On the Exact Amount of Missing Information That Makes Finding Possible Winners Hard

Palash Dey[1] and Neeldhara Misra*[2]

1 **Tata Institute of Fundamental Research, Mumbai, India**
`palash.dey@tifr.res.in`

2 **Indian Institute of Technology, Gandhinagar, India**
`neeldhara.m@iitgn.ac.in`

Abstract

We consider election scenarios with incomplete information, a situation that arises often in practice. There are several models of incomplete information and accordingly, different notions of outcomes of such elections. In one well-studied model of incompleteness, the votes are given by partial orders over the candidates. In this context we can frame the problem of finding a *possible winner*, which involves determining whether a given candidate wins in at least one completion of a given set of partial votes for a specific voting rule.

The Possible Winner problem is well-known to be NP-complete in general, and it is in fact known to be NP-complete for several voting rules where the number of undetermined pairs in every vote is bounded only by some constant. In this paper, we address the question of determining precisely the smallest number of undetermined pairs for which the Possible Winner problem remains NP-complete. In particular, we find the exact values of t for which the Possible Winner problem transitions to being NP-complete from being in P, where t is the maximum number of undetermined pairs in every vote. We demonstrate tight results for a broad subclass of scoring rules which includes all the commonly used scoring rules (such as plurality, veto, Borda, and k-approval), Copeland$^\alpha$ for every $\alpha \in [0, 1]$, maximin, and Bucklin voting rules. A somewhat surprising aspect of our results is that for many of these rules, the Possible Winner problem turns out to be hard even if every vote has at most one undetermined pair of candidates.

1998 ACM Subject Classification F.2.0 Analysis Of Algorithms And Problem Complexity - General

Keywords and phrases Computational Social Choice, Dichotomy, NP-completeness, Maxflow, Voting, Possible winner

Digital Object Identifier 10.4230/LIPIcs.MFCS.2017.57

1 Introduction

In many real life situations including multiagent systems, agents often need to aggregate their preferences and agree upon a common decision (candidate). Voting is an immediate natural tool in these situations. Common and classic applications of voting in multiagent systems include collaborative filtering and recommender systems [26], spam detection [9], computational biology [20], winner determination in sports competition [5] etc. We refer the readers to [25] for an elaborate treatment of computational voting theory.

Usually, in a voting setting, it is assumed that the votes are complete orders over the candidates. However, due to many reasons, for example, lack of knowledge of voters about

* The second author acknowledges support from the DST-INSPIRE faculty grant (IFA12-ENG-31).

42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017).
Editors: Kim G. Larsen, Hans L. Bodlaender, and Jean-Francois Raskin; Article No. 57; pp. 57:1–57:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

some candidates, a voter may be indifferent between some pairs of candidates. Hence, it is both natural and important to consider scenarios where votes are partial orders over the candidates. When votes are only partial orders over the candidates, the winner cannot be determined with certainty since it depends on how these partial orders are extended to linear orders. This leads to a natural computational problem called the Possible Winner problem [21]: given a set of partial votes $\mathcal{P}$ and a distinguished candidate c, is there a way to extend the partial votes to complete votes where c wins? The Possible Winner problem has been studied extensively in the literature [23, 27, 28, 29, 7, 8, 4, 1, 22, 18] following its definition in [21]. Betzler et al. [6] and Baumeister et al. [2] show that the Possible Winner winner problem is NP-complete for all scoring rules except for the plurality and veto voting rules; the Possible Winner winner problem is in P for the plurality and veto voting rules. The Possible Winner problem is known to be NP-complete for many common voting rules, for example, a class of scoring rules, maximin, Copeland, Bucklin etc. even when the maximum number of undetermined pairs of candidates in every vote is bounded above by small constants [29]. Walsh showed that the Possible Winner problem can be solved in polynomial time for all the voting rules mentioned above when we have a constant number of candidates [28].

1.1 Our Contribution

Our main contribution lies in pinning down exactly the minimum number of undetermined pairs allowed per vote so that the Possible Winner winner problem continues to be NP-complete for a large class of scoring rules, Copeland$^\alpha$, maximin, and Bucklin voting rules. To begin with, we describe our results for scoring rules. We work with a class of scoring rules that we call smooth, which are essentially scoring rules where the score vector for $(m+1)$ candidates can be obtained by either duplicating an already duplicated score in the score vector for m candidates, or by extending the score vector for m candidates at one of the endpoints with an arbitrary new value. The smooth rules account for all commonly used scoring rules (such as Borda, plurality, veto, k-approval). Using t to denote the maximum number of undetermined pairs of candidates in every vote, we show the following (note that the Possible Winner problem is in P for all scoring rules when $t = 0$):

- The Possible Winner problem is NP-complete even when $t = 1$ for scoring rules which have two distinct nonzero differences between consecutive coordinates in the score vector (we call them differentiating) and in P when $t = 1$ for other scoring rules [Theorem 7].
- Else the Possible Winner problem is NP-complete when $t \geqslant 2$ and in P when $t \leqslant 1$ for scoring rules that contain $(\alpha+1, \alpha+1, \alpha)$ for any $\alpha \in \mathbb{N}$ [Theorem 8].
- Else the Possible Winner problem is NP-complete when $t \geqslant 3$ and in P when $t \leqslant 2$ for scoring rules which contain $(\alpha+2, \alpha+1, \alpha+1, \alpha)$ for any $\alpha \in \mathbb{N}$ [Theorem 9].
- The Possible Winner problem is NP-complete when $t \geqslant 4$ and in P when $t \leqslant 3$ for k-approval and k-veto voting rules for any $k > 1$ [Theorem 10].
- The Possible Winner problem is NP-complete when $t \geqslant m-1$ and in P when $t \leqslant m-2$ for the scoring rule $(2, 1, 1, \dots, 1, 0)$ [Theorem 10].

We summarize our results for the Copeland$^\alpha$, maximin, and Bucklin voting rules in Table 1. We observe that the Possible Winner problem for the Copeland$^\alpha$ voting rule is NP-complete even when every vote has at most 2 undetermined pairs of candidates for $\alpha \in \{0, 1\}$. However, for $\alpha \in (0, 1)$, the Possible Winner problem for the Copeland$^\alpha$ voting rule is NP-complete even when every vote has at most 1 undetermined pairs of candidates. Our results show that the Possible Winner winner problem continues to be NP-complete

Table 1 Summary and comparison of results from the literature for Copeland$^\alpha$, maximin, and Bucklin voting rules. *The result was proved for the simplified Bucklin voting rule but the proof can be modified easily for the Bucklin voting rule.

Voting rules	NP-complete	Poly time	Known from literature [29]
Copeland^{0,1}	$t \geqslant 2$ [Theorem 11]	$t \leqslant 1$ [Theorem 12]	NP-complete for $t \geqslant 8$
Copeland$^\alpha$ $\alpha \in (0,1)$	$t \geqslant 1$ [Theorem 15]	–	
Maximin	$t \geqslant 2$ [Theorem 17]	$t \leqslant 1$ [Theorem 18]	NP-complete for $t \geqslant 4$
Bucklin	$t \geqslant 2$ [Theorem 19]	$t \leqslant 1$ [Theorem 19]	NP-complete for $t \geqslant 16^\star$

for all the common voting rules studied here (except k-approval) even when the number of undetermined pairs of candidates per vote is at most 2. Other than finding the exact number of undetermined pairs needed per vote to make the Possible Winner problem NP-complete for common voting rules, we also note that all our proofs are much simpler and shorter than most of the corresponding proofs from the literature subsuming the work in [29, 6, 2].

2 Preliminaries

Let us denote the set $\{1, 2, \dots, n\}$ by $[n]$ for any positive integer n. Let $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ be a set of candidates or alternatives and $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ a set of voters. If not mentioned otherwise, we denote the set of candidates by $\mathcal{C}$, the set of voters by $\mathcal{V}$, the number of candidates by m, and the number of voters by n. Every voter v_i has a preference or vote $\succ_i$ which is a complete order over $\mathcal{C}$. We denote the set of complete orders over $\mathcal{C}$ by $\mathcal{L}(\mathcal{C})$. We call a tuple of n preferences $(\succ_1, \succ_2, \cdots, \succ_n) \in \mathcal{L}(\mathcal{C})^n$ an n-voter preference profile. An election is defined as a set of candidates together with a voting profile. It is often convenient to view a preference as a subset of $\mathcal{C} \times \mathcal{C}$ — a preference $\succ$ corresponds to the subset $\mathcal{A} = \{(x, y) \in \mathcal{C} \times \mathcal{C} : x \succ y\}$. For a preference $\succ$ and a subset $\mathcal{A} \subseteq \mathcal{C}$ of candidates, we define $\succ(\mathcal{A})$ be the preference $\succ$ restricted to $\mathcal{A}$, that is $\succ(\mathcal{A}) = \succ \cap (\mathcal{A} \times \mathcal{A})$. Let $\uplus$ denote the disjoint union of sets. A map $r : \uplus_{n,|\mathcal{C}| \in \mathbb{N}^+} \mathcal{L}(\mathcal{C})^n \longrightarrow 2^{\mathcal{C}} \setminus \{\emptyset\}$ is called a *voting rule*. For a voting rule r and a preference profile $\succ = (\succ_1, \dots, \succ_n)$, we say a candidate x wins uniquely if $r(\succ) = \{x\}$ and x co-wins if $x \in r(\succ)$. For a vote $\succ \in \mathcal{L}(\mathcal{C})$ and two candidates $x, y \in \mathcal{C}$, we say x is placed before y in $\succ$ if $x \succ y$; otherwise we say x is placed after y in $\succ$. A candidate is said to be at the i^{th} position from the top (bottom) if there are $(i-1)$ candidates after (before) it. For any two candidates $x, y \in \mathcal{C}$ with $x \neq y$ in an election $\mathcal{E}$, let us define the margin $\mathcal{D}_{\mathcal{E}}(x, y)$ of x from y to be $|\{i : x \succ_i y\}| - |\{i : y \succ_i x\}|$. Examples of some common voting rules are as follows.

Positional scoring rules. A collection $(\overrightarrow{s_m})_{m \in \mathbb{N}^+}$ of m-dimensional vectors $\overrightarrow{s_m} = (\alpha_m, \alpha_{m-1}, \dots, \alpha_1) \in \mathbb{N}^m$ with $\alpha_m \geqslant \alpha_{m-1} \geqslant \dots \geqslant \alpha_1$ and $\alpha_m > \alpha_1$ for every $m \in \mathbb{N}^+$ naturally defines a voting rule — a candidate gets score α_i from a vote if it is placed at the i^{th} position from the bottom, and the score of a candidate is the sum of the scores it receives from all the votes. The winners are the candidates with maximum score. Scoring rules remain unchanged if we multiply every α_i by any constant $\lambda > 0$ and/or add any constant μ. Hence, we assume without loss of generality that for any score vector $\overrightarrow{s_m}$, there exists a j such that $\alpha_k = 0$ for all $k < j$ and the greatest common divisor of $\alpha_1, \dots, \alpha_m$ is one. Such a $\overrightarrow{s_m}$ is called a normalized score vector. Without loss of generality, we will work with normalized scoring rules only in this work. If α_i is 0 for $i \in [m-k]$ and 1 otherwise, then we get the k-approval voting rule. For the k-veto voting rule, α_i is 0 for $i \in [k]$ and 1 otherwise. 1-approval is called the plurality voting rule and 1-veto is called the veto voting rule. We note that our notation is slightly unconventional, this is in the interest of convenience in some of the computations that we will encounter with score vectors.

Copeland$^\alpha$. Given $\alpha \in [0,1]$, the Copeland$^\alpha$ score of a candidate x is $|\{y \neq x : \mathcal{D}_\mathcal{E}(x,y) > 0\}| + \alpha|\{y \neq x : \mathcal{D}_\mathcal{E}(x,y) = 0\}|$. The winners are the candidates with maximum Copeland$^\alpha$ score. If not mentioned otherwise, we will assume α to be zero.

Maximin. The maximin score of a candidate x in an election E is $\min_{y \neq x} \mathcal{D}_\mathcal{E}(x,y)$. The winners are the candidates with maximum maximin score.

Bucklin. Let ℓ be the minimum integer such that there exists at least one candidate $x \in \mathcal{C}$ whom more than half of the voters place in their top ℓ positions. Then the Bucklin winner is the candidate who is placed most number of times within top ℓ positions of the votes.

Elections with Incomplete Information A more general setting is an *election* where the votes are only *partial orders* over candidates. A *partial order* is a relation that is *reflexive, antisymmetric*, and *transitive*. A partial vote can be extended to possibly more than one linear vote depending on how we fix the order for the unspecified pairs of candidates. Given a partial vote $\succ$, we say that an extension $\succ'$ of $\succ$ *places the candidate c as high as possible* if $a \succ' c$ implies $a \succ'' c$ for every extension $\succ''$ of $\succ$.

▶ **Definition 1** (r–Possible Winner)**.** Given a set of partial votes $\mathcal{P}$ over a set of candidates $\mathcal{C}$ and a candidate $c \in \mathcal{C}$, does there exist an extension $\mathcal{P}'$ of $\mathcal{P}$ such that $c \in r(\mathcal{P}')$?

3 Results

For ease of exposition, we present all our results for the co-winner case. All our proofs extend easily to the unique winner case too. We begin with our results for the scoring rules.

3.1 Scoring Rules

In this section, we establish a dichotomous result describing the status of the Possible Winner problem for a large class of scoring rules when the number of undetermined pairs in every vote is at most one, two, three, or four. We begin by introducing some terminology. Instead of working directly with score vectors, it will sometimes be convenient for us to refer to the "vector of differences", which, for a score vector s with m coordinates, is a vector $d(s)$ with $m-1$ coordinates with each entry being the difference between adjacent scores corresponding to that location and the location left to it. This is formally stated below.

▶ **Definition 2.** Given a normalized score vector $\overrightarrow{s_m} = (\alpha_m, \alpha_{m-1}, \ldots, \alpha_1 = 0) \in \mathbb{N}^m$, the associated difference vector $d(\overrightarrow{s_m})$ is given by $(\alpha_m - \alpha_{m-1}, \alpha_{m-1} - \alpha_{m-2}, \ldots, \alpha_2 - \alpha_1) \in \mathbb{N}^{m-1}$. We also employ the following notation to refer to the smallest score difference among all non-zero differences, and the largest score difference, respectively:

- $\delta(\overrightarrow{s_m}) = \min(\{\alpha_i - \alpha_{i-1} \mid 2 \leqslant i \leqslant m \text{ and } \alpha_i - \alpha_{i-1} > 0\})$
- $\Delta(\overrightarrow{s_m}) = \max(\{\alpha_i - \alpha_{i-1} \mid 2 \leqslant i \leqslant m\})$

Note that for every non-trivial normalized score vector $\overrightarrow{s_m}$, $\Delta(\overrightarrow{s_m})$ is always non-zero. We now proceed to defining the notion of smooth scoring rules. Consider a score vector $\overrightarrow{s_m} = (\alpha_m, \alpha_{m-1}, \ldots, \alpha_1)$. For $0 \leqslant i \leqslant m$, we say that $\overrightarrow{s_{m+1}}$ is obtained from s_m by inserting α just before position i from the right if:

$$\overrightarrow{s_{m+1}} = (\alpha_m, \alpha_{m-1}, \ldots, \alpha_{i+1}, \alpha, \alpha_i, \ldots, \alpha_2, \alpha_1).$$

Note that if $i = 0$, we have $\overrightarrow{s_{m+1}} = (\alpha_m, \alpha_{m-1}, \ldots, \alpha_1, \alpha)$, and if $i = m$, then we have $\overrightarrow{s_{m+1}} = (\alpha, \alpha_m, \alpha_{m-1}, \ldots, \alpha_1)$. For $0 \leqslant i \leqslant m$, we say that the position i is *admissible* if $i = 0$, or $i = m$, or $\alpha_{i+1} = \alpha_i$.

▶ **Definition 3** (Smooth scoring rules)**.** We say that a scoring rule s is smooth if there exists some constant $m_0 \in \mathbb{N}^+$ such that for all $m \geqslant m_0$, the score vector $\overrightarrow{s_m}$ can be obtained from $\overrightarrow{s_{m-1}}$ by inserting an additional score value at any position i that is admissible.

Observe that the additional score value is forced to be equal to an existing score value unless it is inserted at one of the endpoints. Intuitively speaking, a smooth scoring rule is one where the score vector for m candidates can be obtained by either extending the one for $(m-1)$ candidates at one of the ends, or by inserting a score between an adjacent pair of ambivalent locations (i.e, consecutive scores in the score vector with the same value). Although at a first glance it may seem that the class of smooth scoring rules involves an evolution from a limited set of operations, we note that all of the common scoring rules, such as plurality, veto, k-approval, Borda, and scoring rules of the form $(2, 1, \ldots, 1, 0)$, are smooth. We now turn to some definitions that will help describe the cases that appear in our classification result.

▶ **Definition 4.** Let $s = (\overrightarrow{s_m})_{m \in \mathbb{N}^+}$ be a scoring rule.

- We say that s is a *Borda-like scoring rule* if there exists some $m_0 \in \mathbb{N}^+$ for which we have that $\Delta(\overrightarrow{s_m}) = \delta(\overrightarrow{s_m})$ for every $m > m_0$.
- Any rule that is not Borda-like is called a *differentiating scoring rule.*
- For any vector t with ℓ coordinates, we say that s is t-difference-free if there exists some $n_0 \in \mathbb{N}^+$ such that for every $m \geqslant n_0$, the vector t does not occur in $d(\overrightarrow{s_m})$. In other words, the vector $\langle d(\overrightarrow{s_m})[i], \ldots, d(\overrightarrow{s_m})[i + \ell - 1]\rangle \neq t$ for any $1 \leqslant i \leqslant m - \ell$.
- For any vector t, we say that s is t-contaminated if it is not t-difference-free. We also say that s is t-contaminated at m if the vector t occurs in $d(\overrightarrow{s_m})$.

We will frequently be dealing with Borda-like score vectors. To this end, the following easy observation will be useful.

▶ **Observation 5.** *If $s = (\overrightarrow{s_m})_{m \in \mathbb{N}^+}$ is a Borda-like scoring rule in its normalized form, then there exists $n_0 \in \mathbb{N}^+$ such that all the coordinates of $d(\overrightarrow{s_m})$ are either zero or one for all $m > n_0$.*

It turns out that if a scoring rule is smooth, then its behavior with respect to some of the properties above is fairly monotone. For instance, we have the following easy proposition. For the interest of space, we omit proofs of some of our results including all our polynomial time algorithms. For a few proofs, we only provide a sketch of the proof deferring the complete proof to the appendix. We mark these results with $\star$. All our polynomial time algorithms are based on reduction to the maximum flow problem in a graph. All the complete proofs can be found here [12].

[$\star$] Let $s = (\overrightarrow{s_m})_{m \in \mathbb{N}^+}$ be a smooth scoring rule that is not Borda-like. Then there exists some $n_0 \in \mathbb{N}^+$ such that $\Delta(\overrightarrow{s_m}) \neq \delta(\overrightarrow{s_m})$ for every $m > n_0$.

We are now ready to state the first classification result of this section, for the scenario where every vote has at most one missing pair. We use $(3, B2)$–SAT to prove some of our hardness results. The $(3, B2)$–SAT problem is the 3-SAT problem restricted to formulas in which each clause contains exactly three literals, and each variable occurs exactly twice positively and twice negatively. We know that $(3, B2)$–SAT is NP-complete [3]. Let us first present a structural result for scoring rules which we will use subsequently.

Suppose we have a set $\mathcal{C} = \{c_1, \ldots, c_{m-1}, g\}$ of m candidates including a "dummy" candidate g. Then it is well known [1, 16, 14, 13, 15, 17, 11], that for a score vector $(\alpha_m, \ldots, \alpha_1)$ and integers $\{k_i^j\}_{i\in[m-1],j\in[m-1]}$, we can add votes polynomially many in $\sum_{i\in[m-1],j\in[m-1]} k_i^j$ so that the score of the candidate c_i is $\lambda + \sum_{j\in[m-1]} k_i^j(\alpha_j - \alpha_{j+1})$ for some λ and the score of g is less than λ. Since the greatest common divisor of non-zero differences of the consecutive entries in a normalized score vector is one, we have the following.

▶ **Lemma 6.** *Let $\mathcal{C} = \{c_1, \ldots, c_m\} \cup D, (|D| > 0)$ be a set of candidates, and $\vec{\alpha}$ a normalized score vector of length $|\mathcal{C}|$. Then for every $\mathbf{X} = (X_1, \ldots, X_m) \in \mathbb{Z}^m$, there exists $\lambda \in \mathbb{N}$ and a voting profile $\mathcal{V}$ such that the $\vec{\alpha}$-score of c_i is $\lambda + X_i$ for all $1 \leqslant i \leqslant m$, and the score of candidates $d \in D$ is less than λ. Moreover, the number of votes in $\mathcal{V}$ is $O(poly(|\mathcal{C}|\cdot\sum_{i=1}^m |X_i|))$.*

▶ **Theorem 7.** *[★] Let s be a smooth scoring rule. If s is differentiating, then the* Possible Winner *problem is* NP*-complete, even if every vote has at most one undetermined pair of candidates. Otherwise, that is, when s is Borda-like, the* Possible Winner *problem for s is in* P *if every vote has at most one undetermined pair of candidates.*

Proof. *(Outline.)* For the hardness result, we reduce from an instance of $(3, B2)$-SAT. Let $\mathcal{I}$ be an instance of $(3, B2)$-SAT, over the variables $\mathcal{V} = \{x_1, \ldots, x_n\}$ and with clauses $\mathcal{T} = \{c_1, \ldots, c_t\}$. To construct the reduced instance $\mathcal{I}'$, we introduce two candidates for every variable, and one candidate for every clause, one special candidate w, and a dummy candidate g to achieve desirable score differences. Notationally, we will use b_i (corresponding to x_i) and b_i' (corresponding to $\bar{x}_i$) to refer to the candidates based on the variable x_i and e_j to refer to the candidate based on the clause c_j. To recap, the set of candidates are given by:

$$\mathcal{C} = \{b_i, b_i' \mid x_i \in \mathcal{V}\} \cup \{e_j \mid c_j \in \mathcal{T}\} \cup \{w, g\}.$$

Consider an arbitrary but fixed ordering over $\mathcal{C}$, such as the lexicographic order. In this proof, the notation $\overrightarrow{\mathcal{C}'}$ for any $\mathcal{C}' \subseteq \mathcal{C}$ will be used to denote the lexicographic ordering restricted to the subset $\mathcal{C}'$. Let m denote $|\mathcal{C}| = 2n+t+2$, and let $\overrightarrow{s_m} = (\alpha_m, \alpha_{m-1}, \ldots, \alpha_1) \in \mathbb{N}^m$. Since s is a smooth differentiating scoring rule, we have that there exist $1 \leqslant p, q \leqslant m$ such that $|p - q| > 1$ and $\alpha_p - \alpha_{p-1} > \alpha_q - \alpha_{q-1} \geqslant 1$.

We use D to refer to the larger of the two differences above, namely $\alpha_p - \alpha_{p-1}$ and d to refer to $\alpha_q - \alpha_{q-1}$. We now turn to a description of the votes. Fix an arbitrary subset $\mathcal{C}_1$ of $(m - p)$ candidates. For every variable $x_i \in \mathcal{V}$, we introduce the following complete and partial votes.

$$\mathfrak{p}_i := \overrightarrow{\mathcal{C}_1} \succ b_i \succ b_i' \succ \overrightarrow{\mathcal{C} \setminus \mathcal{C}_1} \text{ and } \mathfrak{p}_i' := \mathfrak{p}_i \setminus \{(b_i, b_i')\}$$

We next fix an arbitrary subset $\mathcal{C}_2 \subset \mathcal{C}$ of $(m - q)$ candidates. Consider a literal ℓ corresponding to the variable x_i. We use $\ell^\star$ to refer to the candidate b_j if the literal is positive and to refer to the candidate b_j' if the literal is negated. For every clause $c_j \in \mathcal{T}$ given by $c_j = \{\ell_1, \ell_2, \ell_3\}$, we introduce the following complete and partial votes.

$$\mathfrak{q}_{j,1} := \overrightarrow{\mathcal{C}_2} \succ e_j \succ \ell_1^\star \succ \overrightarrow{\mathcal{C} \setminus \mathcal{C}_2} \text{ and } \mathfrak{q}_{j,1}' := \mathfrak{q}_{j,1} \setminus \{(e_j, \ell_1^\star)\}$$

Table 2 Score of candidates from $\mathcal{P} \cup \mathcal{W}$.

$s^+(e_j) = s^+(w) + d \; \forall \; 1 \leqslant j \leqslant t$ $s^+(b_i') = s^+(w) + 1 - d - D \; \forall \; 1 \leqslant i \leqslant n$	$s^+(b_i) = s^+(w) + 1 - d \; \forall \; 1 \leqslant i \leqslant n$ $s^+(g) < s^+(w)$

$$\mathfrak{q}_{j,2} := \overrightarrow{\mathcal{C}_2} \succ e_j \succ \ell_2^\star \succ \overrightarrow{\mathcal{C} \setminus \mathcal{C}_2} \text{ and } \mathfrak{q}_{j,2}' := \mathfrak{q}_{j,2} \setminus \{(e_j, \ell_2^\star)\}$$

$$\mathfrak{q}_{j,3} := \overrightarrow{\mathcal{C}_2} \succ e_j \succ \ell_3^\star \succ \overrightarrow{\mathcal{C} \setminus \mathcal{C}_2} \text{ and } \mathfrak{q}_{j,3}' := \mathfrak{q}_{j,3} \setminus \{(e_j, \ell_3^\star)\}$$

Let us define the following sets of votes:

$$\mathcal{P} = \left(\bigcup_{i=1}^{n} \mathfrak{p}_i \right) \cup \left(\bigcup_{1 \leqslant j \leqslant t, 1 \leqslant b \leqslant 3} \mathfrak{q}_{j,b} \right) \text{ and } \mathcal{P}' = \left(\bigcup_{i=1}^{n} \mathfrak{p}_i' \right) \cup \left(\bigcup_{1 \leqslant j \leqslant t, 1 \leqslant b \leqslant 3} \mathfrak{q}_{j,b}' \right)$$

There exists a set of complete votes $\mathcal{W}$ of size polynomial in m with the following properties due to Lemma 6. Let $s^+ : \mathcal{C} \longrightarrow \mathbb{N}$ be a function mapping candidates to their scores from the set of votes $\mathcal{P} \cup \mathcal{W}$. Then $\mathcal{W}$ can be constructed to ensure the scores as in Table 2. We now define the instance $\mathcal{I}'$ of Possible Winner to be $(\mathcal{C}, \mathcal{P}' \cup \mathcal{W}, w)$. This completes the description of the reduction. We now turn to a proof of the equivalence. Before we begin making our arguments, observe that since w does not participate in any undetermined pairs of the votes in $\mathcal{P}'$, it follows that the score of w continues to be $s^+(w)$ in any completion of $\mathcal{P}'$. The intuition for the construction, described informally, is as follows. The score of every "clause candidate" needs to decrease by d, which can be achieved by pushing it down against its literal partner in the $\mathfrak{q}_j$-votes. However, this comes at the cost of increasing the score of the literals by $2d$ (since every literal appears in at most two clauses). It turns out that this can be compensated appropriately by ensuring that the candidate corresponding to the literal appears in the $(p-1)^{th}$ position among the $\mathfrak{p}$-votes, which will adjust for this increase. Therefore, the setting of the (b_i', b_i) pairs in a successful completion of $\mathfrak{p}_i$ can be read off as a signal for how the corresponding variable should be set by a satisfying assignment. We defer the formal proof of equivalence of the two instances and the polynomial time solvable case to the appendix. ◀

We make a couple of quick remarks before moving on to our next result. Observe that any hardness result that holds for instances where every vote has at most k undetermined pairs also holds for instances where every vote has at most k' undetermined pairs with $k' > k$, by a standard special case argument. Therefore, the next question for us to address is that of whether the Possible Winner problem is in P for all Borda-like scoring rules when the number of undetermined pairs in every vote is at most two. We show that the complexity of the Possible Winner problem for the Borda-like scoring rules crucially depends on the presence (or absence) some particular patterns in the score vector. We begin by stating a hardness result which uses a reduction from the well-known Three Dimensional Matching problem [19].

To help us deal with the nature of the score vectors considered, we will use the following proposition, which again reflects the monotonicity property alluded to earlier.

[$\star$] Let s be a normalized smooth scoring rule that is not $\langle 1, 1 \rangle$-difference-free. Then there exists some $n_0 \in \mathbb{N}^+$ such that for every $m \geqslant n_0$, s is $\langle 1, 1 \rangle$-contaminated at m.

We are now ready to state our next result, which shows that if there are at most 2 undetermined pairs of candidates in every vote, and we are dealing with a smooth Borda-like scoring rule s, then the Possible Winner problem is NP-complete if s is $\langle 1, 1 \rangle$-contaminated, and solvable in polynomial time otherwise.

▶ **Theorem 8.** *[⋆] Let s be a smooth, Borda-like scoring rule. If s is $\langle 1,1\rangle$-contaminated, the* Possible Winner *problem is* NP*-complete, even if every vote has at most* 2 *undetermined pairs of candidates. On the other hand, if s is $\langle 1,1\rangle$-difference-free, then the* Possible Winner *problem for s is in* P *if every vote has at most* 2 *undetermined pairs of candidates.*

We now address the case involving at most three undetermined pairs in every vote. The interesting scoring rules here are smooth Borda-like scoring rules that are $\langle 1,1\rangle$-difference-free. It turns out that here, if the scoring rule is further $\langle 1,0,1\rangle$-difference-free, then the problem again admits a maxflow formulation. On the other hand, s is $\langle 1,0,1\rangle$-contaminated at $m \geqslant N_0$ for some constant N_0, then the Possible Winner problem is NP-complete even with 3 undetermined pairs of candidates per vote.

▶ **Theorem 9.** *[⋆] Let s be a smooth, Borda-like, $\langle 1,1\rangle$-difference-free scoring rule. If there exists a constant $N_0 \in \mathbb{N}^+$ such that s is $\langle 1,0,1\rangle$-contaminated for all $m \geqslant n_0$, then the* Possible Winner *problem is* NP*-complete, even if every vote has at most* 3 *undetermined pairs. On the other hand, if s is $\langle 1,0,1\rangle$-difference-free, then the* Possible Winner *problem for s is in* P *if every vote has at most* 3 *undetermined pairs.*

▶ Remark. Note that unlike the previous two results, this statement is not a complete classification, because we don't have an appropriate analog of Propositions 3.1 and 3.1. Having said that, our result holds for a more general class of scoring rules: those where s is $\langle 1,0,1\rangle$-contaminated at m "sufficiently" often, that is to say that if $\overrightarrow{s_m}$ is $\langle 1,0,1\rangle$-contaminated and $m' > m$ is the smallest natural number for which $\overrightarrow{s_m}$ is $\langle 1,0,1\rangle$-contaminated, then $m' - m$ is bounded by some polynomial function of m, by inserting appropriately many dummy candidates using standard techniques.

We now turn to our final result for scoring rules. Let s be a smooth, Borda-like scoring rule that is $\langle 1,1\rangle$-difference-free. Then we have the following. If s is $\langle 0,1,0\rangle$-contaminated, then the Possible Winner problem for s is NP-complete even when every vote has at most 4 undetermined pairs of candidates. If s is $\langle 0,1,0\rangle$-difference-free, then notice that $d(\overrightarrow{s_m})$ for any suitably large $m \in \mathbb{N}^+$ can contain at most two ones (since s is also $\langle 1,1\rangle$-difference-free). If the number of ones in $d(\overrightarrow{s_m})$ is one, then $d(\overrightarrow{s_m})$ either has a one on the first or the last coordinate (recall that s is $\langle 0,1,0\rangle$-difference-free), corresponding to the plurality and veto voting rules, respectively. On the other hand, if the number of ones is two, $d(\overrightarrow{s_m}) = \langle 1,0,\ldots,0,1\rangle$, which is equivalent (in normal form) to the scoring rule $(2,1,\ldots,1,0)$. The Possible Winner problem is polynomial time solvable for plurality and veto voting rules, and we show here that it is also polynomially solvable for the scoring rule $(2,1,\ldots,1,0)$ as long as the number of undetermined pairs of candidates in any vote is at most $m-1$. We note that the status for the Possible Winner problem for this rule was left unresolved in [6] and was later resolved in [2]. If we allow for m or more undetermined pairs of candidates in every vote, then we show that the Possible Winner problem is NP-complete. As before, we will need the following property of $\langle 1,0,1\rangle$-contaminated vectors.

[⋆] Let s be a normalized smooth scoring rule that is not $\langle 1,0,1\rangle$-difference-free. Then there exists some $n_0 \in \mathbb{N}^+$ such that s is $\langle 0,1,0\rangle$-contaminated at m for every $m > n_0$.

We now state the final result in this section. It is easily checked that the result accounts for all smooth, Borda-like scoring rules that are $\langle 1,1\rangle$-difference-free.

▶ **Theorem 10.** *[⋆] Let s be a smooth, Borda-like scoring rule that is $\langle 1,1\rangle$-difference-free. Then we have the following.*

1. *If s is $\langle 0,1,0\rangle$-contaminated, then the* Possible Winner *problem for s is* NP*-complete even when every vote has at most* 4 *undetermined pairs of candidates.*

2. *If s is equivalent to $(2, 1, \ldots, 1, 0)$, then* Possible Winner *is* NP*-complete even when the number of undetermined pairs of candidates in every vote is at most $m-1$.*
3. *If s is equivalent to $(2, 1, \ldots, 1, 0)$ and the number of undetermined pairs of candidates is strictly less than $m-1$, then* Possible Winner *is in* P*.*
4. *If s is neither $\langle 0, 1, 0\rangle$-contaminated nor equivalent to $(2, 1, \ldots, 1, 0)$, then s is equivalent to either the plurality or veto scoring rules and* Possible Winner *is in* P *for these cases.*

3.2 Copeland$^\alpha$ Voting Rule

We now turn to the Copeland$^\alpha$ voting rule. We show in Theorem 11 below that the Possible Winner problem is NP-complete for the Copeland$^\alpha$ voting rule even when every vote has at most 2 undetermined pairs of candidates for every $\alpha \in [0, 1]$.

▶ **Theorem 11.** *[⋆] The* Possible Winner *problem is* NP*-complete for the Copeland$^\alpha$ voting rule even if the number of undetermined pairs of candidates in every vote is at most 2 for every $\alpha \in [0, 1]$.*

We prove in Theorem 12 that the number of undetermined pairs of candidates in Theorem 11 is tight for the Copeland^0 and Copeland^1 voting rules.

▶ **Theorem 12.** *[⋆] The* Possible Winner *problem is in* P *for the Copeland^0 and Copeland^1 voting rules if the number of undetermined pairs of candidates in every vote is at most 1.*

We show next that the Possible Winner problem is NP-complete for the Copeland$^\alpha$ voting rule even if the number of undetermined pairs of candidates in every vote is at most 1 for $\alpha \in (0, 1)$. We break the proof into two parts — Lemma 13 proves the result for every $\alpha \in (0, 1/2]$ and Lemma 14 proves for every $\alpha \in [1/2, 1)$.

▶ **Lemma 13.** *[⋆] The* Possible Winner *problem is* NP*-complete for the Copeland$^\alpha$ voting rule even if the number of undetermined pairs in every vote is at most 1 for every $\alpha \in (0, 1/2]$.*

Proof. The Possible Winner problem for the Copeland$^\alpha$ voting rule is clearly in NP. To prove NP-hardness of Possible Winner, we reduce Possible Winner from $(3, B2)$–SAT. Let $\mathcal{I}$ be an instance of $(3, B2)$–SAT, over the variables $\mathcal{V} = \{x_1, \ldots, x_n\}$ and with clauses $\mathcal{T} = \{c_1, \ldots, c_m\}$. We construct an instance $\mathcal{I}'$ of Possible Winner from $\mathcal{I}$ as follows.

Set of candidates: $\mathcal{C} = \{x_i, \bar{x}_i, d_i : i \in [n]\} \cup \{c_i : i \in [m]\} \cup \{c\} \cup \mathcal{G}$, where $\mathcal{G} = \{g_1, \ldots, g_{mn}\}$.

For every $i \in [n]$, let us define $\mathfrak{p}^1_{x_i}, \mathfrak{p}^2_{x_i} : x_i \succ d_i \succ$ others and $\mathfrak{p}^1_{\bar{x}_i}, \mathfrak{p}^2_{\bar{x}_i} : \bar{x}_i \succ d_i \succ$ others. Using $\mathfrak{p}^1_{x_i}, \mathfrak{p}^2_{x_i}, \mathfrak{p}^1_{\bar{x}_i}, \mathfrak{p}^2_{\bar{x}_i}$, we define the partial votes $\mathfrak{p}^{1\prime}_{x_i}, \mathfrak{p}^{2\prime}_{x_i}, \mathfrak{p}^{1\prime}_{\bar{x}_i}, \mathfrak{p}^{2\prime}_{\bar{x}_i}$ as follows.

$$\mathfrak{p}^{1\prime}_{x_i}, \mathfrak{p}^{2\prime}_{x_i} : \mathfrak{p}^1_{x_i} \setminus \{(x_i, d_i)\}\ ,\ \mathfrak{p}^{1\prime}_{\bar{x}_i}, \mathfrak{p}^{2\prime}_{\bar{x}_i} : \mathfrak{p}^1_{\bar{x}_i} \setminus \{(\bar{x}_i, d_i)\}$$

Let a clause c_j involves the literals $\ell^1_j, \ell^2_j, \ell^3_j$. For every $j \in [m]$, let us consider the following votes $\mathfrak{q}_j(\ell^1_j), \mathfrak{q}_j(\ell^2_j), \mathfrak{q}_j(\ell^3_j)$.

$$\mathfrak{q}_j(\ell^k_j) : c_j \succ \ell^k_j \succ \text{others}, \forall k \in \{1, 2, 3\}$$

Using $\mathfrak{q}_j(\ell^1_j), \mathfrak{q}_j(\ell^2_j), \mathfrak{q}_j(\ell^3_j)$, we define the partial votes $\mathfrak{q}'_j(\ell^1_j), \mathfrak{q}'_j(\ell^2_j), \mathfrak{q}'_j(\ell^3_j)$ as follows.

$$\mathfrak{q}'_j(\ell^k_j) : \mathfrak{q}_j(\ell^k_j) \setminus \{(c_j, \ell^k_j)\}, \forall k \in \{1, 2, 3\}$$

Let us define

$$\mathcal{P} = \cup_{i\in[n]} \{\mathfrak{p}^1_{x_i}, \mathfrak{p}^2_{x_i}, \mathfrak{p}^1_{\bar{x}_i}, \mathfrak{p}^2_{\bar{x}_i}\} \cup_{j\in[m]} \{\mathfrak{q}_j(\ell^1_j), \mathfrak{q}_j(\ell^2_j), \mathfrak{q}_j(\ell^3_j)\}$$

Table 3 Summary of Copeland$^\alpha$ scores of the candidates from $\mathcal{P} \cup \mathcal{Q}$. All the wins and defeats in the table are by a margin of 2.

Candidates	Copeland$^\alpha$ score	Winning against	Losing against	Tie with
c	$(2n+m)\alpha$ $+n+{}^{3mn}/_4$	$G' \subset \mathcal{G}, \lvert G'\rvert = n + {}^{3mn}/_4$	$\mathcal{G} \setminus G', \lvert G'\rvert = n + {}^{3mn}/_4$ $d_i, \forall i \in [n]$	$x_i, \bar{x}_i \forall i \in [n]$ $c_j \forall j \in [m]$
$x_i, \forall i \in [n]$	$(2n+m)\alpha$ $+n+{}^{3mn}/_4$	$G'' \subset \mathcal{G}, \lvert G''\rvert = {}^{3mn}/_4$ $d_i \forall i \in [n]$	$\mathcal{G} \setminus (G' \cup G'')$	$c, G' \subset \mathcal{G}, \lvert G'\rvert = m$ $x_j, \forall j \in [n] \setminus \{i\}$ $\bar{x}_j \forall j \in [n]$
$\bar{x}_i, \forall i \in [n]$	$(2n+m)\alpha$ $+n+{}^{3mn}/_4$	$G'' \subset \mathcal{G}, \lvert G''\rvert = {}^{3mn}/_4$ $d_i \forall i \in [n]$	$\mathcal{G} \setminus (G' \cup G'')$	$c, G' \subset \mathcal{G}, \lvert G'\rvert = m$ $\bar{x}_j, \forall j \in [n] \setminus \{i\}$ $x_j \forall j \in [n]$
$c_j, \forall j \in [m]$	$(2n+m-1)\alpha$ $+n+{}^{3mn}/_4+1$	$x_i, \bar{x}_i \forall i \in [n]$ $G' \subset \mathcal{G}, \lvert G'\rvert = {}^{3mn}/_4 - n + 1$	$\mathcal{G} \setminus (G' \cup G'')$ $d_i, \forall i \in [n]$	c $c_j \forall j \in [m] \setminus \{i\}$ $G'' \subset \mathcal{G}, \lvert G''\rvert = 2n - 1$
$d_i, i \in [n]$	$(2n+m)\alpha$ $+n+{}^{3mn}/_4-1$	$c, c_j, \forall j \in [m]$ $G'' \subset \mathcal{G}, \lvert G''\rvert = {}^{3mn}/_4 - m + n - 2$	$x_i, \bar{x}_i \forall i \in [n]$ $\mathcal{G} \setminus (G' \cup G'')$	$G' \subset \mathcal{G}, \lvert G'\rvert = 2n + m$
$g_i, \forall i \in [mn]$	$< {}^{3mn}/_4$		$\forall j \in \{i+k : k \in [\lfloor {}^{(mn-1)}/_2 \rfloor]$	

and

$$\mathcal{P}' = \cup_{i\in[n]}\{\mathfrak{p}^{1\prime}_{x_i}, \mathfrak{p}^{2\prime}_{x_i}, \mathfrak{p}^{1\prime}_{\bar{x}_i}, \mathfrak{p}^{2\prime}_{\bar{x}_i}\} \cup_{j\in[m]} \{\mathfrak{q}'_j(\ell^1_j), \mathfrak{q}'_j(\ell^2_j), \mathfrak{q}'_j(\ell^3_j)\}.$$

There exists a set of complete votes $\mathcal{Q}$ of size polynomial in n and m which realizes Table 3 [24]. All the wins and defeats in Table 3 are by a margin of 2. We now define the instance $\mathcal{I}'$ of Possible Winner to be $(\mathcal{C}, \mathcal{P}' \cup \mathcal{Q}, c)$. Notice that the number of undetermined pairs of candidates in every vote in $\mathcal{I}'$ is at most 1. This finishes the description of the Possible Winner instance. We defer the formal proof of equivalence of the two instances and the polynomial time solvable case to the appendix. ◀

Next we present Lemma 14 which resolves the complexity of the Possible Winner problem for the Copeland$^\alpha$ voting rule for every $\alpha \in [{}^1/_2, 1)$ when every partial vote has at most one undetermined pair of candidates.

▶ **Lemma 14.** *[⋆] The* Possible Winner *problem is* NP*-complete for the Copeland$^\alpha$ voting rule even if the number of undetermined pairs in every vote is at most* 1 *for every* $\alpha \in [{}^1/_2, 1)$.

We get the following result for the Copeland$^\alpha$ voting rule from Theorem 13 and 14.

▶ **Theorem 15.** *The* Possible Winner *problem is* NP*-complete for the Copeland$^\alpha$ voting rule even if the number of undetermined pairs of candidates in every vote is at most* 1 *for every* $\alpha \in (0, 1)$.

3.3 Maximin and Bucklin Voting Rules

To prove our hardness result for the maximin voting rule, we reduce the Possible Winner problem from the d–Multicolored Independent Set problem which is defined as below. d–Multicolored Independent Set is known to be NP-complete (for example, see this [10]). We denote arbitrary instance of d–Multicolored Independent Set by $\left(\mathcal{V} = \uplus_{i=1}^k \mathcal{V}_k, \mathcal{E}\right)$.

▶ **Definition 16** (d–Multicolored Independent Set). Given a d-regular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, an integer k, and a partition of the set of vertices $\mathcal{V}$ into k independent sets $\mathcal{V}_1, \ldots, \mathcal{V}_k$, that is $\mathcal{V} = \cup_{i\in[k]}\mathcal{V}_i$ and $\mathcal{V}_i$ is an independent set for every $i \in [k]$, does there exists an independent set $\mathcal{S} \subset \mathcal{V}$ in $\mathcal{G}$ such that $\lvert\mathcal{S} \cap \mathcal{V}_i\rvert = 1$ for every $i \in [k]$.

Table 4 Pairwise margins of candidates from $\mathcal{P} \cup \mathcal{Q}$.

$\forall e \in \mathcal{E}, \mathcal{D}_{\mathcal{P}\cup\mathcal{Q}}(e, c) = \lambda$	$\forall i \in [k], \forall u \in \mathcal{V}_i, \mathcal{D}_{\mathcal{P}\cup\mathcal{Q}}(u, g_i) = \lambda - 2d$
$\forall i \in [k], \forall u \in \mathcal{V}_i, \mathcal{D}_{\mathcal{P}\cup\mathcal{Q}}(g'_i, u) = \lambda + 2d$	$\forall i \in [k], e \in \mathcal{E}, \mathcal{D}_{\mathcal{P}\cup\mathcal{Q}}(e, g'_i) = \lambda$
$\forall e = (u_i, u_j) \in \mathcal{E}, \mathcal{D}_{\mathcal{P}\cup\mathcal{Q}}(u_i, e) = \mathcal{D}_{\mathcal{P}\cup\mathcal{Q}}(u_j, e) = \lambda - 2$	

Table 5 Summary of initial Copeland scores of the candidates.

Candidate	maximin score	worst against	Candidate	maximin score	worst against
c	$-\lambda$	$e \in \mathcal{E}$	$(u_i, u_j) \in \mathcal{E}$	$-(\lambda - 2)$	u_i, u_j
$u \in \mathcal{V}_i$	$-(\lambda + 2d)$	g'_i	g_i	$-(\lambda - 2d)$	$u \in \mathcal{V}_i$
g'_i	$-\lambda$	$e \in \mathcal{E}$			

Now we prove our hardness result for the Possible Winner problem for the maximin voting rule in Theorem 17.

▶ **Theorem 17.** *The* Possible Winner *problem is* NP*-complete for the maximin voting rule even if the number of undetermined pairs of candidates in every vote is at most* 2*.*

Proof. The Possible Winner problem for the maximin voting rule is clearly in NP. To prove NP-hardness of Possible Winner, we reduce Possible Winner from d–Multicolored Independent Set. Let $\mathcal{I} = \left(\mathcal{V} = \cup_{i=1}^{k} \mathcal{V}_k, \mathcal{E}\right)$ be an arbitrary instance of d–Multicolored Independent Set. We construct an instance $\mathcal{I}'$ of Possible Winner from $\mathcal{I}$ as follows.

Set of candidates: $\mathcal{C} = \mathcal{V} \cup \mathcal{E} \cup \{c\} \cup \{g_i, g'_i : i \in [k]\}$

For every $u \in \mathcal{V}_i$ and $\ell \in [d]$, let us consider the following vote $\mathfrak{p}_u$.

$$\mathfrak{p}_u^\ell = \overrightarrow{(\mathcal{C} \setminus \{u, g_i, g'_i\})_u} \succ g_i \succ g'_i \succ u, \text{where } \overrightarrow{(\mathcal{C} \setminus \{u, g_i, g'_i\})_u}$$
is any fixed ordering of $\mathcal{C} \setminus \{u, g_i, g'_i\}$

Using $\mathfrak{p}_u^\ell$, we define a partial vote $\mathfrak{p}_u'^\ell$ as $\mathfrak{p}_u'^\ell = \mathfrak{p}_u^\ell \setminus \{(g_i, u), (g'_i, u)\}$. For every edge $e = (u_i, u_j)$ where $u_i \in \mathcal{V}_i$ and $u_j \in \mathcal{V}_j$, let us consider the following votes $\mathfrak{p}_{e,u_i}$ and $\mathfrak{p}_{e,u_j}$.

$$\mathfrak{p}_{e,u_i} = \overrightarrow{(\mathcal{C} \setminus \{u_i, g'_i, e\})} \succ e \succ g'_i \succ u_i \ , \ \mathfrak{p}_{e,u_j} = \overrightarrow{(\mathcal{C} \setminus \{u_j, g'_j, e\})} \succ e \succ g'_j \succ u_j$$

Using $\mathfrak{p}_{e,u_i}$ and $\mathfrak{p}_{e,u_j}$, we define the partial votes $\mathfrak{p}'_{e,u_i}$ and $\mathfrak{p}'_{e,u_j}$ as follows.

$$\mathfrak{p}'_{e,u_i} = \mathfrak{p}_{e,u_i} \setminus \{(e, u_i), (g'_i, u_i)\} \ , \ \mathfrak{p}'_{e,u_j} = \mathfrak{p}_{e,u_j} \setminus \{(e, u_j), (g'_j, u_j)\}$$

Let us call $\mathfrak{p}_e = \{\mathfrak{p}_{e,u_i}, \mathfrak{p}_{e,u_j}\}$ and $\mathfrak{p}'_e = \{\mathfrak{p}'_{e,u_i}, \mathfrak{p}'_{e,u_j}\}$. Let us define $\mathcal{P} = \cup_{u\in\mathcal{V},\ell\in[d]} \mathfrak{p}_u^\ell \cup_{e\in\mathcal{E}} \mathfrak{p}_e$ and $\mathcal{P}' = \cup_{u\in\mathcal{V},\ell\in[d]} \mathfrak{p}_u'^\ell \cup_{e\in\mathcal{E}} \mathfrak{p}'_e$. There exists a set of complete votes $\mathcal{Q}$ of size polynomial in $|\mathcal{V}|$ and $|\mathcal{E}|$ with the pairwise margins as in Table 4 [24]. Let $\lambda > 3d$ be any positive even integer.

For every pair of candidates $(c_i, c_j) \in \mathcal{C} \times \mathcal{C}$ whose pairwise margin is not defined above, we define $\mathcal{D}_{\mathcal{P}\cup\mathcal{Q}}(c_i, c_j) = 0$. We summarize the maximin score of every candidate in $\mathcal{P} \cup \mathcal{Q}$ in Table 5. We now define the instance $\mathcal{I}'$ of Possible Winner to be $(\mathcal{C}, \mathcal{P}' \cup \mathcal{Q}, c)$. Notice that the number of undetermined pairs of candidates in every vote in $\mathcal{I}'$ is at most 2. This finishes the description of the Possible Winner instance. We claim that $\mathcal{I}$ and $\mathcal{I}'$ are equivalent.

In the forward direction, suppose that $\mathcal{I}$ be a YES instance of d–MULTICOLORED INDEPENDENT SET. Then there exists $u_i \in \mathcal{V}_i$ for every $i \in [k]$ such that $\mathcal{U} = \{u_i : i \in [k]\}$ forms an independent set. We extend the partial vote $\mathfrak{p}_u'^{\ell}$ for every $u \in \mathcal{V}_i, i \in [k], \ell \in [d]$ to $\bar{\mathfrak{p}}_u^{\ell}$ as follows.

$$\bar{\mathfrak{p}}_u^{\ell} = \begin{cases} \overrightarrow{(\mathcal{C} \setminus \{u, g_i, g_i'\})_u} \succ u \succ g_i \succ g_i' & u \in \mathcal{U} \\ \overrightarrow{(\mathcal{C} \setminus \{u, g_i, g_i'\})_u} \succ g_i \succ g_i' \succ u & u \notin \mathcal{U} \end{cases}$$

For every $e = (u_i, u_j)$, we extend $\mathfrak{p}_{e,u_i}'$ and $\mathfrak{p}_{e,u_j}'$ to $\bar{\mathfrak{p}}_{e,u_i}$ and $\bar{\mathfrak{p}}_{e,u_j}$. Since $\mathcal{U}$ is an independent set, at least one of u_i and u_j does not belong to $\mathcal{U}$. Without loss of generality, let us assume $u_i \notin \mathcal{U}$.

$$\bar{\mathfrak{p}}_{e,u_i} = \overrightarrow{(\mathcal{C} \setminus \{u_i, g_i', e\})} \succ u_i \succ e \succ g_i' \ , \ \bar{\mathfrak{p}}_{e,u_j} = \overrightarrow{(\mathcal{C} \setminus \{u_j, g_j', e\})} \succ e \succ g_j' \succ u_j$$

Let us call $\bar{\mathfrak{p}}_e = \{\bar{\mathfrak{p}}_{e,u_i}, \bar{\mathfrak{p}}_{e,u_j}\}$. We consider the extension of $\mathcal{P}$ to $\bar{\mathcal{P}} = \cup_{u \in \mathcal{V}, \ell \in [d]} \bar{\mathfrak{p}}_u^{\ell} \cup_{e \in \mathcal{E}} \bar{\mathfrak{p}}_e$. We claim that c is a co-winner in the profile $\bar{\mathcal{P}} \cup \mathcal{Q}$ since the maximin score of c, g_i, g_i' for every $i \in [k]$, $u \in \mathcal{V}$, and $e \in \mathcal{E}$ in $\bar{\mathcal{P}} \cup \mathcal{Q}$ is $-\lambda$.

In the reverse direction suppose the POSSIBLE WINNER instance $\mathcal{I}'$ be a YES instance. Then there exists an extension of the set of partial votes $\mathcal{P}'$ to a set of complete votes $\bar{\mathcal{P}}$ such that c is a co-winner in $\bar{\mathcal{P}} \cup \mathcal{Q}$. Let us call the extension of $\mathfrak{p}_u'^{\ell}$ in $\bar{\mathcal{P}}$ $\bar{\mathfrak{p}}_u^{\ell}$, $\mathfrak{p}_{e,u_i}'$ and $\mathfrak{p}_{e,u_j}'$ in $\bar{\mathcal{P}}$ $\bar{\mathfrak{p}}_{e,u_i}$ and $\bar{\mathfrak{p}}_{e,u_j}$ respectively. First we notice that the maximin score of c in $\bar{\mathcal{P}} \cup \mathcal{Q}$ is $-\lambda$ since the relative ordering of c with respect to every other candidate is already fixed in $\mathcal{P}' \cup \mathcal{Q}$. Now we observe that, in $\bar{\mathcal{P}} \cup \mathcal{Q}$, the maximin score of g_i for every $i \in [k]$ is $-(\lambda - 2d)$. Hence, for c to co-win, there must exists at least one $u_i^* \in \mathcal{V}_i$ for every $i \in [k]$ such that $u_i^* \succ g_i \succ g_i'$ in $\bar{\mathfrak{p}}_{u_i^*}^{\ell}$ for every $\ell \in [d]$. We claim that $\mathcal{U} = \{u_i^* : i \in [k]\}$ is an independent set in $\mathcal{I}$. If not, then suppose there exists an edge e between u_i^* and u_j^* for some $i, j \in [k]$. Now notice that, for c to co-win either $u_i^* \succ e \succ g_i'$ in $\bar{\mathfrak{p}}_{e,u_i^*}$ or $u_j^* \succ e \succ g_j'$ in $\bar{\mathfrak{p}}_{e,u_j^*}$. However, this makes the maximin score of either u_i^* or u_j^* strictly more than $-\lambda$ contradicting our assumption that c co-wins the election. Hence, $\mathcal{U}$ forms an independent set in $\mathcal{I}$. ◀

We next prove in Theorem 18 that the maximum number of undetermined pairs of candidates in Theorem 17 is tight.

▶ **Theorem 18.** *[⋆] The* POSSIBLE WINNER *problem is in* P *for the maximin voting rule if the number of undetermined pairs of candidates in every vote is at most* 1.

Finally, we state our results for the Bucklin voting rule.

▶ **Theorem 19.** *[⋆] The* POSSIBLE WINNER *problem is* NP*-complete for the Bucklin voting rule even if the number of undetermined pairs of candidates in every vote is at most* 2*, and is in* P *if the number of undetermined pairs of candidates in every vote is at most* 1.

4 Conclusion

We have demonstrated the exact minimum number of undetermined pairs allowed per vote which keeps the POSSIBLE WINNER winner problem NP-complete, and we were able to address a large class of scoring rules, Copeland$^{\alpha}$, maximin, and Bucklin voting rules. Our results generalize many of the known hardness results in the literature, and show that for many voting rules, we need a surprisingly small number of undetermined pairs (often just one or two) for the POSSIBLE WINNER problem to be NP-complete. In the context of scoring rules, it would be interesting to extend these tight results to the class of pure scoring rules, and to extend Theorem 9 to account for all smooth scoring rules.

References

1 Dorothea Baumeister, Magnus Roos, and Jörg Rothe. Computational complexity of two variants of the possible winner problem. In *Proc. International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 853–860, 2011.

2 Dorothea Baumeister and Jörg Rothe. Taking the final step to a full dichotomy of the possible winner problem in pure scoring rules. *Inf. Process. Lett.*, 112(5):186–190, 2012. doi:10.1016/j.ipl.2011.11.016.

3 Piotr Berman, Marek Karpinski, and Alex D. Scott. Approximation hardness and satisfiability of bounded occurrence instances of SAT. *Electronic Colloquium on Computational Complexity (ECCC)*, 10(022), 2003. URL: http://eccc.hpi-web.de/eccc-reports/2003/TR03-022/index.html.

4 Nadja Betzler, Robert Bredereck, and Rolf Niedermeier. Partial kernelization for rank aggregation: theory and experiments. In *Proc. 5th International Symposium on Parameterized and Exact Computation (IPEC)*, pages 26–37. Springer, 2010.

5 Nadja Betzler, Robert Bredereck, and Rolf Niedermeier. Theoretical and empirical evaluation of data reduction for exact kemeny rank aggregation. *Autonomous Agents and Multi-Agent Systems*, 28(5):721–748, 2014. doi:10.1007/s10458-013-9236-y.

6 Nadja Betzler and Britta Dorn. Towards a dichotomy of finding possible winners in elections based on scoring rules. In *Proc. 34th Mathematical Foundations of Computer Science (MFCS)*, pages 124–136. Springer, 2009.

7 Nadja Betzler, Susanne Hemmann, and Rolf Niedermeier. A Multivariate Complexity Analysis of Determining Possible Winners given Incomplete Votes. In *Proc. 21st International Joint Conference on Artificial Intelligence (IJCAI)*, volume 9, pages 53–58, 2009.

8 Yann Chevaleyre, Jérôme Lang, Nicolas Maudet, and Jérôme Monnot. Possible winners when new candidates are added: The case of scoring rules. In *Proc. 24th International Conference on Artificial Intelligence (AAAI)*, 2010.

9 William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. *J. Artif. Int. Res.*, 10(1):243–270, May 1999. URL: http://dl.acm.org/citation.cfm?id=1622859.1622867.

10 Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015. doi:10.1007/978-3-319-21275-3.

11 Palash Dey. Resolving the complexity of some fundamental problems in computational social choice. *CoRR*, abs/1703.08041, 2017. URL: http://arxiv.org/abs/1703.08041.

12 Palash Dey and Neeldhara Misra. On the exact amount of missing information that makes finding possible winners hard. *CoRR*, abs/1610.08407, 2016. URL: http://arxiv.org/abs/1610.08407.

13 Palash Dey, Neeldhara Misra, and Y. Narahari. Kernelization complexity of possible winner and coalitional manipulation problems in voting. In *Proc. 14th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*, pages 87–96, 2015. URL: http://dl.acm.org/citation.cfm?id=2772894.

14 Palash Dey, Neeldhara Misra, and Y. Narahari. Complexity of manipulation with partial information in voting. In *Proc. 25th International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, USA*, pages 229–235, 2016. URL: http://www.ijcai.org/Abstract/16/040.

15 Palash Dey, Neeldhara Misra, and Y. Narahari. Frugal bribery in voting. In *Proc. 30th AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pages 2466–2472, 2016. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12133.

16 Palash Dey, Neeldhara Misra, and Y. Narahari. Kernelization complexity of possible winner and coalitional manipulation problems in voting. *Theor. Comput. Sci.*, 616:111–125, 2016. doi:10.1016/j.tcs.2015.12.023.

17 Palash Dey, Neeldhara Misra, and Y. Narahari. Frugal bribery in voting. *Theor. Comput. Sci.*, 676:15–32, 2017. doi:10.1016/j.tcs.2017.02.031.

18 Piotr Faliszewski, Yannick Reisch, Jörg Rothe, and Lena Schend. Complexity of manipulation, bribery, and campaign management in bucklin and fallback voting. In *Proc. 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1357–1358. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

19 Michael R Garey and David S Johnson. *Computers and Intractability*, volume 174. freeman New York, 1979.

20 Benjamin G. Jackson, Patrick S. Schnable, and Srinivas Aluru. Consensus genetic maps as median orders from inconsistent sources. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 5(2):161–171, 2008. doi:10.1145/1371585.1371586.

21 Kathrin Konczak and Jérôme Lang. Voting procedures with incomplete preferences. In *Proc. 19th International Joint Conference on Artificial Intelligence-05 Multidisciplinary Workshop on Advances in Preference Handling*, volume 20, 2005.

22 Jérôme Lang, Maria Silvia Pini, Francesca Rossi, Domenico Salvagnin, Kristen Brent Venable, and Toby Walsh. Winner determination in voting trees with incomplete preferences and weighted votes. *Auton. Agent Multi Agent Syst.*, 25(1):130–157, 2012.

23 Jérôme Lang, Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. Winner determination in sequential majority voting. In *Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI)*, volume 7, pages 1372–1377, 2007.

24 David C McGarvey. A theorem on the construction of voting paradoxes. *Econometrica*, pages 608–610, 1953.

25 Hervé Moulin, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

26 David M. Pennock, Eric Horvitz, and C. Lee Giles. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *Proc. 17th National Conference on Artificial Intelligence and 12th Conference on on Innovative Applications of Artificial Intelligence, July 30 - August 3, 2000, Austin, Texas, USA.*, pages 729–734, 2000. URL: http://www.aaai.org/Library/AAAI/2000/aaai00-112.php.

27 Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. Incompleteness and incomparability in preference aggregation. In *Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI)*, volume 7, pages 1464–1469, 2007.

28 Toby Walsh. Uncertainty in preference elicitation and aggregation. In *Proc. 22nd International Conference on Artificial Intelligence (AAAI)*, volume 22, page 3, 2007.

29 Lirong Xia and Vincent Conitzer. Determining possible and necessary winners under common voting rules given partial orders. *J. Artif. Intell. Res.*, 41(2):25–67, 2011.