

# Approximating Voting Rules from Truncated Ballots

Manel Ayadi<sup>1,2</sup>, Nahla Ben Amor<sup>1</sup>, and Jérôme Lang<sup>2</sup>

<sup>1</sup> LARODEC, Institut Supérieur de Gestion, Université de Tunis, Tunis, Tunisia

<sup>2</sup> LAMSADE, CNRS, Université Paris-Dauphine, PSL, Paris, France

**Abstract.** Classical voting rules assume that ballots are complete preference orders over candidates. However, when the number of candidates is large enough, it is too costly to ask the voters to rank all candidates. We suggest to fix a rank  $k$ , to ask all voters to specify their best  $k$  candidates, and then to consider “top- $k$  approximations” of rules, which take only into account the *top- $k$*  candidates of each ballot. We consider two measures of the quality of the approximation: the probability of selecting the same winner as the original rule, and the score ratio. We do a worst-case study (for the latter measure only), and for both measures, an average-case study and a study from real data sets.

**Keywords:** Voting rules · Truncated ballots · Approximations.

## 1 Introduction

The input of a voting rule is usually a collection of complete rankings over candidates (although there are exceptions, such as approval voting). However, requiring a voter to provide a complete ranking over the whole set of candidates can be difficult and costly in terms of time and cognitive effort. We suggest to ask voters to report only their *top- $k$*  candidates, for some (small) fixed value of  $k$  (the obtained ballots are then said to be *top- $k$* ). Not only it saves communication effort, but it is also often easier for a voter to find out the top part of their preference relation than the bottom part. However, this raises the issue of how usual voting rules should be adapted to top- $k$  ballots. Reporting top- $k$  ballots is a specific form of *voting with incomplete preferences*, and is highly related to *vote elicitation*. Work on these topics is reviewed in the recent handbook chapter [5]. Existing work on truncated ballots can be classified into two classes according to the type of interaction with the voters:

### (i) Interactive elicitation

An interactive elicitation protocol asks voters to expand their truncated ballots in an incremental way, until the outcome of the vote is eventually determined. This line of research starts with Kalech et al. [14] who start by top-1 ballots, then top-2, etc., until there is sufficient information for knowing the winner. Lu and Boutilier [17,16] propose an incremental elicitation process using *minimax regret* to predict the correct winner given partial information. A more general incremental elicitation framework, with more types of elicitation questions, is cost-effective elicitation [25]. Naamani Dery et al. [10] present two elicitation algorithms for finding a winner with little communication between voters.

**(ii) Non-interactive elicitation**

The central authority elicits the top- $k$  ballots at once, for a fixed value of  $k$ , and outputs a winner without requiring voters to provide extra information. A possibility consists in computing possible winners given these truncated ballots: this is the path followed by Baumeister *et al.* [2] (who also consider double-truncated ballots where each voter ranks some of her top and bottom candidates). Another possibility – which is the one we follow – consists in generalizing the definition of a voting rule so that it takes truncated ballots as input. In this line, Oren *et al.* [21] analyze *top-k* voting by assessing the values of  $k$  needed to ensure the true winner is found with high probability for specific preference distributions. Skowron *et al.* [23] use *top-k* voting as a way to approximate some multiwinner rules. Filmus and Oren [12] study the performance of top- $k$  voting under the impartial culture distribution for the Borda, Harmonic and Copeland rules. They assess the values of  $k$  needed to find the true winner with high probability, and they report on numerical experiments that show that under the impartial culture, top- $k$  ballots for reasonable small values of  $k$  give accurate results.

Bentert and Skowron [3] focus on top- $k$  approximations of voting rules that are defined via the maximization of a score (positional scoring rules and maximin). They evaluate the quality of the approximation of a voting rule by a top- $k$  rule by the worst-case ratio between the scores, with respect to the original profile, of the winner of the original rule and the winner of the approximate rule. They identify the top- $k$  rules that best approximate positional scoring rules (we give more details in Section 5). Their theoretical analysis is completed by numerical experiments using profiles generated from different distributions over preferences: they show that for the Borda rule a small value of  $k$  is needed to achieve a high approximation guarantee while maximin needs more information from a sufficiently many voters to determine the winner.

Ayadi *et al.* [1] evaluate the extent to which STV with *top-k* ballots approximates STV with full information. They show that for small  $k$ , *top-k* ballots are enough to identify the correct winner quite frequently, especially for data taken from real elections. Finally, the recognition of singled-peaked *top-k* profiles is studied in [15] while the computational issues of manipulating rules with *top-k* profiles is addressed in [20].

Our contribution concerns non-interactive elicitation. We adapt different voting rules to truncated ballots: we define approximations of voting rules which take as input the *top-k* candidates of each ballot. The question is then, *are these approximations good predictors of the original rule?* We answer this question by considering two measures: the probability that the approximate rule selects the ‘true’ winner, and the ratio between the scores (for the original rule) of the true winner and the winner of the approximate rule. For the latter measure we give a worst-case theoretical analysis. For both measures we give an empirical study, based on randomly generated profiles and on real-world data. Our findings are that for several common voting rules, both for randomly generated profiles and real data, a very small  $k$  suffices.

Our research can be seen as a continuation of Filmus and Oren [12]. We go further on several points: we consider more voting rules; beyond impartial culture, we consider a large scope of distributions; we study score distortion; and we include experiments using real-world data sets. Our work is also closely related to [3], who have obtained related results independently (see Sections 4 and 5 for a discussion).

Our interpretation of top- $k$  ballots is *epistemic*: the central authority in charge of collecting the votes and computing the outcome ignores the voters' preferences below the *top- $k$*  candidates of each voter, and has to cope with it as much as possible. Voters may very well have a complete preference order in their head (although it does not need to be the case), but they will simply not be asked to report it.

Section 2 gives some background. Section 3 defines top- $k$  approximations of different voting rules. Section 4 analyses empirically the probability that approximate rules select the true winner. Section 5 analyses score distortion, theoretically and empirically.

## 2 Preliminaries

An election is a triple  $E = \langle N, A, P \rangle$  where:  $N = \{1, \dots, n\}$  is the set of *voters*,  $A$  is the set of *candidates*, with  $|A| = m$ ; and  $P = (\succ_1, \dots, \succ_n)$  is the *preference profile* of voters in  $N$ , where for each  $i$ ,  $\succ_i \in P$  is a linear order over  $A$ .  $\mathcal{P}_m$  is the set of all profiles over  $m$  alternatives (for varying  $n$ ).

Given a profile  $P$ ,  $N_P(a, b) = \#\{i, a \succ_i b\}$  is the number of voters who prefer  $a$  to  $b$  in  $P$ . The *majority graph*  $M(P)$  is the graph whose set of vertices is the set of the candidates  $A$  and in which for all  $a, b \in A$ , there is a directed edge from  $a$  to  $b$  (denoted by  $a \rightarrow b$ ) in  $M(P)$  if  $N_P(a, b) > \frac{n}{2}$ .

A resolute voting rule is a function  $f : E \rightarrow A$ . Resolute rules are typically obtained from composing an irresolute rule (mapping an election into a non-empty subset of candidates, called co-winners) with a tie-breaking mechanism.

A *positional scoring rule* (PSR)  $f^s$  is defined by a non-negative vector  $\vec{s} = (s_1, \dots, s_m)$  such that  $s_1 \geq \dots \geq s_m$  and  $s_1 > 0$ . Each candidate receives  $s_j$  points from each voter  $i$  who ranks her in the  $j^{\text{th}}$  position, and the score of a candidate is the total number of points she receives from all voters i.e.  $S(x) = \sum_{i=1}^n s_{j_i}$ . The winner is the candidate with highest total score. Examples of scoring rules are the Borda and Harmonic rules, with  $s_{\text{Borda}} = (m-1, m-2, \dots, 0)$  and  $s_{\text{Harmonic}} = (1, 1/2, \dots, 1/m)$ .

We now define three *pairwise comparison rules*.

The *Copeland* rule outputs the candidate maximizing the *Copeland score*, where the Copeland score of  $x$  is the number of candidates  $y$  with  $x \rightarrow y$  in  $M(P)$ , plus half the number of candidates  $y \neq x$  with no edge between  $x$  and  $y$  in  $M(P)$ .

The *Ranked Pairs* (RP) rule proceeds by ranking all pairs of candidates  $(x, y)$  according to  $N_P(x, y)$  (using tie-breaking when necessary); starting from an empty graph over  $A$ , it then considers all pairs in the described order and includes a pair in the graph if and only if it does not create a cycle in it. At the end of the process, the graph is a complete ranking, whose top element is the winner.

The *maximin* rule outputs the candidates that maximize  $\min_{x \in A(x \neq a)} N_P(a, x)$ .

For the experiments using randomly generated profiles, we use the *Mallows  $\phi$ -model* [18]. It is a (realistic) family of distributions over rankings, parametrized by a modal or reference ranking  $\sigma$  and a dispersion parameter  $\phi \in [0, 1]$ :  $P(r; \sigma, \phi) = \frac{1}{Z} \phi^{d(r, \sigma)}$ , where  $r$  is any ranking,  $d$  is the Kendall tau distance and  $Z = \sum_{r'} \phi^{d(r', \sigma)} = 1 \cdot (1 + \phi) \cdot (1 + \phi + \phi^2) \cdot \dots \cdot (1 + \dots + \phi^{m-1})$  is a normalization constant. With small values of  $\phi$ , the mass is concentrated around  $\sigma$ , while  $\phi = 1$  gives the uniform distribution *Impartial Culture* (IC), where all profiles are equiprobable.

### 3 Approximating Voting Rules from Truncated Ballots

Given  $k \in \{1, \dots, m-1\}$ , a *top-k* election is a triple  $E' = \langle N, A, R \rangle$  where  $N$  and  $A$  are as before, and  $R = (\succ_1^k, \dots, \succ_n^k)$ , where each  $\succ_i^k$  is a ranking of  $k$  out of  $m$  candidates in  $A$ .  $R$  is called a *top-k profile*. If  $P$  is a complete profile,  $\succ_i^k$  is the top- $k$  truncation of  $\succ_i$  (i.e., the best  $k$  candidates, ranked as in  $\succ_i$ ), and  $P_k = (\succ_1^k, \dots, \succ_n^k)$  is the top- $k$ -profile induced from  $P$  and  $k$ . A *top-k (resolute) voting rule* is a function  $f_k$  that maps each *top-k* election  $E'$  to a candidate in  $A$ . We sometimes apply a top- $k$  rule to a complete profile, with  $f_k(P) = f_k(P_k)$ . We now define several *top-k* rules.

#### 3.1 Borda and Positional Scoring Rules

**Definition 1.** A *top-k PSR*  $f_k^s$  is defined by a scoring vector  $s = (s_1, s_2, \dots, s_k, s^*)$  such that  $s_1 \geq s_2 \geq \dots \geq s_k \geq s^* \geq 0$  and  $s_1 > s^*$ . Each candidate in a top- $k$  vote receives  $s_j$  points from each voter  $i$  who ranks her in the  $j^{\text{th}}$  position. A non-ranked candidate gets  $s^*$  points. The winner is the candidate with highest total score.

When starting from a specific PSR for complete ballots, defined by scoring vector  $s = (s_1, \dots, s_m)$ , two choices of  $s^*$  particularly make sense:

- zero score:  $s^* = 0$
- average score:  $s^* = \frac{1}{m-k} (s_{k+1} + \dots + s_m)$

We denote the corresponding approximate rules as  $f_k^0$  and  $f_k^{av}$ .  $Borda_k^{av}$  is known under the name *average score modified Borda Count* [8,13], while  $Borda_k^0$  is known under the name *modified Borda Count* [11]). In the experiments we report only on  $Borda_k^{av}$ , as  $Borda_k^0$  gives very similar results.

Young [24] characterized positional scoring rules by these four properties, which we describe informally (for resolute rules):

- *Neutrality*: all candidates are treated equally
- *Anonymity*: all voters are treated equally
- *Reinforcement*: if  $P$  and  $Q$  are two profiles (on disjoint electorates) and  $x$  is the winner for  $P$  and the winner for  $Q$ , then it is also the winner for  $P \cup Q$ .
- *Continuity*: if  $P$  and  $Q$  are two profiles and  $x$  is the winner for  $P$  but not for  $Q$ , adding sufficiently many votes of  $P$  to  $Q$  leads to elect  $x$ .

$f$  is a PSR if and only if it satisfies neutrality, anonymity, reinforcement and continuity [24].

These four properties still make sense for truncated ballots. It is not difficult to generalize Young's result to *top-k* PSR:

**Theorem 1.** A *top-k* voting rule is a *top-k* PSR if and only if it satisfies neutrality, anonymity, reinforcement, and continuity.

*Proof.* The left-to-right direction is obvious. For the right-to-left direction, let us first define the *top-k-only* property: a standard voting rule is *top-k-only* if for any two complete profiles  $P, P'$ , if  $P_k = P'_k$ , then  $F(P) = F(P')$ . Then (1) a positional scoring rule

$F$  is *top-k-only* if and only if  $s_{k+1} = \dots = s_m$  (if this equality is not satisfied, then it is easy to construct two profiles  $P, P'$  such that  $P_k = P'_k$  and  $F(P) \neq F(P')$ ). Now, assume  $F_k$  is a *top-k* rule satisfying neutrality, anonymity, reinforcement, and continuity. Let  $F$  be the standard voting rule defined by  $F(P) = F_k(P_k)$ . Clearly,  $F$  also satisfies neutrality, anonymity, reinforcement, and continuity, and due to Young's characterization result,  $F$  is a PSR, associated with some vector  $(s_1, \dots, s_m)$ . Because  $F$  is also *top-k-only*, using (1) we have  $s_{k+1} = \dots = s_m$ , therefore,  $F_k$  is a *top-k-PSR*.  $\square$

### 3.2 Rules Based on Pairwise Comparisons

Given a truncated ballot  $\succ_i^k$  and two candidates  $a, b \in A$ , we say that  $a$  dominates  $b$  in  $\succ_i^k$ , denoted by  $a \succ_i^k b$ , if one of these two conditions holds: (1)  $a$  and  $b$  are listed in  $\succ_i^k$ , and  $a \succ_i^k b$ ; (2)  $a$  is listed in  $\succ_i^k$ , and  $b$  is not.

For instance, for  $A = \{a, b, c, d\}$ ,  $k = 2$ , and  $\succ_i^2 = (a \succ b)$ , then  $a$  dominates  $b$ , both  $a$  and  $b$  dominate  $c$  and  $d$ , but  $c$  and  $d$  remain incomparable in  $\succ_i^2$ . Now, the notions of pairwise comparison and majority graph are extended to *top-k* truncated profiles in a straightforward way:

**Definition 2.** Given a *top-k* profile  $R$ ,  $N_R(a, b) = \#\{i, a \succ_i^k b\}$  is the number of voters in  $R$  for whom  $a$  dominates  $b$ . The *top-k* majority graph  $M_k(R)$  induced by  $R$  is the graph whose set of vertices is the set of the candidates  $A$  and in which there is a directed edge from  $a$  to  $b$  if  $N_R(a, b) > N_R(b, a)$ .

The *top-k* rules  $Copeland_k$ ,  $Maximin_k$  and  $RP_k$  are defined exactly as their standard counterparts, but starting from the *top-k* pairwise comparisons and majority graph instead of the standard ones. Note that  $f_{m-1} = f$ , and (for all rules  $f$  we consider)  $f_1$  coincides with plurality.

**Example 1.** Let us consider this 62-voter profile: 20 votes  $a \succ d \succ c \succ b$ , 10 votes  $b \succ c \succ d \succ a$ , 15 votes  $c \succ d \succ b \succ a$  and 17 votes:  $d \succ c \succ a \succ b$ .

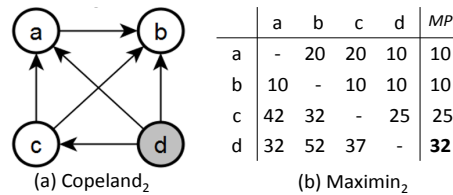


Fig. 1: top-2 approximations of Copeland and Maximin

Fig. 1 (a) shows the *top-k* majority graph and the Copeland winner for  $k = 2$ , and Fig. 1 (b) shows the *top-k* pairwise majority matrix and the Maximin<sub>k</sub> winner for  $k = 2$ . In both cases, the winner for  $k = 1$  (resp.  $k = 3$ ) is  $a$  (resp.  $d$ ). For RP, the winner under  $RP_k$  for  $k \in \{1, 2, 3\}$  is the same as the winner under Copeland<sub>k</sub> since the *k*-truncated majority graph does not create cycles.

## 4 Probability of Selecting the True Winner

The first way of measuring the quality of the *top-k* approximations is to determine the probability that they output the ‘true winner’; that is, the winner of the original voting rule, under various distributions (Subsection 4.1) and for real-world data (Subsection 4.2). In both cases, the procedure is similar: given a voting rule  $f$ , we consider many profiles, and for each profile  $P$  we compare  $f(P)$  to  $f_k(P_k)$  for each  $k = \{1, \dots, m - 2\}$ . The difference between Subsections 4.1 and 4.2 is that in the former we randomly draw profiles according to a given distribution, and for the latter, we draw a profile by selecting  $n$  votes at random in the database. We include in our experiments  $STV_k$  rule defined by Ayadi *et al.* [1], which takes *top-k* ballots as input; and we compared it to our truncated rules.  $STV_k$  proceeds as follows: in each round the candidate with the smallest number of votes is eliminated (using a tie breaking when necessary), if all ranked candidates are eliminated by STV, the vote is then ‘exhausted’ and ignored during further counting.

### 4.1 Experiments Using Mallows Model

Here we follow the research direction initiated by Filmus and Oren [12], but we consider more rules, and beyond *Impartial Culture* we also consider correlated distributions within the *Mallows* model. For each experiment we draw 1000 random preference profiles. In the first set of experiments, we take  $m = 7$ , we let  $n$  and  $\phi$  vary, and we measure the accuracy of the approximate rule for  $k = 1$  and  $k = 2$ . Results are reported on Table 1. Note that for  $k = 1$ , our results can be viewed as answering the question: *with which probability does the true winner with respect to the chosen rule coincide with the plurality winner?*

Table 1: Success rate, Mallows model:  $m = 7$ , varying  $n$ ,  $k$  and  $\phi$ .

| $\phi$                                  | n=100 | n=200 | n=300 | n=400 | n=500        | n=100                                   | n=200 | n=300 | n=400      | n=500        |
|---|-------|-------|-------|-------|--------------|---|-------|-------|------------|--------------|
| <i>Borda</i> <sub>1</sub> <sup>av</sup> |       |       |       |       |              | <i>Borda</i> <sub>2</sub> <sup>av</sup> |       |       |            |              |
| 0.7                                     | 0.902 | 0.958 | 0.986 | 0.992 | <b>1.0</b>   | 0.951                                   | 0.98  | 0.992 | <b>1.0</b> | 1.0          |
| 0.8                                     | 0.77  | 0.855 | 0.9   | 0.94  | 0.963        | 0.853                                   | 0.913 | 0.956 | 0.972      | 0.986        |
| 0.9                                     | 0.588 | 0.694 | 0.685 | 0.718 | <b>0.771</b> | 0.772                                   | 0.805 | 0.827 | 0.846      | 0.873        |
| 1                                       | 0.434 | 0.445 | 0.424 | 0.422 | <b>0.397</b> | 0.576                                   | 0.56  | 0.586 | 0.598      | <b>0.584</b> |
| <i>Copeland</i> <sub>1</sub>            |       |       |       |       |              | <i>Copeland</i> <sub>2</sub>            |       |       |            |              |
| 0.7                                     | 0.908 | 0.968 | 0.991 | 0.994 | <b>1.0</b>   | 0.947                                   | 0.99  | 1.0   | <b>1.0</b> | 1.0          |
| 0.8                                     | 0.736 | 0.847 | 0.891 | 0.934 | 0.949        | 0.822                                   | 0.904 | 0.952 | 0.984      | 0.982        |
| 0.9                                     | 0.497 | 0.567 | 0.655 | 0.684 | <b>0.726</b> | 0.62                                    | 0.69  | 0.77  | 0.805      | 0.838        |
| 1                                       | 0.325 | 0.332 | 0.323 | 0.343 | <b>0.319</b> | 0.458                                   | 0.432 | 0.45  | 0.442      | <b>0.425</b> |
| <i>Maximin</i> <sub>1</sub>             |       |       |       |       |              | <i>Maximin</i> <sub>2</sub>             |       |       |            |              |
| 0.7                                     | 0.908 | 0.969 | 0.986 | 0.99  | <b>1.0</b>   | 0.968                                   | 0.991 | 1.0   | <b>1.0</b> | 1.0          |
| 0.8                                     | 0.787 | 0.856 | 0.915 | 0.939 | 0.955        | 0.872                                   | 0.934 | 0.961 | 0.976      | 0.977        |
| 0.9                                     | 0.57  | 0.633 | 0.691 | 0.717 | <b>0.748</b> | 0.735                                   | 0.76  | 0.794 | 0.838      | 0.869        |
| 1                                       | 0.415 | 0.4   | 0.423 | 0.393 | <b>0.391</b> | 0.52                                    | 0.532 | 0.544 | 0.545      | <b>0.525</b> |
| <i>Harmonic</i> <sub>1</sub>            |       |       |       |       |              | <i>Harmonic</i> <sub>2</sub>            |       |       |            |              |
| 0.7                                     | 0.941 | 0.986 | 0.996 | 1.0   | <b>1.0</b>   | 0.98                                    | 0.992 | 1.0   | <b>1.0</b> | 1.0          |
| 0.8                                     | 0.895 | 0.916 | 0.958 | 0.959 | 0.968        | 0.958                                   | 0.974 | 0.987 | 0.988      | 0.996        |
| 0.9                                     | 0.805 | 0.808 | 0.83  | 0.866 | <b>0.863</b> | 0.895                                   | 0.921 | 0.934 | 0.939      | 0.952        |
| 1                                       | 0.725 | 0.742 | 0.74  | 0.697 | <b>0.737</b> | 0.872                                   | 0.867 | 0.859 | 0.861      | <b>0.859</b> |
| <i>RP</i> <sub>1</sub>                  |       |       |       |       |              | <i>RP</i> <sub>2</sub>                  |       |       |            |              |
| 0.7                                     | 0.926 | 0.972 | 0.995 | 0.995 | <b>1.0</b>   | 0.963                                   | 0.994 | 1.0   | <b>1.0</b> | 1.0          |
| 0.8                                     | 0.778 | 0.856 | 0.908 | 0.939 | 0.957        | 0.871                                   | 0.928 | 0.967 | 0.983      | 0.989        |
| 0.9                                     | 0.587 | 0.64  | 0.674 | 0.718 | <b>0.749</b> | 0.725                                   | 0.765 | 0.777 | 0.838      | 0.862        |
| 1                                       | 0.426 | 0.405 | 0.416 | 0.375 | <b>0.385</b> | 0.558                                   | 0.524 | 0.557 | 0.498      | <b>0.519</b> |
| <i>STV</i> <sub>1</sub>                 |       |       |       |       |              | <i>STV</i> <sub>2</sub>                 |       |       |            |              |
| 0.7                                     | 0.907 | 0.981 | 0.985 | 0.998 | <b>1.0</b>   | 0.959                                   | 0.993 | 0.997 | <b>1.0</b> | 1.0          |
| 0.8                                     | 0.808 | 0.865 | 0.917 | 0.918 | 0.943        | 0.882                                   | 0.933 | 0.962 | 0.966      | 0.974        |
| 0.9                                     | 0.603 | 0.64  | 0.721 | 0.729 | 0.763        | 0.742                                   | 0.776 | 0.792 | 0.855      | 0.846        |
| 1                                       | 0.45  | 0.464 | 0.477 | 0.471 | 0.468        | 0.576                                   | 0.593 | 0.61  | 0.592      | 0.585        |

For  $k = 1$ : when  $n \leq 100$  and  $\phi \leq 0.7$ , prediction reaches 90% for Borda, Copeland, Maximin and STV, 92% for RP, and 94% for Harmonic. When  $n \geq 500$ , the accuracy is perfect for all rules. For  $\phi = 0.8$ , the success rate decreases but results are still good with a large number of voters. For  $\phi = 0.9$  and  $n = 500$ , the rate reaches 86% for Harmonic and 72% for Copeland, with intermediate (and similar) results for Borda, Maximin and  $RP$  and STV. For the  $IC$ , the rate decreases dramatically when  $k$  becomes small, except for Harmonic (73% when  $n = 500$  against 46% for STV, 31% for Copeland and 40% for the remaining rules).

For  $k = 2$ : the probability of selecting the true winner reaches 100% (resp. 98%) when  $\phi \leq 0.7$  (resp.  $\phi \leq 0.8$ ) and  $n \geq 400$  (resp.  $n \geq 500$ ). With high values of  $\phi$ , Harmonic still outperforms other rules followed by  $Borda^{av}$  and STV then the other rules. Consistently with the results obtained by Bentert and Skowron [3] for the  $IC$ , approximating the maximin rule is harder than position scoring rules where maximin needs more information from the voters in order to obtain high approximation guarantees. In all cases, top-2 ballots seem to be always sufficient in practice to predict the winner with 100% accuracy with a low value of  $\phi$ .

In the second set of experiments, we are interested in determining the value of  $k$  needed to predict the correct winner with large elections and with high value of  $\phi$ . We take  $k = \{1, \dots, m\}$ ,  $n = 2000$ ,  $\phi = \{0.9, 1\}$  and  $m = 20$ . Fig. 2 shows depicted results where 1000 random preference profiles are generated for each experiment. Results suggest that in large elections and unless  $\phi$  is very high ( $\phi = 0.9$ ), top- $k$  rules are able to identify the true winner when  $k = 6$  (resp.  $k = 8$ ) for Harmonic (resp. the remaining rules) out of  $m = 20$ . We can also observe the behavior of different truncated rules when  $\phi = 0.9$ : the best accuracy is obtained again by Harmonic and the accuracy of all other rules are very close, which we found surprising. When  $\phi = 1$ , the latter behavior changes: Harmonic still has the best results, followed by  $Borda^{av}$  and STV, then the remaining rules. The good performance of Harmonic in all cases can be explained by the fact that the closer the scoring vector to plurality, the better the prediction.

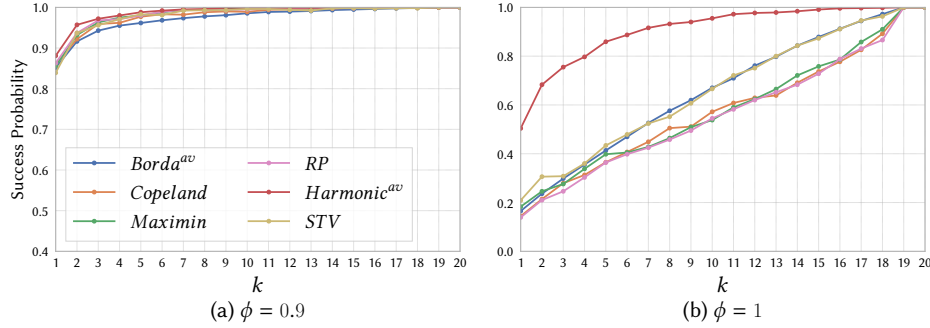


Fig. 2: Success rate, Mallows model:  $n = 2000$ ,  $m = 20$ , varying  $\phi$  and  $k$ .

Next, for each value of  $n \in \{1000, 2000\}$ ,  $\phi \in \{.7, .8, .9, 1\}$ , and  $m \in \{7, 10, 15, 20\}$ , we generated 1000 random profiles, and for each of our rules, we determined the mini-

mal value  $k$  (as a function of  $m$ ) such that the winner is correctly determined from the top- $k$  votes for all generated profiles. The results for  $Borda^{av}$  are:

- for  $\phi = 0.7$ ,  $k = 1$  is always sufficient, whatever  $m$ .
- for  $\phi = 0.8$ ,  $k = 2$  (resp.  $k = 1$ ) is always sufficient for  $n = 1000$  (resp.  $n = 2000$ ), whatever the value of  $m$ .
- for  $\phi = 0.9$ , we observe that the minimal value of  $k$  such that the correct winner is always correctly predicted is around  $\frac{7}{10}m$  (for  $n = 1000$ ) and  $\frac{2}{5}m$  (for  $n = 2000$ ).
- for  $\phi = 1$ , the minimal value of  $k$  is  $m - 1$ : we always find a generated profile for which we get an incorrect result if the profile is not complete.

The results for Copeland, maximin, RP and STV are similar to those for Borda. For Harmonic, we observe that  $k = 1$  is always sufficient for  $\phi \leq 0.8$  and  $n = 2000$ , and that for  $\phi = 0.9$  (resp.  $\phi = 1$ ), the value of  $k$  needed is around  $\frac{1}{3}m$  (resp.  $\frac{2}{3}m$ ).

In order to see how our approximations behave with small number of voters and a high dispersion parameter, we take  $k = \{1, \dots, m\}$ ,  $n = 15$ ,  $m = 7$ , and  $\phi \in \{0.9, 1\}$ . The results are on Fig. 3. The worst performance is obtained with Copeland, while the other rules perform more or less equally well. These results are consistent with the results obtained by Skowron *et al.* [23] for multiwinner rules: elections with few voters and high dispersion appear to be the worst-case scenario for predicting the correct winner using top-truncated ballots. For Harmonic, even with few voters, winner prediction is almost perfect when  $k = 4$  and  $m = 7$ .

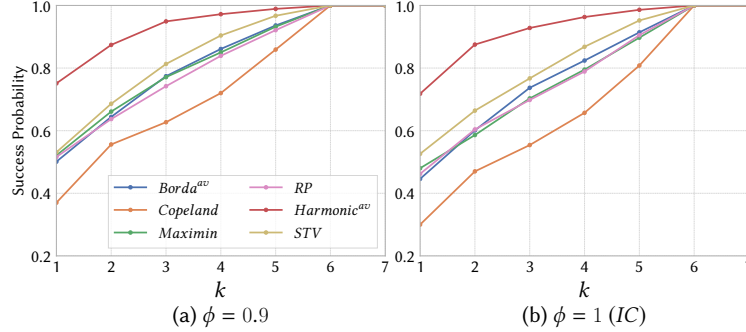


Fig. 3: Success rate, Mallows model:  $m = 7$ ,  $n = 15$ , varying  $\phi$  and  $k$ .

## 4.2 Experiments Using Real Data Sets

We now consider real data set from *Preflib* [19]: 2002 election for Dublin North constituency with 12 candidates and 3662 voters. We consider data with samples of  $n^*$  voters among  $n$  ( $n^* < n$ ), starting by  $n^* = 10$  and increment  $n^*$  in steps of 10. In each experiment, 1000 random profiles are constructed with  $n^*$  voters; then we consider the top- $k$  ballots obtained from these profiles, with  $k = \{1, 2, 3\}$ , and we compute the frequency with which we select the true winner. Fig. 4 shows results for Dublin with small elections ( $n^* = \{10, \dots, 100\}$ ) while Fig. 5 presents results for large elections ( $n^* = \{100, \dots, 2000\}$ ). Arrows indicate the number of voters from which the prediction is perfect.



Consistently with the results of Fig. 3, for small elections; the success rate is low when  $k$  is too small, except for Harmonic where it gives the best performance followed by STV (especially when  $n^* < 60$ ) then the remaining rules, e.g. For Harmonic (resp. STV), 92% (resp. 82%) accuracy is reached with  $k = 3$ ,  $m = 12$  and  $n^* = 50$  against around 75% for the remaining rules.

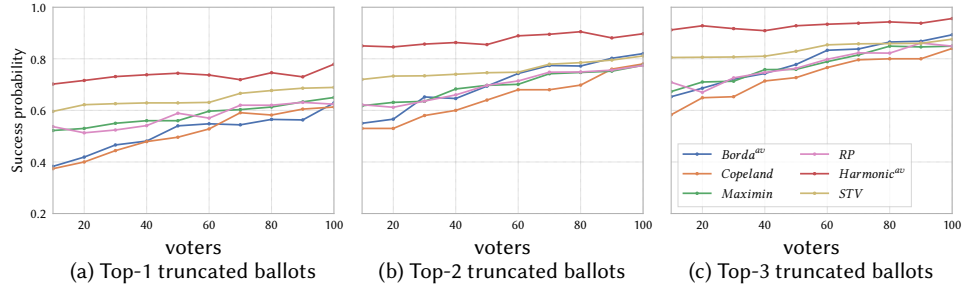


Fig. 4: Success rate, Dublin, varying  $k$ ;  $n^* = \{10, \dots, 100\}$ .

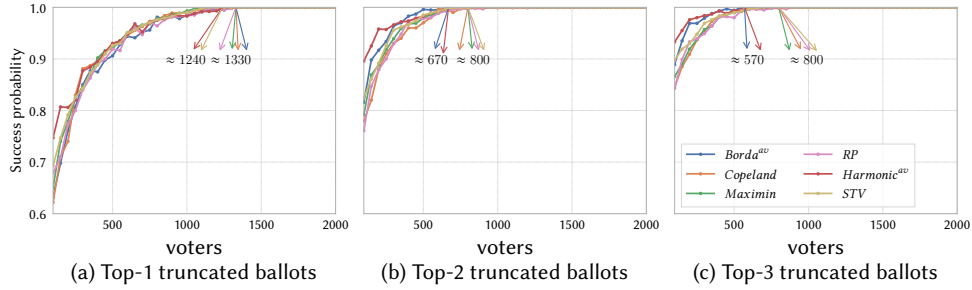


Fig. 5: Success rate, Dublin, varying  $k$ ;  $n^* = \{100, \dots, 2000\}$ .

For large elections, when  $k = 1$ , the different approximations exhibit almost the same behavior except Harmonic, that performs better especially with few voters. Obviously, increasing the value of  $k$  leads to a decrease in the number of voters needed for correct winner selection. In general, the different approximations need a sufficient number of voters to converge to the correct prediction. Scoring rules tend to require less voters.

## 5 Measuring the Approximation Ratio

### 5.1 Worst Case Study

In order to measure the quality of approximate voting rules whose definition is based on score maximization, a classical method consists in computing the worst-case approximation ratio between the scores (for the original rule) of the ‘true’ winner and of the winner of the approximate rule. Using worst-case score ratios is classical: they are defined for measuring the quality of approximate voting rules [7,22], for defining the price of anarchy of a voting rule [6] or for measuring the distortion of a voting rule [4].

Worst-case score ratios particularly make sense if the score of a candidate is meaningful beyond its use for determining the winner. This is definitely the case for Borda, as the Borda count is often seen as a measure of social welfare (see [9]). This worst-case score ratio is called the *price of top- $k$  truncation*.

**Definition 3.** Let  $f$  be a voting rule defined as the maximization of a score  $S$ , and  $f_k$  a top- $k$  approximation of  $f$ . The price of top- $k$ -truncation for  $f$ ,  $f_k$ ,  $m$ , and  $k$ , is defined as:  $R(f, f_k, m, k) = \max_{P \in \mathcal{P}_m} \frac{S(f(P))}{S(f_k(P_k))}$ .

**Positional Scoring Rules:** Let  $f^s$  be a positional scoring rule defined with scoring vector  $s$ . Assume the tie-breaking priority favors  $x_1$ . Let  $f_k^{\bar{s}}$  be a top- $k$  approximation of  $f^s$ , associated with vector  $\bar{s} = (s_1, \dots, s_k, s^*)$ , with the same tie-breaking priority. Let  $s' = (s_1 - s^*, \dots, s_k - s^*, 0) = (s'_1, \dots, s'_k, 0)$ , i.e.,  $s'_i = s_i - s^*$  for  $i = 1, \dots, k$ . Obviously,  $f_k^{\bar{s}} = f_k^{s'}$ . For instance, if  $f^{\bar{s}}$  is the average-score approximation of the Borda rule, then  $\bar{s} = (m-1, \dots, m-k, \frac{m-k-1}{2})$  and  $s' = (m-1 - \frac{m-k-1}{2}, \dots, m-k - \frac{m-k-1}{2}, 0)$ .

Let  $S(x, P)$  be the score of  $x$  for  $P$  under  $f^s$  and  $S'_k(x, P_k)$  be the score of  $x$  for  $P_k$  under  $f_k^{s'}$ . From now on when we write scores we omit  $P$  and  $P_k$ , i.e., we write  $S(x)$  instead of  $S(x, P)$ ,  $S'_k(x)$  instead of  $S'_k(x, P_k)$  etc. In the rest of Subsection 5.1 we assume  $k \geq 2$ . Let  $x_1 = f_k^{s'}(P_k)$  and  $x_2 = f^s(P)$ .

**Lemma 1.**  $R(f^s, f_k^{s'}, m, k) \leq 1 - \frac{s_{k+1}}{s'_1} + \left(1 + \frac{s^*}{s'_1}\right) \frac{ms_{k+1}}{s'_1 + \dots + s'_k}$

*Proof.* The total number of points given to candidates under  $f_k^{s'}$  is  $n(s'_1 + \dots + s'_k)$ , therefore  $S'_k(x_1) \geq \frac{n}{m}(s'_1 + \dots + s'_k)$ .

Let us write  $S(x_2) = S_{1 \rightarrow k}(x_2) + S_{k+1 \rightarrow m}(x_2)$ , where  $S_{1 \rightarrow k}(x_2)$  (resp.  $S_{k+1 \rightarrow m}(x_2)$ ) is the number of points that  $x_2$  gets from the top  $k$  (resp. bottom  $m-k$ ) positions of the ballots in  $P$ . Let  $\gamma$  be the number of ballots in which  $x_2$  is not in the top  $k$  positions. Then  $S_{k+1 \rightarrow m}(x_2) \leq \gamma s_{k+1}$ .

As  $x_2$  appears in at least  $\frac{S'_k(x_2)}{s'_1}$  top- $k$  ballots, we have  $\gamma \leq n - \frac{S'_k(x_2)}{s'_1}$ . Moreover we have  $S(x_1) \geq S_{1 \rightarrow k}(x_1) = S'_k(x_1) + ns^* \geq S'_k(x_2) + ns^* = S_{1 \rightarrow k}(x_2)$ . Now,

$$\begin{aligned} S(x_2) &\leq S_{1 \rightarrow k}(x_2) + \left(n - \frac{S'_k(x_2)}{s'_1}\right) s_{k+1} \\ &\leq S_{1 \rightarrow k}(x_2) + \left(n - \frac{S_k(x_2) - ns^*}{s'_1}\right) s_{k+1} \\ &\leq \left(1 - \frac{s_{k+1}}{s'_1}\right) S_{1 \rightarrow k}(x_2) + ns_{k+1} + \frac{ns^* s_{k+1}}{s'_1} \\ &\leq \left(1 - \frac{s_{k+1}}{s'_1}\right) S(x_1) + ns_{k+1} + \frac{ns^* s_{k+1}}{s'_1} \\ \frac{S(x_2)}{S(x_1)} &\leq 1 - \frac{s_{k+1}}{s'_1} + ns_{k+1} \left(1 + \frac{s^*}{s'_1}\right) \frac{m}{n(s'_1 + \dots + s'_k)} \\ &\leq 1 - \frac{s_{k+1}}{s'_1} + s_{k+1} \left(1 + \frac{s^*}{s'_1}\right) \frac{m}{s'_1 + \dots + s'_k} \quad \square \end{aligned}$$

We now focus on the lower bound. We build the following pathological complete profile  $P$  such that:

- the winner for  $P_k$  (resp.  $P$ ) is  $x_1$  (resp.  $x_2$ ).

- in  $P_k$ , all candidates get the same number of points ( $x_1$  wins thanks to tie-breaking), and  $x_1$  and  $x_2$  get all their points from top-1 positions.
- in  $P$ , the score of  $x_1$  is minimized by ranking it last everywhere where it was not in the top  $k$  positions, and the score of  $x_2$  is maximized by ranking it in position  $k + 1$  everywhere where it was not in the top  $k$  positions.
- $P_k$  is symmetric in  $\{x_3, \dots, x_m\}$ .

Formally,  $P_k$  is defined as follows:

1. for each ranked list  $L$  (resp.  $L'$ ) of  $k - 1$  (resp.  $k$ ) candidates in  $\{x_3, \dots, x_m\}$ :  $\alpha$  votes  $x_1 L$  and  $\alpha$  votes  $x_2 L$  (resp.  $\beta$  votes  $L'$ ).  $\alpha$  and  $\beta$  will be fixed later.
2.  $\alpha$  and  $\beta$  are chosen in such a way that all candidates get the same score  $S'_k(\cdot)$ .

Now,  $P$  is obtained by completing  $P_k$  as follows:

1. each top- $k$  vote  $x_1 L$  is completed into  $x_1 L x_2 -$ . “ $-$ ” means the remaining candidates are in an arbitrary order.
2. each top- $k$  vote  $x_2 L$  is completed into  $x_2 L - x_1$ .
3. each top- $k$  vote  $L'$  is completed into  $L' x_2 - x_1$ .

For instance, for  $m = 5$  and  $k = 3$ ,  $P$  is as follows:

|                                |                                |                               |
|--------------------------------|--------------------------------|-------------------------------|
| $\alpha$ $x_1 x_3 x_4 x_2 x_5$ | $\alpha$ $x_2 x_3 x_4 x_5 x_1$ | $\beta$ $x_3 x_4 x_5 x_2 x_1$ |
| $\alpha$ $x_1 x_3 x_5 x_2 x_4$ | $\alpha$ $x_2 x_3 x_5 x_4 x_1$ | $\beta$ $x_3 x_5 x_4 x_2 x_1$ |
| $\alpha$ $x_1 x_4 x_3 x_2 x_5$ | $\alpha$ $x_2 x_4 x_3 x_5 x_1$ | $\beta$ $x_4 x_3 x_5 x_2 x_1$ |
| $\alpha$ $x_1 x_4 x_5 x_2 x_3$ | $\alpha$ $x_2 x_4 x_5 x_3 x_1$ | $\beta$ $x_4 x_5 x_3 x_2 x_1$ |
| $\alpha$ $x_1 x_5 x_3 x_2 x_4$ | $\alpha$ $x_2 x_5 x_3 x_4 x_1$ | $\beta$ $x_5 x_3 x_4 x_2 x_1$ |
| $\alpha$ $x_1 x_5 x_4 x_2 x_3$ | $\alpha$ $x_2 x_5 x_4 x_3 x_1$ | $\beta$ $x_5 x_4 x_3 x_2 x_1$ |

$$\text{Let } M = \frac{(m-3)!}{(m-k-1)!} \text{ and } Q = \frac{(m-2)!}{(m-k-1)!}.$$

**Lemma 2.**

$$S'_k(x_1) = S'_k(x_2) = \alpha(m-2)s'_1 M$$

$$\text{and for } i \geq 3, S'_k(x_i) = 2\alpha(s'_2 + \dots + s'_k)M + \beta(m-k-1)(s'_1 + \dots + s'_k)M$$

*Proof.* In  $P_k$ ,  $x_1$  and  $x_2$  appear in top position in a number of votes equal to  $\alpha$  times the number of different permutations (ordered lists) of  $(k-1)$  candidates out of  $(m-2)$ , i.e.  $\alpha \frac{(m-2)!}{(m-k-1)!}$  times. Thus  $S'_k(x_1) = S'_k(x_2) = \alpha \frac{(m-2)!}{(m-k-1)!} s'_1$ . For similar reasons, for each  $i \geq 3$ ,

$$S'_k(x_i) = 2\alpha \frac{(m-3)!}{(m-k-1)!} (s'_2 + \dots + s'_k) + \beta \frac{(m-3)!}{(m-k-2)!} (s'_1 + \dots + s'_k). \quad \square$$

As a consequence, all candidates have the same score in  $P_k$  if and only if

$$\frac{\beta}{\alpha} = \frac{(m-2)s'_1 - 2(s'_2 + \dots + s'_k)}{(m-k-1)(s'_1 + \dots + s'_k)}$$

We fix  $\alpha$  and  $\beta$  such that this equality holds. Thanks to the tie-breaking priority, the winner in  $P_k$  is  $x_1$ . In  $P$ , the winner is  $x_2$  and the scores of  $x_1$  and  $x_2$  are as follows:

**Lemma 3.**

$$\begin{aligned} S(x_1) &= Q\alpha s_1 \\ S(x_2) &= Q\alpha s_1 + Q\alpha s_{k+1} + Q(m-k-1)\beta s_{k+1} \end{aligned}$$

*Proof.*  $x_1$  appears at the top of  $\frac{(m-2)!}{(m-k-1)!}\alpha$  votes and at the bottom of all others, hence  $S(x_1) = Q\alpha s_1$ .  $x_2$  appears  $\alpha \frac{(m-2)!}{(m-k-1)!}$  times top position, and in position  $(k+1)$  in the remaining votes, i.e.,  $\alpha \frac{(m-2)!}{(m-k-1)!} + \beta \frac{(m-2)!}{(m-k-2)!}$ . Thus

$$S(x_2) = \alpha \frac{(m-2)!}{(m-k-1)!} (s_1 + s_{k+1}) + \beta \frac{(m-2)!}{(m-k-2)!} s_{k+1} \quad \square$$

**Lemma 4.**  $R(f^s, f_k^{s'}, m, k) \geq 1 - \frac{s_{k+1}}{s_1} + \frac{s_{k+1}}{s_1} \frac{ms'_1}{s'_1 + \dots + s'_k}$ 

*Proof.* From Lemma 3 we get  $\frac{S(x_2)}{S(x_1)} \geq 1 + \frac{s_{k+1}}{s_1} + (m-k-1) \frac{s_{k+1}}{s_1} \frac{\beta}{\alpha}$ .

Finally, using the expression of  $\frac{\beta}{\alpha}$  we get

$$\frac{S(x_2)}{S(x_1)} \geq 1 + \frac{s_{k+1}}{s_1} + (m-k-1) \frac{s_{k+1}}{s_1} \frac{(m-2)s'_1 - 2(s'_2 + \dots + s'_k)}{(m-k-1)(s'_1 + \dots + s'_k)}$$

From this we conclude:

$$\begin{aligned} R(f^s, f_k^{s'}, m, k) &\geq 1 + \frac{s_{k+1}}{s_1} + \frac{s_{k+1}}{s_1} \frac{(m-2)s'_1 - 2(s'_2 + \dots + s'_k)}{s'_1 + \dots + s'_k} \\ &\geq 1 + \frac{s_{k+1}}{s_1} + \frac{s_{k+1}}{s_1} \frac{(m-2)s'_1 + 2s'_1 - 2(s'_2 + \dots + s'_k)}{s'_1 + \dots + s'_k} \\ &\geq 1 + \frac{s_{k+1}}{s_1} + \frac{s_{k+1}}{s_1} \left( \frac{ms'_1}{s'_1 + \dots + s'_k} - 2 \right) \\ &\geq 1 - \frac{s_{k+1}}{s_1} + \frac{s_{k+1}}{s_1} \frac{ms'_1}{s'_1 + \dots + s'_k} \quad \square \end{aligned}$$

Putting Lemmas 1 and 4 together we get

**Proposition 1.**

$$1 - \frac{s_{k+1}}{s_1} + \frac{s_{k+1}}{s_1} \frac{ms'_1}{s'_1 + \dots + s'_k} \leq R(f^s, f_k^{s'}, m, k) \leq 1 - \frac{s_{k+1}}{s_1} + \left(1 + \frac{s^*}{s'_1}\right) \frac{ms_{k+1}}{s'_1 + \dots + s'_k}$$

Note that the lower and upper bound coincide when  $s^* = 0$ , giving a tight worst-case approximation ratio for this class of approximations. This is however not guaranteed when  $s^* > 0$  (the reason being that the pathological profile used in the proof of Lemma 1 may not be the worst). Moreover, when  $s^* = 0$ , our (lower and upper) bound coincides with the optimal ratio given in [3] (Theorem 1).<sup>3</sup> Since the ratio in [3] is shown to be the best possible ratio, this show that taking  $s^* = 0$  gives an optimal top- $k$  approximation of a positional scoring rule.<sup>4</sup>

In particular:

<sup>3</sup> Note that the ratios in our paper are the inverse of the ratios in [3]. That is, the inverse of the ratio given in Theorem 1 of [3] coincides with our ratio for  $s^* = 0$ .

<sup>4</sup> Interestingly, [3] give another optimal rule (thus with same worst-case ratio), which is much more complex, and which is not a top- $k$  PSR. Comparing the average ratio of both rules is left for further study.

- for  $Borda_k^0$  ( $s_i = m - i, s^* = 0$ ), the lower and upper bounds coincide and are equal to  $\frac{k}{m-1} + \frac{2m(m-k-1)}{k(2m-k-1)}$ .
- for  $Borda_k^{av}$  ( $s_i = m - i, s^* = m-k-1/2$ ), the lower bound is  $1 - \frac{m-k-1}{m-1} + \frac{(m-k-1)(m+k-1)}{k(m-1)}$  and the upper bound is  $\frac{k(3k-m+1)+4(m-k-1)(m-1)}{k(m+k-1)}$ .
- for  $Harmonic_k^0$  ( $s_i = 1/i, s^* = 0$ ), the lower and upper bounds are equal to  $\frac{k}{k+1} + \frac{m}{(k+1)(1+\frac{1}{2}+\dots+\frac{1}{k})}$ .

Also, note that for  $k'$ -approval with  $k' > k$  and  $s^* = 0$ , the (exact) worst-case ratio  $\frac{m}{k}$  does not depend on  $k'$ . As a corollary, we get the following order of magnitudes when  $m$  grows:

- $R(Borda, Borda_k^0, m, k) = \Theta\left(\frac{m}{k}\right)$ .
- $R(Borda, Borda_k^{av}, m, k) = \Theta\left(\frac{m}{k}\right)$ .
- $R(Harmonic, Harmonic_k^0, m, k) = \Theta\left(\frac{m}{k \log k}\right)$ .

**Maximin:** Let *Maximin* be the Maximin rule with tie-breaking priority  $x_1 \dots x_m$ , and *Maximin<sub>k</sub>* be the  $k$ -truncated version of the Maximin rule with the same tie-breaking priority order. Let  $S_{Mm}(x_2, P)$  and  $S_{Mm}(x_1, P_k)$  be the Maximin scores of  $x_2$  and  $x_1$  for  $P$  and  $P_k$ , respectively, with  $S_{Mm}(x_2, P) = \min_{y \neq x_2} N_P(x_2, y)$  and similarly for  $P_k$ . Let  $P$  be a profile, and let  $x_1 = \text{Maximin}_k(P_k)$  and  $x_2 = \text{Maximin}(P)$ . All candidates have the same Maximin score in  $P_k$ , therefore, by tie-breaking priority,  $\text{Maximin}_k(P_k) = x_1$ .

**Lemma 5.**  $R(\text{Maximin}, \text{Maximin}_k, m, k) \leq m - k + 1$ .

*Proof.* Because  $x_1 = \text{Maximin}_k(P_k)$ , we must have  $S_{Mm}(x_1, P_k) \geq 1$  (otherwise we would have  $S_{Mm}(x_1, P_k) \geq 0$ , meaning that  $x_1$  does not belong to any top- $k$  ballot, and in this case we cannot have  $x_1 = \text{Maximin}_k(P_k)$ ). Now,  $S_{Mm}(x_2, P) \leq S_{Mm}(x_2, P_k) + (m - k) \leq S_{Mm}(x_1, P_k) + (m - k)$ , therefore,

$$\begin{aligned} \frac{S_{Mm}(x_2, P)}{S_{Mm}(x_1, P)} &\leq \frac{S_{Mm}(x_1, P_k) + (m - k)}{S_{Mm}(x_1, P_k)} \\ &\leq m - k + 1 \quad \square \end{aligned}$$

**Lemma 6.**  $R(\text{Maximin}, \text{Maximin}_k, m, k) \geq m - k$ .

*Proof.* We consider the cyclic profile *Cyc*:

| <i>Cyc</i> |         |         |         |         |       | $P (m = 5, k = 2)$ |       |       |       |       |
|------------|---------|---------|---------|---------|-------|--------------------|-------|-------|-------|-------|
| $x_1$      | $x_2$   | $\dots$ | $m - 1$ | $m$     |       | $x_1$              | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
| $x_2$      | $x_3$   | $\dots$ | $m$     |         | $x_1$ | $x_2$              | $x_3$ | $x_4$ | $x_5$ | $x_1$ |
| $x_3$      | $x_4$   | $\dots$ | $x_1$   |         | $x_2$ | $x_3$              | $x_4$ | $x_2$ | $x_5$ | $x_1$ |
| $\dots$    | $\dots$ | $\dots$ | $\dots$ |         |       | $x_4$              | $x_5$ | $x_2$ | $x_3$ | $x_1$ |
| $m$        | $x_1$   | $\dots$ | $m - 2$ | $m - 1$ |       | $x_5$              | $x_1$ | $x_2$ | $x_3$ | $x_4$ |

Now, let  $P$  be obtained from *Cyc* by the following operations for every vote in *Cyc*:

- if  $x_1$  is not in the top  $k$  positions in the vote, we move it to the last position (and move all candidates who were below  $x_1$  one position upward)
- if  $x_2$  is not in the top  $k$  positions in the vote, we move it to the  $(k + 1)^{th}$  position (and move all candidates who were between position  $k + 1$  and  $x_2$ 's position one position downward).

For instance, for  $m = 5, k = 2$ , we get the profile  $P$  above.

$Maximin(P) = x_2$ , and the Maximin scores of  $x_1$  and  $x_2$  in  $P$  are:

$$S_{Mm}(x_1, P) = 1 \text{ and } S_{Mm}(x_2, P) = m - k.$$

Hence  $\frac{S_{Mm}(x_2, P)}{S_{Mm}(x_1, P)} = m - k$ . □

**Proposition 2.**  $m - k \leq R(Maximin, Maximin_k, m, k) \leq m - k + 1$ .

This worst-case ratio is quite bad, except if  $k$  is close to  $m$ . However, arguably, the maximin score makes less sense *per se* (i.e., as a measure of social welfare) than a positional score such as the Borda count.

**Copeland:** Again, for the Copeland rule, the ratio makes less sense, because the Copeland score is less meaningful as a measure of social welfare.<sup>5</sup> Still, for the sake of completeness we give the following result:

**Proposition 3.**  $R(Copeland, Copeland_k, m, k) = \infty$ .

*Proof.* Let  $P$  be the following profile:

- $P_k$  contains two votes  $x_1 x_2 \dots x_k$ , and one vote  $L$  for each ordered list of  $k$  candidates among  $m$ .
- $P$  is obtained by completing  $P_k$  by adding  $x_1$  (resp.  $x_2$ ) in last position (resp. in position  $k + 1$ ) when it is not in the *top-k* positions.

In  $P_k$ , the winner for  $Copeland_k$  is  $x_1$ . In  $P$ , the Copeland winner is  $x_2$ . Now, with respect to  $P$ , the Copeland score of  $x_1$  (resp.  $x_2$ ) is 0 (resp.  $m - 1$ ), hence the result. □

**Discussion:** The obtained worst-case bounds are rather negative: very negative for Copeland and maximin, less so for Borda, and even less so for Harmonic.<sup>6</sup> However, the maximin and Copeland scores make less sense as a measure of social welfare than positional scores. Note that for maximin rule the obtained lower bound matches the one given by Bentert and Skowron [3] (Section 4.3) which means that our top-k approximation of maximin is optimal.

Now, we may wonder whether these worst cases do occur frequently in practice or if they correspond to rare pathological profiles. The next two subsections show that the latter is the case.

<sup>5</sup> Moreover, there are several ways of defining the Copeland score, all leading to the same rule. However, this has no impact on the negative result below, as long as a Condorcet loser has score 0.

<sup>6</sup> As Ranked Pairs is not based on scores, it was not studied here.

## 5.2 Average Case Evaluation

We present the evaluation of the approximation ratio using data generated from Mallows  $\phi$  model. For each experiment, we draw 10000 random profiles, with  $m = 7$ ,  $n = 15$ , and let  $\phi$  vary.

Fig. 6 shows results reflecting the approximation ratio for truncated rules when using Mallows model. Our results suggest that, in practice, results are much better than in the worst case where best results are obtained by Harmonic, followed by Borda and finally Maximin.

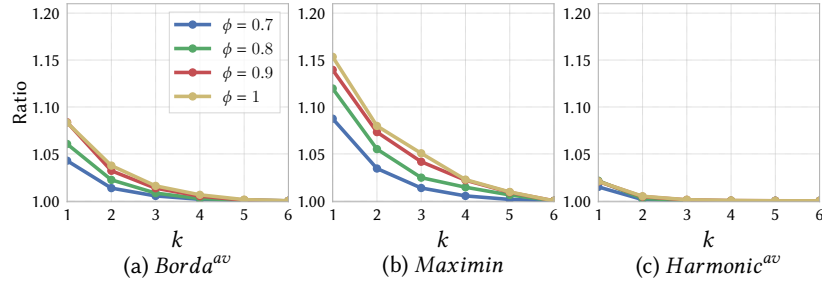


Fig. 6: Mallows model: approximation ratio when  $n = 15$ ,  $m = 7$  and varying  $\phi$ .

## 5.3 Real Data Sets

Again we consider 2002 Dublin North data ( $m = 12$ ,  $n = 3662$ ) with samples of  $n^*$  voters among  $n$  ( $n^* < n$ ) where  $n^* = \{15, 100\}$ . In each experiment 1000 random profiles are constructed with  $n^*$  voters; then we consider the top- $k$  ballots obtained from these profiles with  $k = \{1, \dots, m - 1\}$ . Again, the results are very positive.

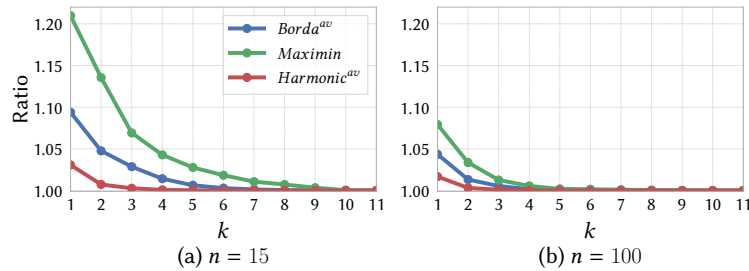


Fig. 7: Approximation ratio with Dublin North data set.

## 6 Conclusion

In this paper we have considered  $k$ -truncated approximations of rules which take only *top- $k$*  ballots as input where we have considered two measures of the quality of the approximation: the probability of selecting the same winner as the original rule, and the score ratio. For the former, our empirical study show that a very small  $k$  suffices. For the latter, while the theoretical bounds are, at best; moderately encouraging, our experiments show that in practice the approximation ratio is much better than in the worst case: our results suggest that a very small value of  $k$  works very well in practice. Many issues remain open. Especially, it would be interesting to consider *top- $k$*  approximations as voting rules on their own, and to study their normative properties.

## Acknowledgements

This work was supported by Agence Nationale de la Recherche under the “Programme d’Investissements d’Avenir” ANR-19-P3IA-0001 (PRAIRIE).

## References

1. Ayadi, M., Ben Amor, N., Lang, J., Peters, D.: Single transferable vote: Incomplete knowledge and communication issues. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. pp. 1288–1296 (2019)
2. Baumeister, D., Faliszewski, P., Lang, J., Rothe, J.: Campaigns for lazy voters: Truncated ballots. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2. pp. 577–584 (2012)
3. Bentert, M., Skowron, P.: Comparing election methods where each voter ranks only few candidates. arXiv preprint arXiv:1901.10848 (2019)
4. Boutilier, C., Caragiannis, I., Haber, S., Lu, T., Procaccia, A.D., Sheffet, O.: Optimal social choice functions: A utilitarian view. *Artif. Intell.* **227**, 190–213 (2015)
5. Boutilier, C., Rosenschein, J.S.: Incomplete information and communication in voting. In: Handbook of Computational Social Choice, pp. 223–258 (2016)
6. Brânzei, S., Caragiannis, I., Morgenstern, J., Procaccia, A.D.: How bad is selfish voting? In: AAAI. vol. 13, pp. 138–144 (2013)
7. Caragiannis, I., Kaklamanis, C., Karanikolas, N., Procaccia, A.D.: Socially desirable approximations for dodgson’s voting rule. *ACM Transactions on Algorithms (TALG)* **10**(2), 6 (2014)
8. Cullinan, J., Hsiao, S.K., Polett, D.: A Borda count for partially ordered ballots. *Social Choice and Welfare* **42**(4), 913–926 (2014)
9. d’Aspremont, C., Gevers, L.: Chapter 10 social welfare functionals and interpersonal comparability. In: Handbook of Social Choice and Welfare, Handbook of Social Choice and Welfare, vol. 1, pp. 459 – 541. Elsevier (2002)
10. Dery, L.N., Kalech, M., Rokach, L., Shapira, B.: Reaching a joint decision with minimal elicitation of voter preferences. vol. 278, pp. 466–487. Elsevier (2014)
11. Emerson, P.J.: The politics of consensus: For the resolution of conflict and reform of majority rule (1994)
12. Filmus, Y., Oren, J.: Efficient voting via the top- $k$  elicitation scheme: a probabilistic approach. In: Proceedings of the fifteenth ACM conference on Economics and computation. pp. 295–312. ACM (2014)



13. Grandi, U., Loreggia, A., Rossi, F., Saraswat, V.: A Borda count for collective sentiment analysis. *Annals of Mathematics and Artificial Intelligence* **77**(3-4), 281–302 (Aug 2016)
14. Kalech, M., Kraus, S., Kaminka, G.A., Goldman, C.V.: Practical voting rules with partial information. vol. 22, pp. 151–182. Springer (2011)
15. Lackner, M.: Incomplete preferences in single-peaked electorates. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. pp. 742–748 (2014)
16. Lu, T., Boutilier, C.: Robust approximation and incremental elicitation in voting protocols. In: *Proceedings of IJCAI International Joint Conference on Artificial Intelligence*. vol. 22, pp. 287–293 (2011)
17. Lu, T., Boutilier, C.: Vote elicitation with probabilistic preference models: Empirical estimation and cost tradeoffs. In: *Algorithmic Decision Theory*. pp. 135–149. Springer (2011)
18. Mallows, C.L.: Non-null ranking models. i. *Biometrika* pp. 114–130 (1957)
19. Mattei, N., Walsh, T.: Preflib: A library for preferences <http://www.preflib.org>. In: *Algorithmic Decision Theory*. pp. 259–270. Springer (2013)
20. Narodytska, N., Walsh, T.: The computational impact of partial votes on strategic voting. In: *ECAI 2014 - 21st European Conference on Artificial Intelligence*. pp. 657–662 (2014)
21. Oren, J., Filmus, Y., Boutilier, C.: Efficient vote elicitation under candidate uncertainty. In: *Proceedings of the Twenty-Third international Joint Conference on Artificial Intelligence*. pp. 309–316. AAAI Press (2013)
22. Service, T.C., Adams, J.A.: Communication complexity of approximating voting rules. In: *International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012*. pp. 593–602 (2012)
23. Skowron, P., Faliszewski, P., Slinko, A.: Achieving fully proportional representation: Approximability results. *Artificial Intelligence* **222**, 67–103 (2015)
24. Young, H.P.: Social choice scoring functions. *SIAM Journal on Applied Mathematics* **28**(4), 824–838 (1975)
25. Zhao, Z., Li, H., Wang, J., Kephart, J.O., Mattei, N., Su, H., Xia, L.: A cost-effective framework for preference elicitation and aggregation. In: *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018*. pp. 446–456 (2018)