49]. While not polynomial time, this algorithm easily outperforms algorithms based on enumerating all possible 3-voter preference profiles, since there are many more profiles than 2-partitions (also see Table 2).

We compared the running times of 2-Partition-Check-3-Majority with the ones of our implementation via Sat as described in Section 5.1.[6] It turns out that Sat-Check-3-Majority offers significantly better running times (see Table 3), even though it is much more universal. Moreover, Sat-Check-k-Majority directly returns a resulting preference profile with k voters.

Further runtimes, which exhibit the practical power of our SAT approach (and its limits), can be obtained from Table 4. All experiments were run on an Intel Core i5, 2.66 GHz (quad-core) machine with 12 GB RAM using the SAT solver plingeling [5]. Interestingly, an integer programming (IP) formulation of the problem by Eggermont et al. [23] appears to perform worse than our SAT-based formalization: Eggermont et al. [23] report that for $n > 20$ runtimes are prohibitively large.

[6] As a programming language, Java was used in both cases.

Algorithm 2: 2-PARTITION-CHECK-3-MAJORITY.

```
Input: tournament (V, E)
Output: whether (V, E) is a 3-majority digraph
foreach 2-partition {E1, E2} of E do
    if E1 is transitive and E2 is acyclic and E2 has a transitive reorientation then
        return true;
    end
end
return false;
```

Table 3
Comparison of average runtimes of SAT-CHECK-3-MAJORITY and 2-PARTITION-CHECK-3-MAJORITY for randomly sampled tournaments of size n with a cutoff time of one hour.

Algorithm	5	6	7	8	9	10	20	50	100
SAT	< 0.1s	< 0.1s	< 0.1s	< 0.1s	< 0.1s	< 0.1s	0.1s	1.5s	12.5s
2-PARTITION	< 0.1s	< 0.1s	0.1s	530s	–	–	–	–	–

Table 4
Runtime in seconds of SAT-CHECK-k-MAJORITY for different number of alternatives and different number of voters k when average runtimes did not exceed 20 seconds. For this table, averages were taken over 5 samples from the uniform random tournament model.

$n\backslash k$	3	4	5	6	7	8	9	10	11	12
3	.04	.04	.03	.04	.04	.04	.04	.05	.08	.10
4	.03	.04	.03	.04	.04	.04	.05	.07	.10	.18
5	.03	.04	.03	.04	.06	.05	.06	.09	.16	.35
6	.03	.04	.04	.04	.05	.06	.08	.12	.27	.63
7	.04	.04	.04	.05	.05	.07	.10	.17	.45	1.10
8	.04	.05	.05	.05	.07	.08	.13	.23	.69	1.80
9	.04	.05	.05	.64	.07	.10	.17	.33	1.06	2.83
10	.05	.05	.06	.67	.09	.12	.23	.46	1.56	4.25
11	.06	.06	.06	1.92	.10	.14	.30	.63	2.22	6.37
12	.06	.07	.07	3.35	.12	.19	.40	.85	3.18	8.48
13	.07	.07	.09	3.93	.15	.27	.52	1.16	4.44	12.30
14	.07	.09	.10	4.15	.18	.36	.64	1.51	5.99	16.84
15	.08	.10	.13	3.89	.21	.88	.79	2.22	7.67	–
16	.09	.11	.14	4.12	.25	4.55	.99	2.90	9.80	–
17	.10	.12	.19	4.41	.29	7.15	1.23	4.69	12.48	–
18	.11	.14	.23	4.76	.35	17.51	1.53	8.25	15.97	–
19	.12	.15	.35	4.97	.43	–	1.80	–	19.99	–
20	.13	.17	.54	5.04	.47	–	2.21	–	–	–
21	.14	.18	5.87	6.15	.63	–	2.71	–	–	–
22	.16	.20	11.07	5.43	.96	–	3.24	–	–	–
23	.17	.23	18.95	5.76	1.57	–	4.12	–	–	–
24	.20	.26	–	5.87	2.56	–	4.60	–	–	–
25	.22	.29	–	6.12	4.21	–	5.85	–	–	–

6. Analyzing majority dimensions

Using the algorithm described in the previous section, we are now in a position to analyze the majority dimension of digraphs. In this section, we report on our findings for different sources of digraphs.

6.1. Exhaustive analysis

Using the tournament generator from the NAUTY toolkit [47], we generated all tournaments with up to 10 alternatives and found that all of them are 5-inducible. In fact, all tournaments of size up to seven are even 3-inducible, confirming a conjecture by Shepardson and Tovey [51]. Shepardson and Tovey [51] also showed that there exist tournaments of size 8 that are not 3-inducible. We also found that the exact number of such tournaments is 96 (out of 6880), replicating a result of Eggermont et al. [23]. One of these tournaments is depicted in Fig. 2.

Like Eggermont et al. [23], we have not encountered a single tournament for which we could show that it is not 5-inducible. Since quadratic residue tournaments of enormous size are the only concrete tournament of which we know that they have higher majority dimension (see Section 3), we examined small tournaments of this kind as well and found that

$$\dim(Q_3) = \dim(Q_7) = 3 \quad \text{and} \quad \dim(Q_{11}) = \dim(Q_{19}) = 5.$$

Unfortunately, we were not able to check whether the majority dimension of Q_{23} is equal to 5 or larger as the SAT solver did not terminate within a total of six weeks.[7]

6.2. Empirical analysis

In the preference library PrefLib [42], scholars have contributed data sets from real world scenarios ranging from preferences over movies or sushi to real election data, and even Formula 1 championship results. The number of voters whose preferences originally induced these data sets vary heavily between 4 and 44,000. At the time of writing, PrefLib contained 354 tournaments induced from pairwise majority comparisons as well as 185 incomplete majority digraphs.

Among the tournaments in PrefLib, 58 are 3-inducible. The two largest tournaments in the data set have 240 and 242 vertices, respectively. The first one is a 5-majority tournament and the Sat solver did not terminate on the second one within one day. The remaining tournaments are transitive and thus 1-inducible. Therefore, all checkable tournaments in PrefLib are 5-inducible.

For the non-complete majority digraphs in PrefLib, we found that the indifference constraints which are imposed on missing arcs change the picture. In comparison to tournaments, Sat-Check-k-Majority is slower on non-complete digraphs, and so we had to restrict our attention to instances with at most 40 alternatives. Non-complete digraphs also had higher majority dimension: among the 85 instances on which computation was feasible, we found examples with a majority dimension of up to 8.

6.3. Stochastic analysis

Additionally, we consider *stochastic* models to generate tournaments of a given size n. Many different models for linear preferences (or orderings) have been proposed in the literature. We refer the interested reader to Critchlow et al. [20], Marden [41], Mattei et al. [43], and Brandt and Seedig [7]. In this work, we decided to examine tournaments generated with five different stochastic models.

In the *uniform random tournament model*, the same probability is assigned to each *labeled* tournament of size n, i.e.,

$$\Pr(T) = 2^{-\binom{n}{2}} \text{ for each } T \text{ with } |T| = n.$$

In all of the remaining models, we sample preference profiles and work with the tournament induced by the majority relation. In accordance with McCabe-Dansted and Slinko [44] and Brandt and Seedig [7], we generated profiles with 51 voters.

The *impartial culture model* (IC) is the most widely-studied model for individual preferences in social choice. It assumes that every possible preference ordering has the same probability of $\frac{1}{n!}$. If we add anonymity by having indistinguishable voters, the set of profiles is partitioned into equivalence classes. In the *impartial anonymous culture* (IAC), each of these equivalence classes is chosen with equal probability.

In the *Mallows-ϕ model* [40], the distance to a reference ranking is measured by means of the Kendall-tau distance which counts the number of pairwise disagreements. Let R_0 be the reference ranking. Then, the Kendall-tau distance of a preference ranking R to R_0 is

$$\tau(R, R_0) = \binom{n}{2} - (|R \cap R_0| - n).$$

According to the model, this induces the probability of a voter having R as his preferences to be

$$\Pr(R) = \frac{\phi^{\tau(R,R_0)}}{C}$$

where C is a normalization constant and $\phi \in (0, 1]$ is a dispersion parameter. Small values for ϕ put most of the probability on rankings very close to R_0. For $\phi = 1$, the model coincides with IC.[8]

A very different kind of model is the *spatial model*. Here, alternatives and voters are placed uniformly at random in multi-dimensional space and the voters' preferences are determined by the (Euclidean) distances to the alternatives. The spatial model plays an important role in political and social choice theory where the dimensions are interpreted as different aspects or properties of the alternatives (see, e.g., [48,3]).

[7] Another specific tournament that we considered is a tournament on 24 alternatives used by Brandt et al. [14] to disprove a conjecture in social choice theory [10]. We found that this tournament is 5-inducible, which implies that the negative consequences of the counterexample already hold for settings with only 5 voters (and at least 24 alternatives).

[8] An interpretation of distance-based models such as Mallows-ϕ model is that there exists a pre-existing truth in the form of a reference ordering and the agents report noisy estimates of said truth as their preferences. For these models, Laslier [39] has introduced the term *Rousseauist cultures* [39].

Table 5
Average majority dimension in tournaments generated by stochastic (preference) models. The given values are averaged over 30 samples each.

n	Uniform	IC	IAC	Mallows-ϕ ($\phi = 0.95$)	Spatial (dim = 2)
3	1.40	1.13	1.13	1.13	1.00
5	3.00	1.67	2.13	1.33	1.13
7	3.00	2.67	2.67	2.47	1.33
9	3.13	3.00	3.00	2.67	1.60
11	3.93	3.07	3.00	2.87	2.33
13	4.80	3.07	3.20	2.93	2.53
15	5.00	3.27	3.40	3.00	2.67
17	5.00	3.40	3.80	2.93	2.80
19	5.00	4.27	4.20	3.00	2.80
21	5.00	4.47	4.33	3.00	2.87

For up to 21 alternatives, we sampled preference profiles (each consisting of 51 voters[9]) from the aforementioned stochastic models and examined the corresponding majority digraphs for their majority dimension using Sat-Check-k-Majority. For each size, Table 5 shows the majority dimension averaged over 30 instances that we sampled. We see that the unbiased models (IC, IAC, uniform) tend to induce digraphs with higher majority dimension.

Again, we encountered no tournament that was not a 5-majority tournament, even after additionally checking more than 8 million uniform random tournaments with 12 alternatives.

7. Hardness of voting with few voters

In this section, we show that the winner determination problem of six well-studied voting rules remains NP-hard even if the number of voters is a small constant. Our general method is to analyze existing hardness constructions for these rules with respect to their susceptibility to the sufficient conditions in Lemma 8 or Lemma 9. In all cases, we slightly modify the hardness constructions to get better bounds on the number of voters. Before we proceed, we introduce two new constrained classes of propositional formulae (Ordered3-CNF, to be used for the results in Sections 7.1 and 7.2, and ReducedFew-CNF, to be used for the result in Section 7.3) and show for both that the problem of deciding whether a given formula is satisfiable is NP-complete.

A formula of propositional logic in conjunctive normal form (CNF) is in 3-CNF if each clause has at most three literals. We say that a formula φ from 3-CNF is in Ordered3-CNF if its clauses all contain exactly three distinct literals and are ordered within φ in such a way that for each propositional variable p, all clauses containing the literal p precede all clauses containing $\neg p$. It is known that 3Sat, the problem of deciding whether a given formula in 3-CNF is satisfiable, is NP-complete [35]. For formulae in Ordered3-CNF, we call the corresponding decision problem Ordered3Sat.

Lemma 11. Ordered3Sat *is NP-complete.*

Proof. Membership in NP is obvious. For hardness, we reduce from 3Sat. Let φ be some formula in 3-CNF. Let P denote the set of variables of the propositional language in which φ is formulated and let $C = (c_1, \dots, c_{|C|})$ denote the clause set of φ. We may assume without loss of generality that no clause contains the same variable twice, that all literals in a clause are ordered according to a fixed ordering $(p_1, p_2, \dots)$, and that every clause is of size three. The latter is due to the fact that clauses of size one can be easily used to simplify φ and the remaining clauses $(p \vee q)$ of size two can be padded with a new variable x to $(p \vee q \vee x) \wedge (p \vee q \vee \neg x)$. We call all variables that occur at least once in φ *original* variables.

For the reduction, we construct an ordered formula φ' in 3-CNF with $6 \cdot |C|$ clauses and $4 \cdot |C|$ additional variables that is satisfiable if and only if φ is. For every clause $c_i = (\ell_1 \vee \ell_2 \vee \ell_3)$, define a set of new propositional variables x_i, x'_i, y_i, and z_i, and a set of new clauses $\varphi_i = \bigwedge_{j=1}^{6} c_i^j$ with

$$
\begin{aligned}
c_i^1 &= (\ell_1 \vee x_i \vee x'_i), & c_i^2 &= (\ell_2 \vee \neg x_i \vee y_i),\\
c_i^3 &= (\ell_2 \vee \neg x'_i \vee y_i), & c_i^4 &= (\ell_3 \vee \neg y_i \vee z_i),\\
c_i^5 &= (\neg x_i \vee \neg y_i \vee \neg z_i), \text{ and} & c_i^6 &= (\neg x'_i \vee \neg y_i \vee \neg z_i)
\end{aligned}
$$

These six clauses will replace c_i in the constructed formula φ'. It is easy to check that a variable assignment to the original variables satisfies c_i if and only if there is a way to extend this variable assignment to the new variables such that φ_i is satisfied. Hence $\bigwedge_i \varphi_i$ is satisfiable if and only if φ is.

[9] In another study [7], this size turned out to be sufficiently large to discriminate the different underlying stochastic models.

What remains to be shown is that the clauses c_i^j of the formula $\bigwedge_i \varphi_i$ can be arranged in such a way that the resulting formula is ordered. To this end, we partition the set of these clauses into subsets. For each original variable p and $j \in \{1, \dots, 4\}$, define the clause sets

$$C^{p,j} = \bigcup_i \{c_i^j : p \in c_i^j\} \qquad \text{and} \qquad C^{\neg p,j} = \bigcup_i \{c_i^j : \neg p \in c_i^j\}.$$

Also define

$$C^5 = \bigcup_i c_i^5 \qquad \text{and} \qquad C^6 = \bigcup_i c_i^6.$$

We are now in a position to define φ' to be

$$\varphi' = \bigwedge_{i=1}^{|P|} \left(\left(\bigwedge_{j=1}^{4} \bigwedge_{c \in C^{p_i,j}} c \right) \wedge \left(\bigwedge_{j=1}^{4} \bigwedge_{c \in C^{\neg p_i,j}} c \right) \right) \wedge \bigwedge_{c \in C^5 \cup C^6} c.$$

We claim that φ' is ordered. We show this for original and new variables separately. For each original variable p, all positive occurrences are in the $C^{p,j}$, preceding the negative occurrences in the $C^{\neg p,j}$.

For new variables, we consider each original clause $c_i = (\ell_1 \vee \ell_2 \vee \ell_3)$ separately. Because its literals are ordered according to the fixed ordering of P, the clause c_i^1 (corresponding to ℓ_1) occurs in φ' before clauses c_i^2 and c_i^3 (corresponding to ℓ_2), which occur before c_i^4 (corresponding to ℓ_3). Indeed, from the definition of φ', the clauses in φ_i occur in φ' in the order $c_i^1, c_i^2, c_i^3, c_i^4, c_i^5, c_i^6$. Hence, for the new variables occurring in these clauses, positive occurrences precede negative occurrences in φ'. $\square$

We say that a formula from 3-CNF is in Few-CNF if each literal appears at most twice, and each variable appears at most thrice. We call the problem of checking whether a formula given in Few-CNF is satisfiable FewSat. Tovey [54] has shown that FewSat is NP-complete [54, Thm. 2.1]. We follow his proof to show that this still holds for formulae in ReducedFew-CNF where we additionally require that every variable occurs in at most one three-literal clause and every literal in at most one two-literal clause. Denote the corresponding decision problem by ReducedFewSat.

Lemma 12. ReducedFewSat *is NP-complete.*

Proof. Membership in NP is obvious. For hardness, we reduce from 3Sat. Let $\varphi := \bigwedge_{i=1}^n (x_i \vee y_i \vee z_i)$ be some formula in 3-CNF where no clause contains the same variable twice. For every variable v occurring in φ, replace each of its L occurrences with a new variable v_j where $1 \le j \le L$. Now for every v occurring in φ, add the clauses

$$\varphi_v = (\neg v_L \vee v_1) \wedge \bigwedge_{j=1}^{L-1} (\neg v_j \vee v_{j+1})$$

which are equivalent to $(v_L \Rightarrow v_1) \wedge \bigwedge_{j=1}^{L-1} (v_j \Rightarrow v_{j+1})$. Call the formula resulting from these transformations $red(\varphi)$. Note that $red(\varphi)$ only contains clauses with three literals (original clauses with replaced variables) or two literals (the new clauses); denote these clause sets by C^3 and C^2, respectively. Also observe that every variable occurs exactly once in C_3 and every literal exactly once in C^2, i.e., $red(\varphi)$ is in ReducedFew-CNF.

For every old variable v, we can only satisfy φ_v by setting all v_j to the same value. Since setting all v_j to the same value t satisfies φ_v and has the same effect on the original part of $red(\varphi)$ that setting v to t has on φ, it follows that φ is satisfiable if and only if $red(\varphi)$ is satisfiable. $\square$

7.1. The Banks set

The *Banks set* associates with each majority tournament the maximal elements of its maximal (with respect to set-inclusion) transitive subtournaments (see, e.g., [38,11]).

Although *some* alternative in the Banks set can be found in polynomial time using a greedy algorithm [32], deciding whether a *specific* alternative belongs to the Banks set is NP-complete as shown by Woeginger [56] by a reduction from 3-colorability. Brandt et al. [8] gave an arguably simpler proof of this result by a reduction from 3Sat: every formula φ in 3-CNF can be transformed in polynomial time into a tournament T_φ^{BA} with a decision vertex c_0 such that c_0 is in the Banks set of T_φ^{BA} if and only if φ is satisfiable. Due to Lemma 11, this reduction works just as well if φ is assumed to be ordered. Again, we have P denote the set of variables of the propositional language in which φ is formulated.

A tournament (V, E) is in *the class* $\mathcal{G}^{BA}$ if it satisfies the following properties. There is an odd integer m such that,

$$V = C \cup U_1 \cup \dots \cup U_m,$$

where $C, U_1, \ldots, U_m$ are pairwise disjoint and $C = \{c_0, \ldots, c_m\}$. We have C_i denote the singleton $\{c_i\}$ and $U = \bigcup_{i=1}^m U_i$. If i is odd, $U_i = \{u_i^1, u_i^2, u_i^3\}$ whereas if i is even U_i is a singleton $\{u_i\}$. Let $X = \bigcup\{U_i : i \text{ is odd}\}$ and $Y = \bigcup\{U_i : i \text{ is even}\}$. Intuitively, (V, E) is T_φ^{BA} for some φ in ordered 3-CNF with $\frac{1}{2}(m+1)$ clauses. If i is odd, U_i corresponds to a clause of φ and the vertices it contains represent (tokens of) literals. We assume each of these vertices u_i^j to be labeled by the literal $\lambda(u_i^j)$ it represents. For *odd* $i \in \{1, \ldots, m\}$ and $j \in \{1, 2, 3\}$ we define,

$$U_i^j = \{u_i^j\},$$
$$U_i^p = \{u \in U_i : \lambda(u) = p\}, \text{ and}$$
$$U_i^{\neg p} = \{u \in U_i : \lambda(u) = \neg p\}.$$

Moreover, for *even* $i \in \{1, \ldots, m\}$ and $j \in \{1, 2, 3\}$, we let

$$U_i^j = U_i^p = U_i^{\neg p} = \emptyset.$$

Observe that $\bigcup_{\substack{p \in P\\ 1 \le i \le m}} (U_i^p \cup U_i^{\neg p}) = X$.

We are now in a position to define the arc set E, almost as in Brandt et al. [8].[10] Let

$$\begin{aligned} E = {} & \bigcup_{i<j} (C_j \times C_i) \cup \bigcup_{i<j} \big((U_i \times U_j) \setminus \overline{E^\varphi}\big) \cup \\ & \bigcup_{1 \le i \le m} \big((U_i^1 \times U_i^2) \cup (U_i^2 \times U_i^3) \cup (U_i^1 \times U_i^3)\big) \cup \\ & \bigcup_{i \ne j} (C_i \times U_j) \cup \bigcup_i (U_i \times C_i) \cup E^\varphi, \end{aligned}$$

where

$$E^\varphi = \bigcup_{\substack{p \in P\\ i<j}} (U_j^p \times U_i^{\neg p}) \cup \bigcup_{\substack{p \in P\\ i<j}} (U_j^{\neg p} \times U_i^p).$$

Fig. 6 illustrates this type of tournament. The set E^φ is the part of the tournament T_φ^{BA} that depends on the input formula. The arc set

$$(E \setminus E^\varphi) \cup \overline{E^\varphi}$$

we refer to as its *skeleton*.

We will show that the skeleton of each tournament T_φ^{BA} is induced by a 3-voter profile such that the arcs in $\overline{E^\varphi}$ all get a weight of one. At the same time, E^φ is 2-inducible such that the weight on all arcs is two. A little reasoning and an application of Lemma 9 then gives us the desired result.

Theorem 2. *Computing the Banks set is NP-hard if the number of voters is at least* 5.

Proof. Let (V, E) be a tournament in $\mathcal{G}^{BA}$. It suffices to show that (V, E) is induced by a 5-voter profile. To this end define

$$E_1 = \bigcup_i (U_i \times C_i),$$
$$E_2 = E^\varphi, \text{ and}$$
$$E_3 = E \setminus (E_1 \cup E_2).$$

Observe that $E = E_1 \cup E_2 \cup E_3$ and that E_1, E_2, and E_3 are pairwise disjoint. In virtue of Lemma 9, it therefore suffices to show that (V, E_1) and (V, E_2) are induced by 2-voter profiles and that (V, E_3) is acyclic.

For (V, E_1) it is easy to see that it is a union of unidirected stars and therefore 2-inducible. For (V, E_2), let

$$E_2^p = \bigcup_{i<j} \left(U_j^p \times U_i^{\neg p}\right) \cup \left(U_j^{\neg p} \times U_i^p\right)$$

[10] There is only a slight change compared to the original construction by Brandt et al. [8]. Specifically, we now have arcs $U_i^1 \times U_i^3$ instead of the other way around. It is not difficult to check that the argument of the reduction is not affected—it is irrelevant whether the crucial transitive subtournament with c_0 as its maximal element may contain one, two, or three vertices from each U_i.

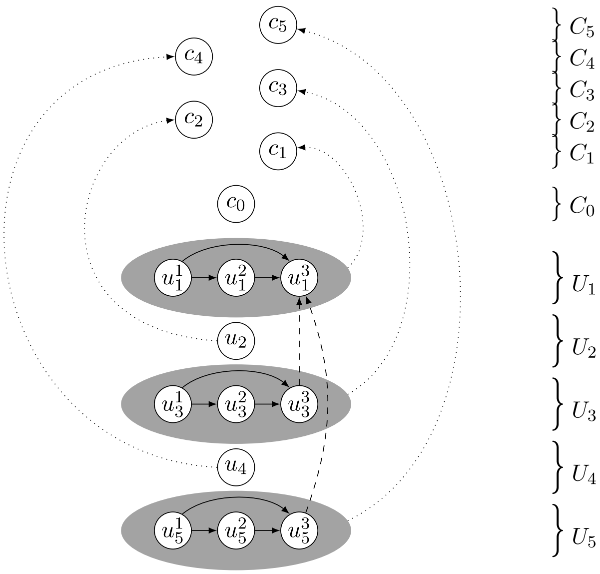


Fig. 6. A tournament $T_\varphi^{BA} = (V, E)$ in the class $\mathcal{G}^{BA}$. Omitted arcs point downwards. Moreover, $\lambda(u_5^3) = \lambda(u_3^3) = \overline{\lambda}(u_1^3)$. The dotted and dashed upward arcs correspond to the arc sets E_1 and E_2 in Theorem 2, respectively. The remaining arcs, i.e., all downward arcs and the arcs within the U_i form an acyclic arc set and correspond to E_3.

be the arcs in E_2 associated with a variable p. Note that $E_2 = \bigcup_{p\in P} E_2^p$ and that all E_2^p are vertex-disjoint from each other. Recall that (V, E) was in induced through a construction that was based on an *ordered* formula. This implies that whenever $U_j^p \neq \emptyset \neq U_i^{\neg p}$ we have that i is greater than j. Therefore, E_2^p can also be written as $\bigcup_{i,j}(U_i^{\neg p} \times U_j^p)$. In this representation, it is clear that E_2^p is a unidirected bipartite digraph. Thus, E_2 is a *bilevel graph* and hence 2-inducible by Lemmas 5 and 6.

To see that E_3 is acyclic, note that it forms a subset of

$$C \times U \cup \bigcup_{i<j}(U_i \times U_j) \cup \bigcup_{i>j}(C_i \times C_j) \cup \bigcup_{1\le i\le m} \left((U_i^1 \times U_i^2) \cup (U_i^2 \times U_i^3) \cup (U_i^1 \times U_i^3)\right)$$

and corresponds to all (shown) horizontal and (missing) downward arcs in Fig. 6. $\square$

7.2. The tournament equilibrium set and the minimal extending set

The *tournament equilibrium set* (*TEQ*) is another voting rule that, like the Banks set, selects a subset of alternatives from each tournament (see, e.g., [38,11]). Its recursive definition is based on the notion of retentiveness. Given a tournament (V, E), a non-empty subset $X \subseteq V$ is said to be *TEQ-retentive* if for all $v \in X$ all alternatives chosen by *TEQ* from the subtournament of (V, E) induced by $\{w \in V : (w, v) \in E\}$ are contained in X. *TEQ* is then defined so as to select the union of inclusion-minimal *TEQ*-retentive sets from each tournament.

A related voting rule called the *minimal extending set* (*ME*) was proposed by Brandt [6]. It is defined as the union of inclusion-minimal sets $X \subseteq V$ such that for all $v \in V \setminus X$, v is not contained in the Banks set of the subtournament of (V, E) induced by $X \cup \{v\}$.

Using the same construction, it was shown that both computing *TEQ* and computing *ME* is NP-hard by reduction from 3Sat [8,13]. In virtue of Lemma 11, this construction is also a valid reduction from Ordered3Sat. For every formula φ in ordered 3-CNF a tournament T_φ^{TEQ} can be constructed such that *TEQ* (and *ME*) selects a decision vertex c_0 from T_φ^{TEQ} if and only if φ is satisfiable. The class of these tournaments T_φ^{TEQ} is denoted by $\mathcal{G}^{TEQ}$. The tournaments in $\mathcal{G}^{TEQ}$ bear a strong structural similarity to those in $\mathcal{G}^{BA}$, which can be exploited to show that every tournament in $\mathcal{G}^{TEQ}$ is induced by a 7-voter profile.

A tournament (V, E) is in *the class* $\mathcal{G}^{TEQ}$ if it satisfies the following properties. There is an odd integer m with $m \equiv 1 \pmod 4$ such that,

$$V = C \cup U_1 \cup \cdots \cup U_m,$$

where $C, U_1, \ldots, U_m$ are defined the same as in $\mathcal{G}^{BA}$. We have C_i denote the singleton $\{c_i\}$. Moreover, let $X = \bigcup\{U_i : i \equiv 1 \pmod 4\}$, $Y = \bigcup\{U_i : i \text{ is even}\}$, and $Z = \bigcup\{U_i : i \equiv 3 \pmod 4\}$.

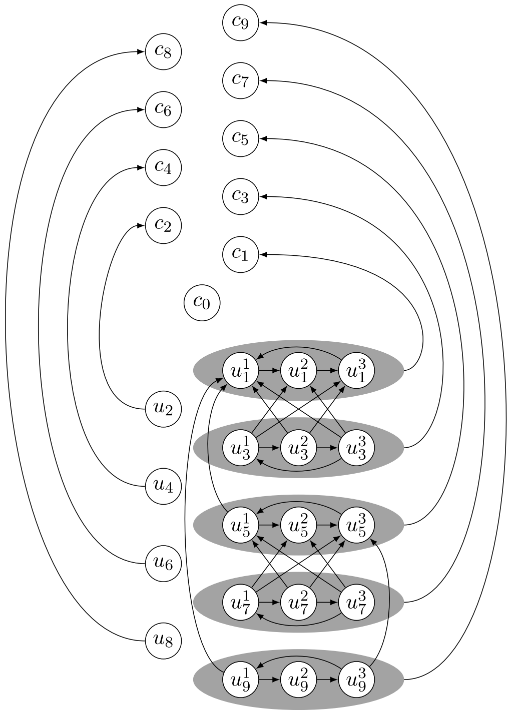


Fig. 7. A tournament $T_\varphi^{TEQ} = (V, E)$ in the class $\mathcal{G}^{TEQ}$. Omitted arcs point downwards.

Intuitively, (V, E) is T_φ^{TEQ} for some φ in ordered 3-CNF with $\frac{1}{4}(m+3)$ clauses. Every $U_i \in X$ corresponds to a clause of φ and the vertices it contains represent (tokens of) literals. Again, we assume each of these vertices u_i^j to be labeled by the literal $\lambda(u_i^j)$ it represents. For $i \in \{1, 5, \dots, m\}$ and $j \in \{1, 2, 3\}$ we define,

$$
\begin{aligned}
U_i^j &= \{u_i^j\}, \\
U_i^p &= \{u \in U_i : \lambda(u) = p\}, \text{ and} \\
U_i^{\neg p} &= \{u \in U_i : \lambda(u) = \neg p\}.
\end{aligned}
$$

Moreover, for the other values of i, and $j \in \{1, 2, 3\}$, we stipulate,

$$U_i^j = U_i^p = U_i^{\neg p} = \emptyset.$$

Observe that $\bigcup_{\substack{p \in P \\ 1 \le i \le m}} (U_i^p \cup U_i^{\neg p}) = X$.

We are now in a position to define the arc set E.

$$
\begin{aligned}
E = {} & \bigcup_{i<j} (C_j \times C_i) \cup \bigcup_{i \neq j} (C_i \times U_j) \cup \bigcup_{i=j} (U_j \times C_i) \cup \\
& \bigcup_{1 \le i \le m} \big((U_i^1 \times U_i^2) \cup (U_i^2 \times U_i^3) \cup (U_i^3 \times U_i^1)\big) \cup \\
& \bigcup_{i<j} \big((U_i \times U_j) \setminus (\overline{E^\varphi} \cup \overline{E^z})\big) \cup E^\varphi \cup E^z,
\end{aligned}
$$

where

$$E^\varphi = \bigcup_{\substack{p \in P \\ i>j}} (U_i^p \times U_j^{\neg p}) \cup \bigcup_{\substack{p \in P \\ i>j}} (U_i^{\neg p} \times U_j^p) \qquad \text{and} \qquad E^z = \bigcup_{\substack{l \neq l' \\ i=j+2}} (U_i^l \times U_j^{l'}).$$

An example of such a tournament is depicted in Fig. 7. The notable structural differences to $\mathcal{G}^{BA}$ are the cycles in U_i for odd i and the arcs E^z between Z and X. Next, we show that every tournament T_φ^{TEQ} is induced by a 7-voter profile, using the same approach as in Theorem 2.

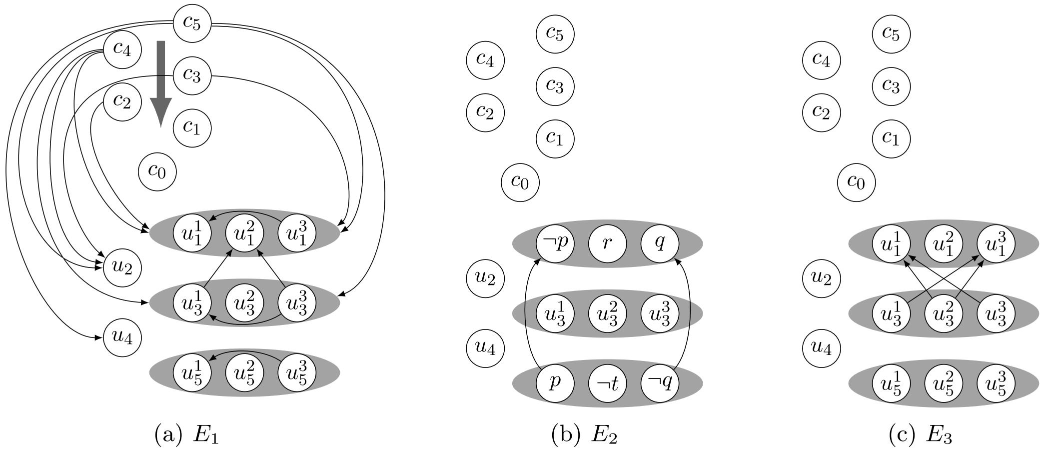


Fig. 8. Illustration of the arc sets $E_1, E_2, E_3 \subset E$ in $T_\varphi^{TEQ} = (V, E)$. The thick arrow on the left represents all arcs $\bigcup_{i<j}\{(c_j, c_i)\}$ being part of E_1.

Theorem 3. *The problems of computing the tournament equilibrium set and of computing the minimal extending set are NP-hard if the number of voters is at least* 7.

Proof. Similar to the proof for Theorem 2, it suffices to show that every tournament (V, E) in $\mathcal{G}^{TEQ}$ is induced by a 7-voter profile. To achieve this, we partition E into four disjoint arc sets $E_1, E_2, E_3, E_4 \subseteq E$ and show that the digraphs (V, E_1), (V, E_2), and (V, E_3) are each induced by 2-voter profiles as well as that (V, E_4) is acyclic. Then the result follows from Lemma 9.

While the tournaments in $\mathcal{G}^{TEQ}$ are very similar to the ones in $\mathcal{G}^{BA}$, the introduction of new vertices and arcs makes finding an appealing partition a bit trickier. We define

$$E_1 = \bigcup_{i>j}\big(C_i \times (C_j \cup U_j)\big) \cup \bigcup_i (U_i^3 \times U_i^1) \cup \bigcup_{\substack{i\equiv 3 \\ \text{mod } 4}} \big((U_i^1 \cup U_i^3) \times U_{i-2}^2\big),$$

$$E_2 = E^\varphi,$$

$$E_3 = E^z \setminus E_1, \text{ and}$$

$$E_4 = E \setminus (E_1 \cup E_2 \cup E_3).$$

It can readily be appreciated that E_1, E_2, and E_3 are contained in E (see Fig. 8). Also, they are pairwise disjoint and therefore $\{E_1, E_2, E_3, E_4\}$ is a proper partition of E.

To show that (V, E_1) is 2-inducible, we define

$$E_1' = \bigcup_{\substack{i\le j \\ U_j \subset X\cup Z}} (C_i \times U_j) \cup \bigcup_{\substack{i\le j \\ U_j\subset Y}} (U_j \times C_i) \cup \bigcup_{\substack{i<j \\ U_i, U_j \subset X\cup Z}} \big((U_i \times U_j) \setminus E_1\big) \cup \bigcup_{\substack{i<j \\ U_i,U_j\subset Y}} (U_j \times U_i) \cup$$
$$\bigcup_{i \text{ odd}} \big((U_i^1 \cup U_i^3) \times U_i^2\big) \cup (Y \times (X \cup Z)).$$

It is straightforward to check that E_1' is a reorientation of $\tilde{E}_1$. Also, it is easy but tedious, by making the obvious case distinctions, to show for E_1 and E_1' that the out-neighborhood of each vertex is contained in the out-neighborhood of each of its in-neighbors, implying that E_1 and E_1' are both transitive.[11] For example, consider a vertex $u_i^1 \in X$ in E_1' for which

$$D = U_i^2 \cup \bigcup_{\substack{j>i \\ j \text{ odd}}} U_j \qquad \text{and} \qquad \overline{D} = Y \cup \bigcup_{\substack{j<i \\ j \text{ odd}}} U_j \cup \bigcup_{j\le i} C_j$$

denote the set of all out-neighbors and all in-neighbors of u_i^1 in (V, E_1'), respectively. It is straightforward to check that every vertex in $\overline{D}$ also has an arc in E_1' to every vertex in D.

Thus, in virtue of Lemma 3, (V, E_1) is induced by a 2-voter profile.

[11] In a digraph (V, E), the out-neighbors of a vertex x are given by $D_x = \{y \in V : (x, y) \in E\}$. Analogously, the in-neighbors of x are defined as $\overline{D}_x = \{y \in V : (y, x) \in E\}$.

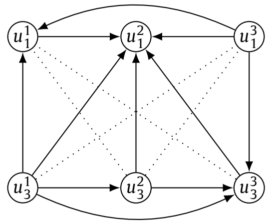


Fig. 9. The arc set $E'_{3,3}$ which is part of the reorientation E'_3 of $\tilde{E}_3$ in the proof of Theorem 3. Dotted arcs denote the incomparability subgraph of $E'_{3,3}$.

The proof for (V, E_2) being 2-inducible is analogous to the proof of the same statement in the Banks construction (see Theorem 2). This is also where the orderedness of φ is exploited.

The digraph (V, E_3) is obviously transitive. We also observe that it consists of isomorphic and vertex-disjoint subgraphs $(U_i \cup U_{i-2}, E_{3,i})$ for $i \equiv 3 \pmod 4$ with $E_i = (U_i^l \times U_{i-2}^{l'})$ for $l \neq l'$. It is sufficient to find a general transitive reorientation $E'_{3,i}$ on such a subgraph because then every completion of $\bigcup_{i \equiv 3 (\mathrm{mod}\ 4)} E'_{3,i}$ is a transitive reorientation of $\tilde{E}_3$. We define

$$E'_{3,i} = (U_i \times U_{i-2}^2) \cup \left((U_{i-2}^1 \cup U_{i-2}^3) \times U_{i-2}^2\right) \cup \left((U_{i-2}^3 \cup U_i^1) \times (U_{i-2}^1 \cup U_i^3)\right) \cup (U_i^1 \times U_i^2) \cup (U_i^2 \times U_i^3).$$

This subgraph set is also shown in Fig. 9 and it is easy to verify that it is indeed transitive.

Finally, to see the acyclicity of (V, E_4), observe that

$$E_4 = \bigcup_{i<j} (C_i \times U_j) \cup \bigcup_{i<j} \left((U_i \times U_j) \setminus (\overline{E^{\varphi}} \cup \overline{E^{z}})\right) \cup \bigcup_{i} \left((U_i^1 \times U_i^2) \cup (U_i^2 \times U_i^3)\right)$$

and that E_4 is thereby contained in the transitive closure of the ordering

$$(c_0, u_1^1, u_1^2, u_1^3, c_1, u_2, c_2, u_3^1, u_3^2, u_3^3, c_3, u_4, c_4, u_5^1, \ldots, c_m).$$

This completes the proof. □

7.3. Slater's rule and Kemeny's rule

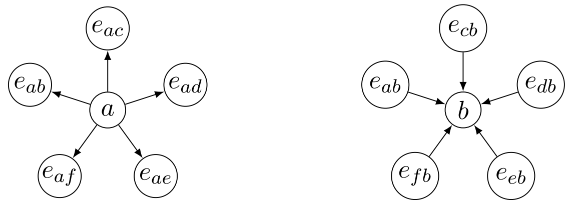
Slater's rule seeks linear rankings over alternatives that minimally conflict with the pairwise majority relation and returns the maximal elements of these rankings (see, e.g., [38,11]). Formally, given a tournament (V, E), the *Slater score* of a linear ranking $\succ$ of V is the number of pairs $(x, y) \in V \times V$ such that both $x \succ y$ and $(x, y) \in E$. A *Slater ranking* is a ranking $\succ$ with maximum Slater score. The *Slater set* consists of all those alternatives $v \in V$ that occur at the top of a Slater ranking. There is a close relationship between Slater rankings and feedback arc sets: maximizing the number of agreed-upon pairwise comparison is the same as minimizing the number of arcs of (V, E) that need to be turned around so as to produce a transitive (and so acyclic) tournament. This connection makes it easy to show that computing Slater rankings is NP-hard in general digraphs, since it is well-known that the feedback arc set problem is NP-hard [29]. Whether the feedback arc set problem restricted to tournaments is NP-hard was a long-standing open problem, which was solved independently by Alon [1], Conitzer [18], and Charbit et al. [16]. As a consequence, we now know that computing Slater rankings and the Slater set is NP-hard for tournaments (see also [34]).

A close relative to Slater's rule is *Kemeny's rule*. While Slater's rule only uses the information contained in the pairwise majority relation, Kemeny's rule also takes into account the magnitude of majority comparison, so that the input to Kemeny's rule is a *weighted* majority tournament. A further difference is that, while Slater's rule is typically used so as to produce a *set of winners*, Kemeny's rule is typically used to find *consensus rankings*. Kemeny's rule has very appealing axiomatic properties [58]. Let us now formally define Kemeny's rule. Given a weighted digraph (V, w), the *Kemeny score* of a linear ranking $\succ$ of V is $\sum_{x \succ y} \mathrm{w}(x, y)$, and a *Kemeny ranking* is a ranking with maximum Kemeny score. The Kemeny rule just returns all Kemeny rankings. Again, notice the close connection to the (weighted) feedback arc set problem. Further, notice that Kemeny's and Slater's rules coincide on tournaments where every arc has weight 1.

Let us now analyze the complexity of these two rules in a setting where there is a constant number of voters. For Kemeny's rule, Dwork et al. [22] showed the problem to be hard even for weighted digraphs induced by a profile of 4 voters. Their reduction contained a small error that was fixed by Biedl et al. [4]. With the tools we developed in Section 3, we can give a short exposition of this reduction.

Theorem 4. *Computing Kemeny's rule is NP-hard if the number of voters is even and at least 4.*

Proof. As we noted above, computing Kemeny's rule is equivalent to solving the feedback arc set problem, so we just need to show that this problem remains hard for digraphs inducible by 4-voter profiles. (The case for even $n > 4$ can be seen by just adding two completely reversed orders to this profile.)



(a) The arcs E_1. (b) The arcs E_2.

Fig. 10. Decomposition of a subdivided digraph into two forests of stars.

We show this by reduction from feedback arc set on general digraphs. Let (V, E) be an instance of this problem. Now produce a new digraph (V', E') from (V, E) by *subdividing* every arc. Thus, for each arc $(a, b) \in E$, introduce a new vertex $e_{ab} \in V'$ and arcs $(a, e_{ab}) \in E'$ and $(e_{ab}, b) \in E'$. Formally, $V' = V \cup S$, where $S = \{e_{ab} : (a, b) \in E\}$ is the set of subdividers, and $E' = \{(a, e_{ab}) : (a, b) \in E\} \cup \{(e_{ab}, b) : (a, b) \in E\}$.

$$a \to b \text{ in } (V, E) \text{ becomes } a \to e_{ab} \to b \text{ in } (V', E').$$

This already completes our description of the reduction. We now claim that (V', E') is 4-inducible, and that the size of the minimum feedback arc set of (V', E') is the same as that of the original graph (V, E).

To see that (V', E') is 4-inducible, we partition its arcs into two arc-disjoint forests of unidirected stars. Therefore, by Lemmas 4, 6, and 8, we deduce that (V', E') is 4-inducible. The promised partition is $E' = E_1 \cup E_2$, where

$$E_1 = E' \cap (V \times S),$$
$$E_2 = E' \cap (S \times V).$$

The set E_1 contains arcs from original vertices to subdividers, while the set E_2 contains arcs from subdividers to original vertices (see Fig. 10).

It is also easy to see that subdivision preserves the size of the minimal feedback arc set. If $F \subseteq E$ is a feedback arc set of (V, E), then the set $F' = \{(a, e_{ab}) : (a, b) \in F\}$ is a feedback arc set of (V', E') of the same size (that is, we delete 'half' of every arc of F). Conversely, a minimal feedback arc set F' of (V', E') will only ever delete one half of an original arc; deleting an arc of (V, E) whenever half of it is deleted in F' gives us a feedback arc set of (V, E) of the same size as F'. $\square$

It is easy to see that Kemeny's rule can be computed in polynomial time for only 2 voters (e.g. the preference rankings of the voters are also optimal Kemeny rankings). Thus, the complexity of Kemeny's rule is settled for every constant *even* number of voters. The complexity for a constant *odd* number of voters was open. We now establish that both Kemeny's and Slater's rules are hard for 7 voters or more.

To do this, we will analyze the reduction by Conitzer [18] showing hardness of the feedback arc set problem restricted to tournaments. Conitzer gives a reduction from MaxSat, which asks for an assignment to the propositional variables in a Boolean formula φ such that at least a given number s_1 of clauses is satisfied. Due to Lemma 12, we can constrain φ to be in ReducedFew-CNF without affecting the correctness of Conitzer's [18] reduction. The reduction is based on tournaments T_φ^{SL} that admit a Slater ranking with at most s_2 inconsistent arcs if and only if an assignment for φ with at least s_1 satisfied clauses exists, where s_2 depends (polynomially) on φ and s_1.

Let $\mathcal{G}^{SL}$ denote the class of all tournaments T_φ^{SL} obtained from a Boolean formula φ in ReducedFew-CNF according to this construction. A tournament (V, E) is in the class $\mathcal{G}^{SL}$ if it satisfies the following properties. There exist integers $m, l \geq 1$, such that

$$V = C \cup \bigcup_{\substack{1 \leq i \leq m \\ 1 \leq j \leq 6}} T_i^j,$$

where C and all T_i^j are pairwise disjoint and for $1 \leq i \leq m$

$$C = \{c_1, \ldots, c_{|C|}\},$$
$$T_i^j = \{t_i^{j,1}, \ldots, t_i^{j,l}\}.$$

Each subtournament $(T_i^j, E \cap (T_i^j \times T_i^j))$ has to be a transitive component, i.e., it is a linear order and for a vertex $v \in V \setminus T_i^j$ and vertices $v_1, v_2 \in T_i^j$, either $\{(v_1, v), (v_2, v)\}$ or $\{(v, v_1), (v, v_2)\}$ have to be in E. For our purposes, we can treat T_i^j as

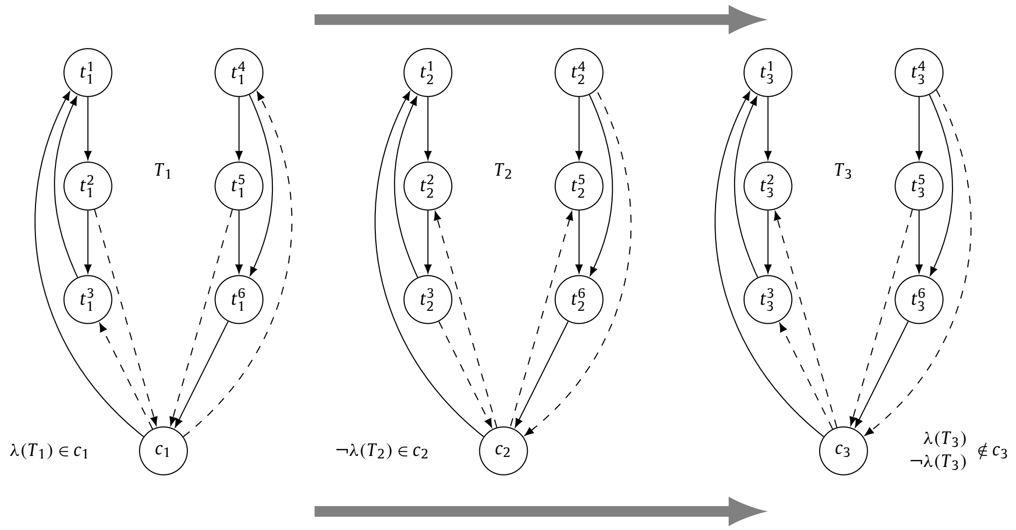


Fig. 11. A schematic of a tournament T_φ^{SL} to illustrate the three different cases for the arcs between $T^2 \cup T^3 \cup T^4 \cup T^5$ and C. These arcs are shown as dashed and are the only ones that depend on φ. The thick arrows below and above indicate the fixed order between and within the T_i^j, and in between the c_i—they stand for the following implicit, undepicted arcs: the arcs (c_i, c_j) for $i < j$, the arcs in $T_i \times T_j$ for $i < j$, and the arcs in $\{t_i^1, t_i^2, t_i^3\} \times \{t_i^4, t_i^5, t_i^6\}$.

a single vertex denoted by t_i^j. Every c_i is associated with a clause in φ. Abusing notation, we denote this clause with c_i as well. Every T_i corresponds to a variable $\lambda(T_i)$ in φ. For notational convenience, let

$$T^j = \bigcup_{1 \le i \le m} t_i^j \quad \text{and} \quad T_i = \bigcup_{1 \le j \le 6} t_i^j.$$

For (V, E) to be in $\mathcal{G}^{SL}$, the arc set has to be of the form

$$E = E_A \cup \bigcup_i \{(t_i^1, t_i^2), (t_i^2, t_i^3), (t_i^3, t_i^1)\} \cup$$
$$(T^6 \times C) \cup (C \times T^1) \cup E^\varphi$$

where

$$E_A = \bigcup_{i<j} \{(c_i, c_j)\} \cup \bigcup_{i<j} (T_i \times T_j) \cup \bigcup_i \bigcup_{\substack{J \in \{4,5,6\} \\ 1 \le j < J}} \{(t_i^j, t_i^J)\}, \text{ and}$$

$$\begin{aligned} E^\varphi = \{&(t_i^2, c_j), (c_j, t_i^3), (c_j, t_i^4), (t_i^5, c_j) \colon \lambda(T_i) \in c_j, c_j \in C\} \cup \\ \{&(c_j, t_i^2), (t_i^3, c_j), (t_i^4, c_j), (c_j, t_i^5) \colon \neg\lambda(T_i) \in c_j, c_j \in C\} \cup \\ \{&(c_j, t_i^2), (c_j, t_i^3), (t_i^4, c_j), (t_i^5, c_j) \colon \lambda(T_i), \neg\lambda(T_i) \notin c_j, c_j \in C\}. \end{aligned}$$

So for every clause c_j and every variable $\lambda(T_i)$, the c_j vertex points to exactly three of the six vertices in T_i, but which three vertices are pointed to depends on whether and how the variable appears in the clause.[12]

An illustration of a tournament in $\mathcal{G}^{SL}$ is depicted in Fig. 11.

Theorem 5. *The problems of computing the Slater set and of computing a Kemeny ranking are NP-hard if the number of voters is at least* 7.

Proof. Let (V, E) be a tournament in $\mathcal{G}^{SL}$ that is constructed from a formula φ in ReducedFew-CNF whose set of clauses $C = C^2 \cup C^3$ is partitioned into clauses in C^2 that contain exactly two literals and clauses in C^3 that contain exactly three literals. We give a transitive (1-inducible) arc set E_1 and 2-inducible arc sets E_2, E_3, E_4 such that putting these four profiles together yields a profile with 7 voters that induces (V, E) as its majority relation, and moreover, so that the majority margins are all equal to 1 (see Fig. 12). This gives the desired result.

[12] The tournaments T_φ^{SL} that we have defined differ very slightly from those used in the original reduction [18]. In the case that $\neg\lambda(T_i) \in c_j$, the original reduction uses arcs $(c_j, t_i^4), (t_i^5, c_j)$, while we direct them as $(t_i^4, c_j), (c_j, t_i^5)$. Conitzer's correctness proof goes through with minor changes to the sentence beginning with "Because d_v and e_v always have arcs into c_k..." [18, Theorem 3].

$$E_1 = E_A \cup \bigcup_i \{(t_i^1, t_i^2), (t_i^2, t_i^3), (t_i^1, t_i^3)\} \cup \left((T^1 \cup \cdots \cup T^6) \times (C^2 \cup C^3)\right)$$

$$E_2 = (C^2 \cup C^3) \times (T^1 \cup T^2 \cup T^3)$$

$$E_3 = \bigcup_{c_j \in C^2} \{(t_i^3, t_i^1), (t_i^2, c_j) \colon \lambda(T_i) \in c_j\} \qquad E_4 = \bigcup_{c_j \in C^3} \{(t_i^2, c_j) \colon \lambda(T_i) \in c_j\}$$

$$\cup \bigcup_{c_j \in C^2} \{(t_i^3, t_i^1), (t_i^3, c_j) \colon \neg\lambda(T_i) \in c_j\} \qquad \cup \bigcup_{c_j \in C^3} \{(t_i^3, c_j) \colon \neg\lambda(T_i) \in c_j\}$$

$$\cup \bigcup_{c_j \in C^3} \{(c_j, t_i^4) \colon \lambda(T_i) \in c_j\} \qquad \cup \bigcup_{c_j \in C^2} \{(c_j, t_i^4) \colon \lambda(T_i) \in c_j\}$$

$$\cup \bigcup_{c_j \in C^3} \{(c_j, t_i^5) \colon \neg\lambda(T_i) \in c_j\} \qquad \cup \bigcup_{c_j \in C^2} \{(c_j, t_i^5) \colon \neg\lambda(T_i) \in c_j\}$$

The set E_1 is complete and transitive: it is induced by the 1-voter profile R_1 whose voter has order $t_1^1 > t_1^2 > t_1^3 > t_1^4 > t_1^5 > t_1^6 > t_2^1 > \cdots > t_m^1 > \cdots > t_m^6 > c_1 > \cdots > c_{|C|}$. The set E_2 is a bilevel graph, and thus 2-inducible (see Lemmas 5 and 6) by a profile R_2 of 2 voters. The sets E_3 and E_4 are also 2-inducible by profiles R_3 and R_4, since E_2 and E_3 both consist of vertex-disjoint unidirected stars (to see disjointness one appeals to the definition of ReducedFew-CNF: every variable occurs at most once in C^3, and every literal occurs at most once in C^2).

Now, notice that the sets E_2, E_3, E_4 are pairwise disjoint, although certain arcs (from $(C^2 \cup C^3) \times (T^2 \cup T^3)$) in E_2 occur once in the opposite direction in E_3 or E_4. Thus, we can see that the union of the profiles R_2, R_3, and R_4 induces the following arcs with weight 2, and all other arcs with weight 0:

$$\begin{aligned}\bigcup_i \{t_i^3, t_i^1\} \cup \{(c_j, t_i^1), (c_j, t_i^3), (c_j, t_i^4) \colon \lambda(T_i) &\in c_j, c_j \in C\} \cup \\ \{(c_j, t_i^1), (c_j, t_i^2), (c_j, t_i^5) \colon &\neg\lambda(T_i) \in c_j, c_j \in C\} \cup \\ \{(c_j, t_i^1), (c_j, t_i^2), (c_j, t_i^3) \colon \lambda(T_i), &\neg\lambda(T_i) \notin c_j, c_j \in C\}.\end{aligned}$$

But this is precisely the set of arcs on which the intended arc set E and the transitive set E_1 disagree! Hence adding R_2, R_3, and R_4 to the 1-voter profile R_1 precisely fixes these disagreements, and induces our target tournament (V, E). Also notice that all majority margins have weight 1. The profile constructed has 7 voters, as required. $\square$

7.4. Ranked pairs

The final voting rule we investigate is called *ranked pairs (RP)* (see, e.g., [28]). Just like Kemeny's rule, it operates on *weighted* digraphs. Hence, we have separate results for an odd and for an even number of voters.

There are two versions of *RP* commonly discussed in the literature. The one we are concerned with is the *neutral* one, i.e., the one that does not differentiate among alternatives. Deciding whether a given alternative is a winner according to this version of *RP* is NP-complete [15].

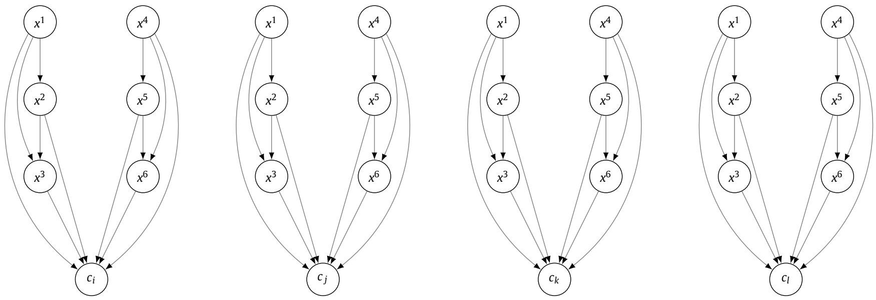
Usually, *RP* is regarded as a procedure. First, one defines priorities for all pairs of alternatives, and then ranks the alternatives iteratively in the order of their priority. The priority over pairs (a, b) of alternatives is given by the number of voters who prefer a to b. To avoid creating cycles, any pair whose addition would yield a cycle is discarded in the procedure. The neutral version of *RP*, which was defined by Tideman [53] and considered by Brill and Fischer [15], returns the set of all rankings generated by the above procedure for *some* tie breaking rule. From this point on, we refer to this variant by *RP*.

The NP-hardness proof by Brill and Fischer [15] is by reduction from Sat. For each Boolean formula φ in CNF they constructed a weighted digraph G_φ^{RP} such that a decision vertex d is selected by *RP* from G_φ^{RP} if and only if φ is satisfiable. The construction, of course, works just as well for a reduction from 3Sat. We may also assume that in every formula φ in 3-CNF no variable occurs more than once in each clause.

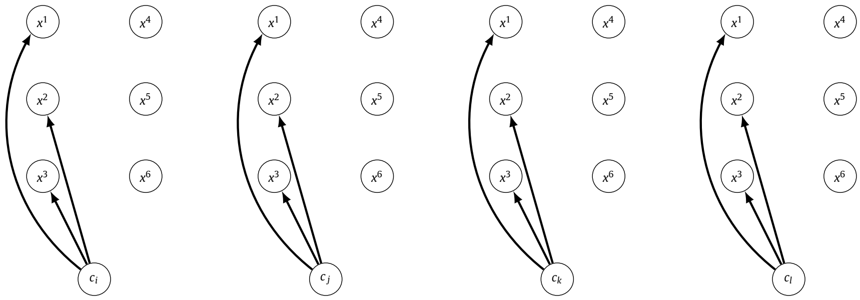
Since the original construction by Brill and Fischer [15] does not yield a tournament, investigating it would give only results involving an even number of voters. However, a minor modification of the argument results in a tournament, which allows to consider the case of an odd number of voters. We first define the class $\mathcal{G}^{RP}$ in which the weighted digraphs G_φ^{RP} for formulae φ in 3-CNF are contained. Then, we prove that every digraph in this class is induced by an 8-voter profile, showing that deciding whether a given alternative is a ranked pairs winner is already NP-complete for eight voters. Later, we define the tournament class $\mathcal{T}^{RP}$ and show the same result for an odd number of voters. Finally we combine these two results into a corollary.

A weighted digraph (V, E) (with weight function w) belongs to $\mathcal{G}^{RP}$ if and only if it satisfies the following conditions. There are some integers $m, l \geq 1$ such that

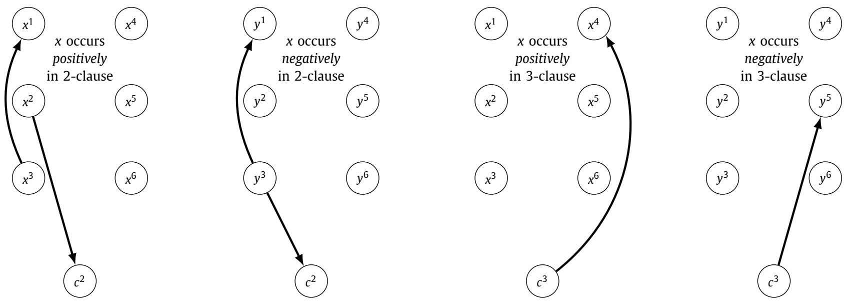
$$V = D \cup U_1 \cup \cdots \cup U_m \cup X_1 \cup \cdots \cup X_l,$$



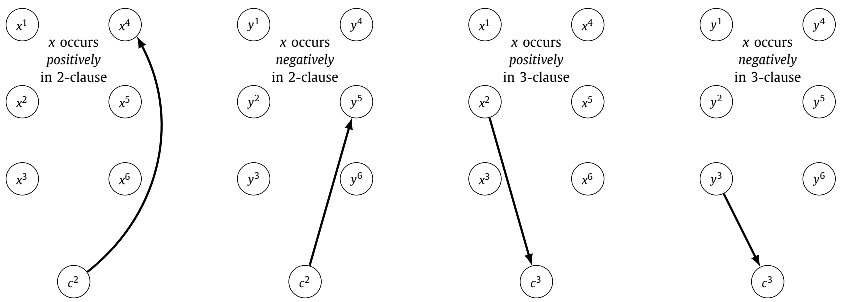
(a) The arcs in E_1 give a transitive tournament, inducible by 1 voter.



(b) The arcs in E_2 form a bilevel graph, and are thus 2-inducible.



(c) The arcs in E_3 form a (2-inducible) forest of unidirected stars, centered at x^2, y^3, and c^3.



(d) The arcs in E_4 form a (2-inducible) forest of unidirected stars, centered at x^2, y^3, and c^2.

Fig. 12. The decomposition of the arcs of $\mathcal{G}^{SL}$ used in the proof of Theorem 5.

where, for $1 \le i \le m$ and $1 \le j \le l$,

$$D = \{d\},$$
$$U_i = \{u_i^1, u_i^2, u_i^3, u_i^4\}, \text{ and}$$
$$X_j = \{x_j\}.$$

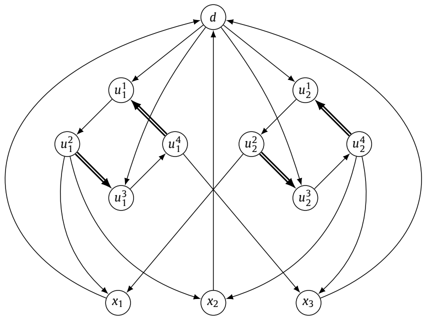


Fig. 13. A digraph (V, E) in the class $\mathcal{G}^{RP}$. Double arcs have weight 4 whereas normal arcs have weight 2.

If (V, E) is obtained as the digraph G_φ^{RP} for some φ in 3-CNF, l is the number of clauses, m the number of variables occurring in φ, the U_is are the variable gadgets, the X_js the clause gadgets, and, finally, D the decision vertex. Let $U_i^j = \{u_i^j\}$, $U^j = \bigcup_{i=1}^m \{u_i^j\}$, $U = \bigcup_{i=1}^m U_i$ and $X = \bigcup_{i=1}^l X_i$. Moreover, $E = E^\sigma \cup E^\varphi$, where E^σ (the *skeleton*) and E^φ (the *formula dependent part*) are disjoint such that

$$E^\sigma = (D \times (U^1 \cup U^3)) \cup (X \times D) \cup \bigcup_{i=1}^{m} \left\{(u_i^1, u_i^2), (u_i^2, u_i^3), (u_i^3, u_i^4), (u_i^4, u_i^1)\right\}$$

and E^φ is such that for all $1 \le i \le m$ and all $1 \le j \le l$:

$$E^\varphi \subset (U^2 \cup U^4) \times X,$$
$$|E^\varphi \cap (U^2 \cup U^4) \times X_j)| \le 3, \text{ and}$$
$$|E^\varphi \cap (U_i^2 \cup U_i^4) \times X_j)| \le 1,$$

i.e., every vertex in X has at most three incoming arcs (intuitively corresponding to the literals x contains) and at most one from every U_i (intuitively corresponding to the fact that no propositional variable occurs more than once in each clause). Finally, we check that the weight function w is defined such that all arcs in $E \cap ((U^2 \times U^3) \cup (U^4 \times U^1))$ have weight 4 and all arcs in $E \setminus ((U^2 \times U^3) \cup (U^4 \times U^1))$ have weight 2. An example illustrating this definition of the class $\mathcal{G}^{RP}$ is shown in Fig. 13.

Since G_φ^{RP} is incomplete, it can only be induced by a profile involving an even number of voters. In fact, we will prove that only eight voters suffice to induce any digraph in $\mathcal{G}^{RP}$.

Theorem 6. *Deciding whether a given alternative is a ranked pairs winner is NP-complete if the number of voters is even and at least* 8.

Proof. Membership in NP follows from the fact that it is easy to verify whether a given ranking can be the outcome of the *RP* procedure, independently of the number of voters.

For hardness, let (V, E) be a digraph (with weight function w) in $\mathcal{G}^{RP}$. Intuitively, $(V, E) = G_\varphi^{RP}$ for some formula φ in 3-CNF. It suffices to show that (V, E) is induced by an 8-voter profile. As an auxiliary notion, let for each $1 \le j \le l$,

$$E^\varphi \cap ((U^2 \cup U^4) \times X_j) = E_{j,1}^\varphi \cup E_{j,2}^\varphi \cup E_{j,3}^\varphi,$$

where $|E_{j,i}^\varphi| \le 1$ for all $1 \le i \le 3$. Intuitively, $E_{j,1}^\varphi$, $E_{j,2}^\varphi$, and $E_{j,3}^\varphi$ impose an ordering on the incoming arcs of vertex x_j. Also set

$$E_i^\varphi = \bigcup_{j=1}^{l} E_{j,i}^\varphi$$

for each $1 \le i \le 3$, i.e., E_i^φ collects the i-th incoming arcs of the vertices in X. Now define the following arc sets.

$$E_1 = E_1^\varphi \cup \bigcup_{i=1}^{m} \left((U_i^2 \times U_i^3) \cup (U_i^4 \times U_i^1)\right),$$

$$E_2 = E_2^\varphi \cup \bigcup_{i=1}^{m} \left((U_i^2 \times U_i^3) \cup (U_i^4 \times U_i^1)\right),$$

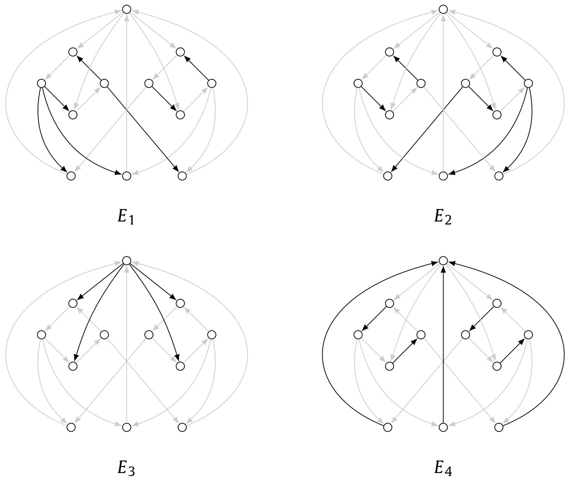


Fig. 14. The sets E_1, E_2, E_3, and E_4 for the digraph of Fig. 13 as defined in the proof of Theorem 6.

$$E_3 = E_3^{\varphi} \cup (D \times (U^1 \cup U^3)), \text{ and}$$

$$E_4 = (X \times D) \cup \bigcup_{i=1}^{m} \big((U_i^1 \times U_i^2) \cup (U_i^3 \times U_i^4)\big).$$

Observe that $E = E_1 \cup E_2 \cup E_3 \cup E_4$ (see Fig. 14). Moreover, each of (V, E_1), (V, E_2), (V, E_3), and (V, E_4) is a vertex-disjoint union of unidirected stars. Hence, by Lemma 6 we may assume they are induced by the 2-voter profiles (R_1^1, R_2^1), (R_1^2, R_2^2), (R_1^3, R_2^3), and (R_1^4, R_2^4), respectively. Moreover, E_1, E_2, E_3, and E_4 all contained in E and therefore also pairwise orientation compatible. By Lemma 8 it thus follows that (V, E) is induced by the 8-voter profile

$$R = (R_1^1, R_2^1, R_1^2, R_2^2, R_1^3, R_2^3, R_1^4, R_2^4).$$

Moreover, E_1, E_3, and E_4 as well as E_2, E_3, and E_4 are pairwise disjoint whereas $E_1 \cap E_2 = \bigcup_{i=1}^{m} \big((U_i^2 \times U_i^3) \cup (U_i^4 \times U_i^1)\big)$. Thus, all arcs in $E \setminus \bigcup_{i=1}^{m} \big((U_i^2 \times U_i^3) \cup (U_i^4 \times U_i^1)\big)$ have weight 2, whereas those in $\bigcup_{i=1}^{m} \big((U_i^2 \times U_i^3) \cup (U_i^4 \times U_i^1)\big)$ have weight 4. We may conclude that also the digraph (V, E) with its weights is induced by the 8-voter profile R. □

The original hardness construction contained arcs with weights 2 or 4 and unspecified arcs, defining a priority over the arcs. It is easy to see that increasing all weights in such a digraph by 1 to 3 and 5 does not change this priority. Similarly, adding arcs with weight 1 is not harmful as the corresponding pairs are added to the bottom of the priority, making them irrelevant to determining whether d is an *RP* winner or not. Therefore, by incorporating these observations into $\mathcal{G}_{\varphi}^{RP}$, for each Boolean formula φ in 3-CNF, we can create a weighted tournament (call it T_{φ}^{RP}) from which d is selected by *RP* if and only if φ is satisfiable. We denote the class of weighted tournaments that consist of these T_{φ}^{RP} by $\mathcal{T}^{RP}$.

We adopt the same notation as for $\mathcal{G}^{RP}$. A weighted tournament (V, E') (with weight function w$'$) belongs to $\mathcal{T}^{RP}$ if and only if it satisfies the following conditions. The set of alternatives can be written as

$$V = D \cup U_1 \cup \cdots \cup U_m \cup X_1 \cup \cdots \cup X_l$$

whereas the arc set E' is the union of two disjoint sets E'^{σ} (the *skeleton*) and E'^{φ} (the *formula dependent part*). Assuming that E is the arc set of $\mathcal{G}_{\varphi}^{RP}$, then $E'^{\varphi} = E^{\varphi}$ and $E'^{\sigma} = E^{\sigma} \cup E_c'^{\sigma}$ where

$$E_c'^{\sigma} = \big(((D \times U) \cup (U \times X)) \setminus E\big) \cup \bigcup_{i=1}^{m} \big((U_i^1 \times U_i^3) \cup (U_i^2 \times U_i^4)\big) \cup \bigcup_{i<j} (U_i \times U_j) \cup \bigcup_{i<j} (X_i \times X_j).$$

$E_c'^{\sigma}$ can be equivalently described as a reorientation of $\tilde{E}$. Moreover, we check that w$'$ is defined such that all arcs in $E'^{\sigma} \cap ((U^2 \times U^3) \cup (U^4 \times U^1))$ have weight 5, all arcs in $(E'^{\varphi} \cup E'^{\sigma}) \setminus ((U^2 \times U^3) \cup (U^4 \times U^1))$ have weight 3, and all arcs in $E_c'^{\sigma}$ have weight 1.

Now we can give the second result of this section.

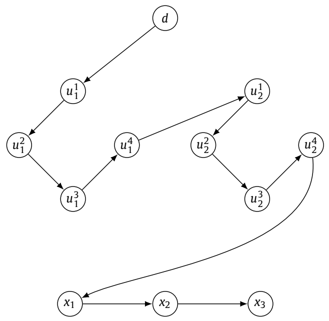


Fig. 15. The order implied by the arc set E'_6 over the alternatives of a tournament (V, E') in the class $\mathcal{T}^{RP}$.

Theorem 7. *Deciding whether a given alternative is a ranked pairs winner is NP-complete if the number of voters is odd and at least* 11.

Proof. The proof here is similar to that of the previous theorem. Let (V, E') be a tournament with weight function w' in $\mathcal{T}^{RP}$. Intuitively, $(V, E') = T^{RP}_{\varphi}$ for some formula φ in 3-CNF. It suffices to show that (V, E') is induced by an 11-voter profile. Using the notation provided in the proof of Theorem 6, we define the following arc sets.

$$\begin{aligned}
E'_1 &= E'^{\varphi}_1 \cup \bigcup_{i=1}^{m} \big((U_i^2 \times U_i^3) \cup (U_i^4 \times U_i^1)\big),\\
E'_2 &= E'^{\varphi}_2 \cup \bigcup_{i=1}^{m} \big((U_i^2 \times U_i^3) \cup (U_i^4 \times U_i^1)\big),\\
E'_3 &= E'^{\varphi}_3 \cup (D \times (U^1 \cup U^3)),\\
E'_4 &= (X \times D) \cup \bigcup_{i=1}^{m} \big((U_i^1 \times U_i^2) \cup (U_i^3 \times U_i^4)\big),\\
E'_5 &= (X \times D) \cup \bigcup_{i=1}^{m} \big\{(u_i^4, u_i^1)\big\}, \text{ and}\\
E'_6 &= (D \times U) \cup (D \times X) \cup (U \times X) \cup \bigcup_{\substack{1 \le i \le m\\ j<l}} (U_i^j \times U_i^l) \cup\\
&\quad \bigcup_{i<j} (U_i \times U_j) \cup \bigcup_{i<j} (X_i \times X_j).
\end{aligned}$$

Observe that E'_1, E'_2, E'_3, E'_4, and E'_5 are contained in E', making them pairwise orientation compatible, and that each of (V, E'_1), (V, E'_2), (V, E'_3), (V, E'_4), and (V, E'_5) is a forest of stars. Therefore, in virtue of Lemma 6 we may assume that they are induced by the 2-voter profiles (R^1_1, R^1_2), (R^2_1, R^2_2), (R^3_1, R^3_2), (R^4_1, R^4_2), and (R^5_1, R^5_2). Moreover, it can readily be appreciated that $E'_6 \supseteq E' \setminus (E'_1 \cup \ldots \cup E'_5)$. As E'_6 defines a transitive closure for an order over all of the alternatives in V (see Fig. 15), (V, E'_6) is acyclic, and we may assume that it is induced by a voter with the preference relation $R^6 = E'_6$. Thus by Lemma 9, (V, E') is induced by the 11-voter profile

$$R = (R^1_1, R^1_2, R^2_1, R^2_2, R^3_1, R^3_2, R^4_1, R^4_2, R^5_1, R^5_2, R^6).$$

Furthermore, note that there are some arcs in common among the arc sets and that E'_6 is not orientation compatible with E'. Arcs in $E'^{\sigma} \cap (U^2 \times U^3)$ occur in E'_1, E'_2, and E'_6; arcs in $E'^{\sigma} \cap (U^4 \times U^1)$ occur in E'_1, E'_2, and E'_5 while E'_6 includes arcs in the opposing direction or, equivalently, includes $\bigcup_{i=1}^{m} (U_i^1 \times U_i^4)$; each arc in $(E'^{\varphi} \cup E'^{\sigma}) \setminus ((U^2 \times U^3) \cup (U^4 \times U^1))$ occurs in E'_6 and exactly one of the other arc sets; and, finally, arcs in E'^{σ}_c occur only in E'_6. Thus, arcs in $E'^{\sigma} \cap ((U^2 \times U^3) \cup (U^4 \times U^1))$ have weight 5, arcs in $(E'^{\varphi} \cup E'^{\sigma}) \setminus ((U^2 \times U^3) \cup (U^4 \times U^1))$ have weight 3, and arcs in E'^{σ}_c have weight 1. Therefore, we may conclude that (V, E') together with its weights is induced by the 11-voter profile R. □

Corollary 1. *Deciding whether a given alternative is a ranked pairs winner is NP-complete if the number of voters is either* 8 *or at least* 10.

Proof. The statement follows from Theorems 6 and 7. □

Table 6
Numbers of voters for which winner determination is NP-hard. The Banks set and the tournament equilibrium set are defined for an odd number of voters only.

Voting rule	NP-hard for $n \geq$
Banks set	5 voters
Tournament equilibrium set	7 voters
Minimal extending set	7 voters
Slater's rule	7 voters
Kemeny's rule	7 voters
Ranked pairs	8 voters ($n \neq 9$)

8. Conclusion and future work

Many hardness results in computational social choice only hold if the number of voters is roughly of the same order as the number of alternatives. In some applications of voting, however, the number of voters can be much smaller than the number of alternatives and it is unclear whether hardness still holds.

We gave complete characterizations of 2-inducible and 3-inducible majority digraphs, respectively, and provided sufficient conditions for k-inducible majority digraphs. We then considered majority digraphs of real-world and of generated preference profiles. Using an implementation based on SAT solving, we computed their majority dimension and found that, in each case, they are inducible by at most eight voters. We did not encounter a single tournament that is not 5-inducible.[13]

We then leveraged the sufficient conditions we obtained earlier to show that winner determination for the Banks set, the tournament equilibrium set, the minimal extending set, Slater's rule, Kemeny's rule, and ranked pairs remains hard even when there is only a small constant number of voters. This was achieved by analyzing existing hardness proofs and checking whether the class of majority digraphs used in these constructions can be induced by small constant numbers of voters. Our hardness results are summarized in Table 6.

We believe there is interesting potential for future work. It would be desirable to characterize the sets of digraphs inducible by four, five, or more voters in graph-theoretic terms. The computational complexity of checking whether a given majority digraph is k-inducible is wide open for any fixed $k \geq 3$, or even without fixing k. We have shown that evaluating several common voting rules remains hard for a constant number of voters, but we do not know whether our bounds are tight. It would be interesting to obtain any lower bounds, and in particular to study the complexity of the Banks set for 3 voters and that of Kemeny's rule for 5 voters. Finally, our techniques might allow checking whether other hardness proofs in computational social choice remain intact for a bounded number of voters, most notably hardness shields against manipulation, bribery, and control. Some advances in this direction have recently been made by Chen et al. [17].

Acknowledgments

This material is based on work supported by the Deutsche Forschungsgemeinschaft under grants BR 2312/7-2 and BR 2312/9-1. Additionally, the work by Paul Harrenstein was supported by the ERC under Advanced Grant 291528 ("RACE"). Dominik Peters has been supported by EPSRC and by COST Action IC1205 on Computational Social Choice. Preliminary results of this paper were presented at the 12th International Conference on Autonomous Agents and Multi-Agent Systems (Saint Paul, USA, May 2013), the 1st Workshop on Exploring Beyond the Worst Case in Computational Social Choice (Paris, France, May 2014), and at an open poster session at the 6th International Workshop on Computational Social Choice (Toulouse, France, June 2016). The authors thank Martin Bullinger and Christian Stricker for many helpful comments, and Olivier Hudry and Rolf Niedermeier for pointing us to useful references.

References

[1] N. Alon, Ranking tournaments, SIAM J. Discrete Math. 20 (1) (2006) 137–142.
[2] N. Alon, G. Brightwell, H.A. Kierstead, A.V. Kostochka, P. Winkler, Dominating sets in k-majority tournaments, J. Comb. Theory, Ser. B 96 (2006) 374–387.
[3] D. Austen-Smith, J.S. Banks, Positive Political Theory I: Collective Preference, University of Michigan Press, 1999.
[4] T. Biedl, F.J. Brandenburg, X. Deng, On the complexity of crossings in permutations, Discrete Math. 309 (7) (2009) 1813–1823.
[5] A. Biere, Lingeling, plingeling and treengeling entering the SAT competition 2013, in: Proceedings of the SAT Competition 2013, 2013, pp. 51–52.
[6] F. Brandt, Minimal stable sets in tournaments, J. Econ. Theory 146 (4) (2011) 1481–1499.
[7] F. Brandt, H.G. Seedig, On the discriminative power of tournament solutions, in: Selected Papers of the International Conference on Operations Research, OR2014, Operations Research Proceedings, Springer-Verlag, 2016, pp. 53–58.
[8] F. Brandt, F. Fischer, P. Harrenstein, M. Mair, A computational analysis of the tournament equilibrium set, Soc. Choice Welf. 34 (4) (2010) 597–609.
[9] F. Brandt, M. Brill, H.G. Seedig, On the fixed-parameter tractability of composition-consistent tournament solutions, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI, AAAI Press, 2011, pp. 85–90.

[13] In Section 3, we saw that the smallest concrete tournament known not to be 5-inducible consists of more than 600 million vertices. Non-constructive arguments entail the existence of such a tournament with at most 41 vertices.

[10] F. Brandt, M. Chudnovsky, I. Kim, G. Liu, S. Norin, A. Scott, P. Seymour, S. Thomassé, A counterexample to a conjecture of Schwartz, Soc. Choice Welf. 40 (3) (2013) 739–743.
[11] F. Brandt, M. Brill, P. Harrenstein, Tournament solutions, in: F. Brandt, V. Conitzer, U. Endriss, J. Lang, A.D. Procaccia (Eds.), Handbook of Computational Social Choice, Cambridge University Press, 2016, Chapter 3.
[12] F. Brandt, V. Conitzer, U. Endriss, J. Lang, A. Procaccia (Eds.), Handbook of Computational Social Choice, Cambridge University Press, 2016.
[13] F. Brandt, P. Harrenstein, H.G. Seedig, Minimal extending sets in tournaments, Math. Soc. Sci. 87 (2017) 55–63.
[14] F. Brandt, M. Brill, H.G. Seedig, W. Suksompong, On the structure of stable tournament solutions, Econ. Theory 65 (2) (2018) 483–507.
[15] M. Brill, F. Fischer, The price of neutrality for the ranked pairs method, in: Proceedings of the 26th AAAI Conference on Artificial Intelligence, AAAI, AAAI Press, 2012, pp. 1299–1305.
[16] P. Charbit, S. Thomassé, A. Yeo, The minimum feedback arc set problem is NP-hard for tournaments, Comb. Probab. Comput. 16 (1) (2007) 1–4.
[17] J. Chen, P. Faliszewski, R. Niedermeier, N. Talmon, Elections with few voters: candidate control can be easy, J. Artif. Intell. Res. 60 (2017) 937–1002.
[18] V. Conitzer, Computing Slater rankings using similarities among candidates, in: Proceedings of the 21st National Conference on Artificial Intelligence, AAAI, AAAI Press, 2006, pp. 613–619.
[19] V. Conitzer, T. Sandholm, J. Lang, When are elections with few candidates hard to manipulate?, J. ACM 54 (3) (2007).
[20] D.E. Critchlow, M.A. Fligner, J.S. Verducci, Probability models on rankings, J. Math. Psychol. 35 (1991) 294–318.
[21] B. Dushnik, E.W. Miller, Partially ordered sets, Am. J. Math. 63 (3) (1941) 600–610.
[22] C. Dwork, R. Kumar, M. Naor, D. Sivakumar, Rank aggregation methods for the web, in: Proceedings of the 10th International Conference on the World Wide Web, WWW, ACM Press, 2001, pp. 613–622.
[23] C. Eggermont, C. Hurkens, G.J. Woeginger, Realizing small tournaments through few permutations, Acta Cybern. 21 (2) (2013) 267–271.
[24] P. Erdős, L. Moser, On the representation of directed graphs as unions of orderings, Publ. Math. Inst. Hung. Acad. Sci. 9 (1964) 125–132.
[25] P. Faliszewski, E. Hemaspaandra, L. Hemaspaandra, J. Rothe, A richer understanding of the complexity of election systems, in: S. Ravi, S. Shukla (Eds.), Fundamental Problems in Computing: Essays in Honor of Professor Daniel J. Rosenkrantz, Springer-Verlag, 2009.
[26] D. Fidler, A recurrence for bounds on dominating sets in k-majority tournaments, Electron. J. Comb. 18 (1) (2011).
[27] M.A. Fiol, A note on the voting problem, Stochastica XIII (1) (1992) 155–158.
[28] F. Fischer, O. Hudry, R. Niedermeier, Weighted tournament solutions, in: F. Brandt, V. Conitzer, U. Endriss, J. Lang, A.D. Procaccia (Eds.), Handbook of Computational Social Choice, Cambridge University Press, 2016, Chapter 4.
[29] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman, 1979.
[30] C. Geist, U. Endriss, Automated search for impossibility theorems in social choice theory: ranking sets of objects, J. Artif. Intell. Res. 40 (2011) 143–174.
[31] R.L. Graham, J.H. Spencer, A constructive solution to a tournament problem, Can. Math. Bull. 14 (1) (1971) 45–48.
[32] O. Hudry, A note on "Banks winners in tournaments are difficult to recognize" by G. J. Woeginger, Soc. Choice Welf. 23 (1) (2004) 113–114.
[33] O. Hudry, NP-hardness results for the aggregation of linear orders into median orders, Ann. Oper. Res. 163 (1) (2008) 63–88.
[34] O. Hudry, On the complexity of Slater's problems, Eur. J. Oper. Res. 203 (1) (2010) 216–221.
[35] R.M. Karp, Reducibility among combinatorial problems, in: R.E. Miller, J.W. Thatcher (Eds.), Complexity of Computer Computations, Plenum Press, 1972, pp. 85–103.
[36] J.G. Kemeny, Mathematics without numbers, Daedalus 88 (1959) 577–591.
[37] H.A. Kierstead, W.T. Trotter Jr., P. Charbit, K. Milans, P. Wenger, k-majority digraphs, http://www.math.uiuc.edu/~west/regs/k-majority.html, 2009.
[38] J.-F. Laslier, Tournament Solutions and Majority Voting, Springer-Verlag, 1997.
[39] J.-F. Laslier, In silico voting experiments, in: J.-F. Laslier, M.R. Sanver (Eds.), Handbook on Approval Voting, Springer-Verlag, 2010, pp. 311–335, Chapter 13.
[40] C.L. Mallows, Non-null ranking models, Biometrika 44 (1/2) (1957) 114–130.
[41] J.I. Marden, Analyzing and Modeling Rank Data, Monographs on Statistics and Applied Probability, vol. 64, Chapman & Hall, 1995.
[42] N. Mattei, T. Walsh, PrefLib: a library for preference data, in: Proceedings of the 3rd International Conference on Algorithmic Decision Theory, ADT, in: Lecture Notes in Computer Science (LNCS), vol. 8176, Springer-Verlag, 2013, pp. 259–270.
[43] N. Mattei, J. Forshee, J. Goldsmith, An empirical study of voting rules and manipulation with large datasets, in: Proceedings of the 4th International Workshop on Computational Social Choice, COMSOC, 2012.
[44] J.C. McCabe-Dansted, A. Slinko, Exploratory analysis of similarities between social choice rules, Group Decis. Negot. 15 (1) (2006) 77–107.
[45] R.M. McConnell, F. de Montgolfier, Linear-time modular decomposition of directed graphs, Discrete Appl. Math. 145 (2) (2005) 198–209.
[46] D.C. McGarvey, A theorem on the construction of voting paradoxes, Econometrica 21 (4) (1953) 608–610.
[47] B.D. McKay, A. Piperno, Practical graph isomorphism, II, J. Symb. Comput. 60 (2014) 94–112.
[48] P.C. Ordeshook, The Spatial Analysis of Elections and Committees: Four Decades of Research, Technical report, California Institute of Technology, 1993.
[49] A. Pnueli, A. Lempel, S. Even, Transitive orientation of graphs and identification of permutation graphs, Can. J. Math. 23 (1971) 160–175.
[50] J. Rothe (Ed.), Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division, Springer, 2015.
[51] D. Shepardson, C.A. Tovey, Smallest tournament not realizable by $\frac{2}{3}$-majority voting, Soc. Choice Welf. 33 (3) (2009) 495–503.
[52] R. Stearns, The voting problem, Am. Math. Mon. 66 (9) (1959) 761–763.
[53] T.N. Tideman, Independence of clones as a criterion for voting rules, Soc. Choice Welf. 4 (3) (1987) 185–206.
[54] C.A. Tovey, A simplified NP-complete satisfiability problem, Discrete Appl. Math. 8 (1) (1984) 85–89.
[55] G.S. Tseitin, On the complexity of derivation in propositional calculus, in: Automation of Reasoning, Springer, 1983, pp. 466–483.
[56] G.J. Woeginger, Banks winners in tournaments are difficult to recognize, Soc. Choice Welf. 20 (3) (2003) 523–528.
[57] M. Yannakakis, The complexity of the partial order dimension problem, SIAM J. Algebraic Discrete Methods 3 (3) (1982) 351–358.
[58] H.P. Young, A. Levenglick, A consistent extension of Condorcet's election principle, SIAM J. Appl. Math. 35 (2) (1978) 285–300.