A Quantitative and Qualitative Analysis of the Robustness of (Real-World) Election Winners

Niclas Boehmer
Technische Universität Berlin
Berlin, Germany
niclas.boehmer@tu-berlin.de

Robert Bredereck
TU Clausthal
Clausthal-Zellerfeld, Germany
robert.bredereck@tu-clausthal.de

Piotr Faliszewski
AGH University
Kraków, Poland
faliszew@agh.edu.pl

Rolf Niedermeier
Technische Universität Berlin
Berlin, Germany

ABSTRACT

Contributing to the toolbox for interpreting election results, we evaluate the robustness of election winners to random noise. We compare the robustness of different voting rules and evaluate the robustness of real-world election winners from the Formula 1 World Championship and some variant of political elections. We find many instances of elections that have very non-robust winners and numerous delicate robustness patterns that cannot be identified using classical and simpler approaches.

ACM Reference Format:
Niclas Boehmer, Robert Bredereck, Piotr Faliszewski, and Rolf Niedermeier. 2022. A Quantitative and Qualitative Analysis of the Robustness of (Real-World) Election Winners. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), October 6–9, 2022, Arlington, VA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3551624.3555292

1 INTRODUCTION

Voting is a convenient and powerful framework to aggregate preferences. It has many real-world applications ranging from political elections, through evaluation panels deciding on which research projects to fund, and televoting in TV shows, to aggregating the results of sport competitions. Interestingly, independent of the application, one regularly, and emotionally, debated matter is by "how much" the winning candidate had won the election. Remarkably, studies have found that there are more extremely close elections than one might intuitively expect. For instance, there is a list of 313 political elections on Wikipedia where the election was decided by less than 0.1% of all voters [18]. Moreover, Mulligan and Hunter [23] reported that in state elections in the United States one in every 15.000 voters casts a decisive vote. Motivated by this, there is a rich body of theoretical literature on the likelihood that elections are decided by a single vote [1, 3, 11, 16, 17, 21, 22, 33].

EAAMO '22, October 6–9, 2022, Arlington, VA, USA

ACM ISBN 978-1-4503-9477-2/22/10…$15.00
https://doi.org/10.1145/3551624.3555292

But what is the relevance of close elections beyond being a topic people like to argue about? The main underlying assumption here is that the recorded votes in the election capture reality only approximately. For instance, it might be the case that some voters cast their votes in a rush or without having enough information available to them, voters were unable to participate in the election, or votes were incorrectly recorded due to technical errors (which indeed happen sometimes [24, 31]). If an election is detected to be close, various counter-measures can be taken: For instance, it is possible to do a recount or to audit the election results [25, 27–29], to continue discussions about the election issue, or to collect further votes. Even if one assumes that the election result is "correct", by how much a candidate won also influences its legitimacy and credibility, in particular considering that voters might change their mind over time. To sum up, a reliable estimate for the lead of an election winner has the potential to increase the fairness and transparency of elections, as it allows for a better interpretation of election results and the initiation of possible countermeasures.

But what does it mean for an election to be close? In political elections, Plurality voting is often used. Here, each voter awards one point to its most preferred candidate and the candidate with the most points wins. For Plurality (and also for arbitrary scoring-based rules) a natural and common measure to assess the closeness of an election is the difference between the score of the election winner and the candidate finishing in the second place. A more fine grained version of this notion is the *margin of victory*, which is defined as the minimum number of voters that need to change their votes to change the election outcome [10, 13, 19, 32]. However, both of these concepts are too "coarse" in certain situations. To illustrate this, consider an election E containing 50 times the vote $a \succ b \succ \dots$ and 49 times the vote $b \succ \dots \succ a$, and an election E' containing 50 times the vote $a \succ \dots \succ b$ and 49 times the vote $b \succ a \succ \dots$ (we write $a \succ b$ to indicate that a is preferred to b and "…" to indicate that we rank all remaining candidates in some arbitrary ordering). While in both elections the score difference and margin of victory under Plurality is one, examining the votes more closely, the situation in these two elections is quite different: In E, in order for b to win the election, only one of 50 voters needs to slightly change its mind (by swapping its two most preferred candidates). In contrast, in E', in order for b to win the election, at least one of 50 voters needs to drastically change its mind by ranking its previously last-ranked candidate in the first position (plus there

are 49 voters where a can easily gain a point). To sum up, both the score difference and the margin of victory do not take into account that small changes are more likely than large ones. Further, both measures suffer from the drawback that they focus on the worst case (e.g., for the margin of victory it does not make a difference whether there is only one specific voter that can change the election outcome by modifying its vote or whether multiple voters have this power) and not on the average case, which is probably the practically more relevant one.

In this paper, following the works of Boehmer et al. [4] and Baumeister and Hogrebe [2], we study a more fine-grained robustness measure: We analyze how the winning probabilities of candidates behave if we start to perturb the election by performing some swaps of adjacent candidates in some votes. How quickly the winning probability of the original election winner decreases as we move further and further away from the original election sheds some light on this winner's robustness. Herein, we assume that changes are equiprobable and, thus, affect each voter and each part of the vote with the same probability. Moreover, we assume that the probability of replacing a vote by a new one is anti-proportional to the swap distance between the two. For this, we make use of the famous Mallows noise model [20] (see Section 3 for a detailed description of our approach). One can thus interpret our approach as a tool to measure the robustness of election winners against random equiprobable noise.

Recently, Boehmer et al. [4] conducted some experiments on the winning probabilities of candidates under Plurality and Borda on synthetic elections if elections are perturbed. Their results indicate that some elections have extremely non-robust winners and that a winning-probability based approach offers a different and more nuanced view on the robustness of election winners than established, simpler measures. In this work, we aim for comparing the robustness of different voting rules and conducting an in-depth analysis of the robustness of winners in real-world elections.

1.1 Related Work

Closest works related to ours are the papers of Baumeister and Hogrebe [2] and Boehmer et al. [4]. Both study the computational complexity of computing the winning probabilities of candidates if we replace each vote with one sampled from some distribution. Among other models, Baumeister and Hogrebe [2] considered a Mallows-based approach as used in this paper from a theoretical perspective: For some given $\phi \in [0, 1]$, each vote v is replaced by a vote sampled from a Mallows distribution with center vote v and dispersion parameter ϕ, i.e., a vote v' is sampled with probability proportional to $\phi^{\kappa(v,v')}$ (the swap distance $\kappa(v, v')$ between v and v' is the number of swaps of adjacent candidates that are needed to transform v into v'). Boehmer et al. [4] studied the related computational problem of counting elections at some given swap distance from a given initial election where some given candidate wins (see Section 3 for details). In fact, Baumeister and Hogrebe [2] showed these two problems to be equivalent from the computational perspective. Together, Baumeister and Hogrebe [2] and Boehmer et al. [4] proved strong (parameterized) intractability results for these problems for Plurality and Borda.

The problem of Boehmer et al. [4] can be phrased as the counting variant of the Swap Bribery problem. In Swap Bribery, we are given an election, a designated candidate p, and a budget k, and the question is whether we can perform k swaps of adjacent candidates in some votes to make p an election winner. Bribery problems in elections have been introduced by Faliszewski et al. [14] and have been extensively studied since then (see the overview of Faliszewski and Rothe [15]). The idea to use swap bribery for evaluating the robustness of election winners is due to Shiryaev et al. [26] (and has also been used in other contexts [6, 7, 9]): In the Destructive Swap Bribery problem we want to prevent a given candidate from winning the election by performing as few swaps as possible. The minimum cost of a successful destructive swap bribery can then act as a robustness measure (however, like the margin of victory and score difference, this measure is focused on the worst case). Boehmer et al. [4] observed that for Borda and Plurality the minimum cost of a destructive swap bribery might be disconnected from their winning-probability based approach (see Section 3).

1.2 Our Contributions

The main goal of this paper is to better understand the robustness of election winners against random equiprobable noise. In particular, we analyze what makes an election winner robust and how this is influenced by the voting rule used. We address this goal in multiple steps, thereby significantly extending the experimental work of Boehmer et al. [4], who only considered the robustness of winners under the Plurality and Borda voting rules in synthetic elections: In Section 3, we present our approach for measuring the robustness of election winners and compare it to the approach used by Boehmer et al. [4]. In essence, our measure is very similar but easier to handle and compute. In Section 4, we compare the robustness of different voting rules on synthetic data. Generally speaking, out of the considered rules, Copeland tends to produce the most robust winners, then comes Borda, then Bucklin, then STV, and Plurality produces the least robust winners. In Section 5, we analyze the robustness of real-world elections from two different sources, i.e., the Formula 1 World Championship and some form of political elections. We identify many elections with winners that are remarkably sensitive to random equiprobable noise. For example, in some editions of the Formula 1 World Championship the original winner loses with 22% probability if we make only an expected number of 5 *random* swaps of adjacent candidates in the whole election.

Furthermore, throughout the whole paper, we observe in different places that our approach allows one to identify patterns that are invisible when considering simpler robustness measures such as the score difference, and that the non-robustness of winners can be of different types. Moreover, we describe how our approach can be used to distinguish between tied election winners, thereby serving as a potential tie-breaking mechanism.

2 PRELIMINARIES

Elections. An election is a pair (C, V) where $C = \{c_1, \ldots, c_m\}$ is a set of candidates and $V = (v_1, \ldots, v_n)$ is a collection of votes. Each vote is a strict total order over all candidates. We write $v : c_1 \succ c_2$ to denote that v prefers c_1 to c_2, and for a candidate $c \in C$ we say that

v ranks c in the ith position if v prefers exactly $i-1$ candidates to c. In Section 5, we allow for top-truncated votes, i.e., strict total orders over subsets of candidates. The implicit meaning of a top-truncated vote is that the voter prefers all the ranked candidates to all the unranked ones. For a top-truncated vote, we refer to the number of candidates the voter ranks as the vote length.

Voting Rules. A voting rule is a function that maps an election to a subset of candidates that tie as winners of this election. A *scoring vector* is a vector $\mathtt{s} = (s_1, \ldots, s_m)$ with $s_i \in \mathbb{R}$ for all $i \in [m]$ and $s_1 \geq s_2 \geq \cdots \geq s_m$. A *positional scoring rule* is defined by a scoring vector $\mathtt{s}$: Each voter awards s_i points to the candidate it ranks in the ith position for each $i \in [m]$.[1] All candidates with the maximum summed score win. Two rules that are of particular importance in our analysis are Plurality, which corresponds to the scoring vector $(1, 0, \ldots, 0)$, and Borda, which corresponds to the scoring vector $(m-1, m-2, \ldots, 1, 0)$.

Under the *Copeland* voting rule, we compute a score for each candidate and all candidates with the highest score win. A candidate c gets a point for each candidate $d \in C \setminus \{c\}$ for which more than half of the voters prefers c to d and loses a point for each candidate $d \in C \setminus \{c\}$ where more than half of the voters prefers d to c.

Under the *Bucklin* voting rule, for each candidate let i_c be the minimum $i \in [m]$ such that more than half of the voters rank c in one of the first i positions. The candidate for which i_c is smallest wins. If multiple candidates have the minimum i_c, say i^*, then the candidate(s) that appear in the most votes in one of the first i^* positions win.

Under the *single transferable vote* (STV) voting rule, we are given a strict total order $\succ_t$ of the candidates as the tie-breaking order. In each round, we delete the candidate with the minimum Plurality score. If multiple candidates have the minimum Plurality score, then the candidate that is ranked last in $\succ_t$ is deleted. The last remaining candidate is the winner of the election. Because we actively apply a tie-breaking rule for STV, there are no tied winners under STV.

Swap Distance. Given two votes v and v' over the same candidate set, their swap distance $\kappa(v, v')$ is the number of candidate pairs on whose ordering v and v' disagree (equivalently, this is the minimum number of swaps of adjacent candidates that are needed to transform v into v'). Note that the maximum swap distance between two votes over m candidates is $\frac{m(m-1)}{2}$. The swap distance between two elections $E = (C, V)$ and $E' = (C, V')$ where $V = (v_1, \ldots, v_n)$ and $V' = (v'_1, \ldots, v'_n)$ is $\sum_{i=1}^{n} \kappa(v_i, v'_i)$.

(Normalized) Mallows Distribution. For a set C of m candidates, the Mallows distribution [20] is parameterized by a central strict total order v^* over C and a dispersion parameter $\phi \in [0, 1]$. It assigns to each strict total order v over C a probability $\mathcal{D}^{v^*;\phi}_{\text{Mallows}}(v)$ that depends on the swap distance between v and v^*. Specifically, we have: $\mathcal{D}^{v^*;\phi}_{\text{Mallows}}(v) = \frac{1}{Z}\phi^{\kappa(v;v^*)}$ with normalizing constant $Z = 1 \cdot (1+\phi) \cdot (1+\phi+\phi^2) \cdot \cdots \cdot (1+\cdots+\phi^{m-1})$. For $\phi = 0$, vote v^* has probability one and all other voters have probability zero. For $\phi = 1$, all votes are drawn with the same probability. Note that we use the Mallows distribution in two different ways. On the one hand, as part of our robustness measure, we use it to perturb a given vote v', which typically means that we replace v' by a vote sampled from $\mathcal{D}^{v';\phi}_{\text{Mallows}}$. On the other hand, we use it as a model to generate elections in which case we create an election by drawing multiple votes from $\mathcal{D}^{v^*;\phi}_{\text{Mallows}}$ where v^* is the lexicographic ordering of candidates.

Unfortunately, as argued by Boehmer et al. [5], the dispersion parameter ϕ is not easy to interpret. Moreover, elections with different numbers of candidates sampled from Mallows distributions with the same fixed value of ϕ are in some sense of a fundamentally different nature. That is why we use the normalization of Mallows model proposed by Boehmer et al. [5]: Here, the Mallows distribution is parameterized by a normalized dispersion parameter norm-$\phi \in [0, 1]$, which is internally converted to a value of ϕ, such that the expected swap distance between a sampled vote and the central vote is norm-$\phi \cdot \frac{m(m-1)}{4}$. Again, norm-$\phi = 0$ leads to v^* being sampled all the time (the expected swap distance is zero) and for norm-$\phi = 1$ all votes have the same probability (the expected swap distance is $\frac{m(m-1)}{4}$). However, here, norm-$\phi = 0.5$ leads to a distribution that is in some sense in the middle between these two extremes, as the expected swap distance between the sampled and central vote is $\frac{m(m-1)}{8}$. Moreover, one value of norm-ϕ leads to the same expected relative number of swaps for different numbers of candidates, which will be vital for our purposes.

Thus, using norm-ϕ instead of ϕ basically leads to a rescaling of the considered range of perturbation. That is, there is a one-to-one mapping of values of norm-ϕ and values of ϕ. However, as argued above, norm-ϕ is much easier to use and allows for a more natural interpretation and comparison of results.

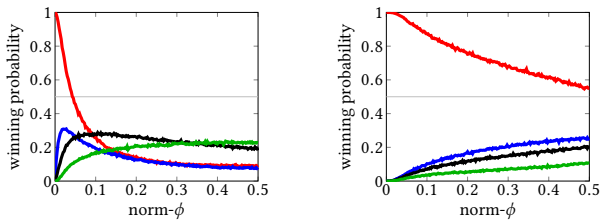
Pearson Correlation Coefficient. The Pearson correlation coefficient (PCC) is a measure of a linear correlation between two quantities, where 1 means perfect linear proportional correlation, 0 means no linear correlation and -1 means perfect linear anti-proportional correlation.

3 ASSESSING WINNER ROBUSTNESS

In this section, we describe how we asses the robustness of election winners by computing the candidate's probabilities to win the election if voters partly and randomly change their preferences. In this section, to validate our approach, we briefly mention the results of some experiments which we conducted on a diverse collection of 800 synthetic elections with 100 voters and 10 candidates collected by Szufa et al. [30] (this dataset has also been also used by Boehmer et al. [4]).

Our approach relies on the Mallows model, which is typically considered as a natural way to add random noise to an election. To model this, we replace each vote v from the election with a vote sampled from the Mallows distribution with central vote v and some normalized dispersion parameter norm-$\phi \in [0, 1]$. Specifically, for an election $E = (C, V)$, a candidate $c \in C$, and norm-$\phi \in [0, 1]$, we let $P_{E;c}(\text{norm-}\phi)$ be the probability that candidate c is a winner of an election that results from replacing each vote $v \in V$ with a vote sampled from $\mathcal{D}^{v;\text{norm-}\phi}_{\text{Mallows}}$. We refer to $P_{E;c}(\text{norm-}\phi)$ as c's winning probability at norm-ϕ and to $1 - P_{E;c}(\text{norm-}\phi)$ as

[1] For top-truncated votes in an election with m candidates, we still use the original scoring vector containing m entries. A voter which ranks $j \in [m]$ candidates then awards $\mathtt{s}_i$ points to the candidate it ranks in the ith position for each $i \in [j]$.



(a) Election generated from Mallows model with lexicographic central order and norm-$\phi = 1$

(b) Election generated from Mallows model with lexicographic central order and norm-$\phi = 0.6$

Figure 1: We plot $P_{E;c}$(norm-ϕ) (y-axis) for the Plurality voting rule as a function of norm-ϕ (x-axis) for the four most successful candidates.

c's loosing probability, i.e., the probability that c is not a winner. Notably, for norm-$\phi = 1$ each vote has the same probability under the Mallows distribution and, thus, each election has the same probability of being sampled. This implies that, assuming votes are complete, all candidates have the same probability of being a winner at norm-$\phi = 1$. In the following, we say that a winner is robust if $P_{E;c}$(norm-ϕ) does not "quickly" decrease. We often visualize the winning probabilities of different candidates as line plots. In those plots, each line represents one candidate and depicts its winning probability $P_{E;c}$(norm-ϕ) (y-axis) for different values of norm-ϕ (x-axis). We only depict the range norm-$\phi \in [0, 0.5]$ as for larger values of norm-ϕ the sampled elections have less and less similarities to the given one. We depict two example plots in Figure 1. The election displayed in Figure 1a is sampled from the Mallows model with norm-$\phi = 1$ (so each vote had the same probability of being sampled). This is also clearly visible in the plot: The winning probability of the initially winning red candidate quickly decreases. In Figure 1b, we show a more structured election (sampled from the Mallows model with norm-$\phi = 0.6$), where the winning probability of the initially winning red candidate stays high even if substantial random noise is introduced.

Comparing our approach to previous works, Boehmer et al. [4] followed a related path by computing for a given election $E = (C, V)$ and candidate $c \in C$ the probability $Q_{E;c}(r)$ that c is a winner of an election at swap distance r from E. $P_{E;c}$(norm-ϕ) and $Q_{E;c}(r)$ are indeed closely related because $P_{E;c}$(norm-ϕ) is a weighted average over $Q_{E;c}(r)$ for different values of r, as shown by Baumeister and Hogrebe [2]. From a computational perspective, computing $Q_{E;c}(r)$ (and thus $P_{E;c}$(norm-ϕ)) exactly is equivalent to solving an instance of #Swap-Bribery (simply take the number of elections at swap distance r from E where c wins and divide it by the total number of elections at swap distance r).

Unfortunately, from the results of Boehmer et al. [4] and Baumeister and Hogrebe [2] it follows that solving #Swap-Bribery and, thus, computing $P_{E;c}$(norm-ϕ) is intractable. This is why we resort to a sampling approach: To compute $P_{E;c}$(norm-ϕ) for some $E = (C, V)$, we sample an election by replacing each vote $v \in V$ by a vote sampled from $\mathcal{D}^{v;\text{norm-}\phi}_{\text{Mallows}}$. We repeat this multiple times and record for each candidate the fraction of sampled elections in which c is a winner.[2] To quantify the robustness of a non-tied election E, we use the 50%-winner threshold introduced by Boehmer et al. [4], which is the smallest value of norm-ϕ such that the winning probability of the winner of E is smaller than 50%.[3] The 50%-winner threshold thus quantifies how fast the winning probability of the initial winner declines when we move further and further away from the initial election and can be easily used to compare the robustness of election winners in different elections. Of course, instead of considering the 50%-winner threshold, one could also consider the x%-winner threshold (i.e., the smallest value of norm-ϕ such that the winning probability $P_{E;c}$(norm-ϕ) of the winner c of E is smaller than x%) for other values of x. However, for all considered voting rules, the 50%-winner threshold is strongly correlated with the 25%-winner and 75%-winner threshold on the diverse synthetic dataset of Szufa et al. [30] (the PCC is typically between 0.85 and 0.95). As fixing a single value is advantageous for clarity, we picked the 50%-winner threshold, since it has a special appeal as it quantifies the perturbation level until which the initial winner is stronger than all other candidates combined.

The main reason why we use $P_{E;c}$(norm-ϕ) instead of $Q_{E;c}(r)$, as done by Boehmer et al. [4], is that to compute $Q_{E;c}(r)$ we need to sample elections that are exactly at a given swap distance from E. Unfortunately, this sampling procedure is non-trivial and for more than 20 candidates already takes quite some time to compute [4]. In contrast to this, the approach used in this paper is much faster. Furthermore, as already argued above, both $P_{E;c}$(norm-ϕ) and $Q_{E;c}(r)$ are conceptually closely related. In particular, each value of norm-ϕ corresponds to making some expected number of swaps of adjacent candidates; of course, there is naturally some variance around this average. However, typically, for some fixed norm-ϕ, for all swap distances with a non-negligible probability of getting sampled for this norm-ϕ, in sampled elections at this distance the winning probabilities of candidates are typically quite similar (the only exception are very small values of norm-ϕ). To further analyze the relationship between $P_{E;c}$(norm-ϕ) and $Q_{E;c}(r)$, we computed the PCC of the 50%-winner threshold output by the two approaches on the synthetic dataset of Szufa et al. [30]. For Plurality, Borda, Copeland, Bucklin, and STV, the correlation is 0.991, 0.982, 0.987, 0.989, and 0.989 respectively. So, overall, for both approaches the 50%-winner thresholds are very strongly correlated.

[2] By default, for each election we computed $P_{E;c}$(norm-ϕ) for norm-$\phi \in \{0; 0.1; \dots; 1\}$. For each value of norm-ϕ, we did so by sampling 500 elections and recording for each candidate the fraction of these elections where it is a winner. To evaluate whether 500 elections are sufficient here, we also reran some of our experiments with 4000 elections sampled for each value of norm-ϕ and found that the results only marginally changed (in particular, in all elections, the 50%-winner threshold changed by at most 0.1, which is the smallest observable change). For all visualized elections, we used a finer resolution by computing $P_{E;c}$(norm-ϕ) for norm-$\phi = 0.0025 \cdot i$ for $i \in \{0; 1; 2; \dots; 200\}$ by sampling for each value of norm-ϕ 10 000 elections.

[3] For STV, we cannot simply compute $P_{E;c}$(norm-ϕ) by sampling some elections and recording in how many of them c is a winner, because deciding whether some candidate is a winner under STV in some given election is NP-hard [12]. Thus, a tie-breaking rule needs to be specified. To deal with this issue, for each run of STV on some election, we sample a strict total order $\succ_t$ over all candidates uniformly at random from the set of all strict total orders and break ties according to $\succ_t$. This in particular implies that as we do 500 runs at norm-$\phi = 0$, i.e., we apply STV 500 times to the initial election with different tie-breaking orders, multiple candidates may have a non-zero winning probability in the initial election. We consider as the initial winner the candidate having the highest winning probability at norm-$\phi = 0$ and for elections where no candidate has a winning probability over 50% at norm-$\phi = 0$, we set the 50%-winner threshold to 0.

Finally, note that if we consider an election containing some top-truncated vote v (such votes appear in our real-world data), then we do not adjust our procedure and still replace v by a vote sampled from the Mallows distribution with v as the central vote and the given normalized dispersion parameter. This means that candidates that do not appear in the vote in the original election will never be added to it. We proceed in this way because otherwise we would need to make some (artificial) assumptions about the insertion probabilities of the non-ranked candidates. Moreover, in most of our applications, non-ranked candidates are not included in some vote "by design". For instance, in our political elections not all parties nominate a candidate in each voting district. Note also that in elections with top-truncated votes, winners can be both particularly robust and particularly non-robust: If the winner appears in all votes and all other candidates only appear in few votes, then the winner will still be the (by far) most probable winner for any value of norm-ϕ, as even if each vote is replaced with a uniformly at random sampled one, the winner is most likely to have the strongest standing in the election. In contrast, top-truncated votes also open up the possibility for very non-robust winners: Consider as an example a Plurality election consisting of two candidates c and d, where x voters rank c in the first position (and do not rank d at all) and $x+1$ voters rank d in the first position and c in the second position. Then d wins the election; however, in each election at swap distance $r > 0$, c wins.

4 COMPARING THE ROBUSTNESS OF DIFFERENT VOTING RULES ON MALLOWS ELECTIONS

In this section, we conduct a comparison of the robustness of different voting rules using synthetic elections generated from a variant of the Mallows model.

Setup. For different voting rules, for norm-$\phi = \frac{1}{15} \cdot i$ with $i = 0, \ldots, 15$, we sampled 500 elections with 10 candidates and 100 voters from the Mallows model with lexicographic central order and normalized dispersion parameter norm-ϕ. We reversed each sampled vote with probability x% for $x \in \{0, 30, 50\}$. The intuitive meaning of this model is that the electorate is split into two groups and the "ground truth" (the central vote in the Mallows model) of one group is the reversed "ground truth" of the other. Subsequently, for each sampled election E, for norm-$\phi = \{0, 0.1, 0.2, \ldots, 1\}$, we estimated $P_{E;c}$(norm-ϕ) using 500 samples.[4]

Results. In Figure 2, we compare the robustness of five voting rules. The results draw a mixed picture: For reversion probability 0% which means that we simply consider elections sampled from the Mallows model, all voting rules become less robust as norm-ϕ grows (which is also quite intuitive, as the votes in the sampled elections become more and more different from each other). Moreover, there is a clear ranking of the voting rules in terms of their robustness independent of norm-ϕ: Copeland and Borda produce the most robust results. Bucklin is the third most robust rule, then STV, and

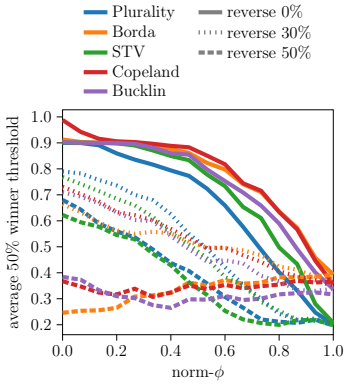


Figure 2: Average 50%-winner threshold of different voting rules for elections sampled from the Mallows model with varying norm-ϕ where each sampled vote is reversed with some probability.

Plurality is the least robust rule. The results highlight that rules are most robust if they take into account the "full election" without local focus, as this prevents the existence of strong "hidden" contenders.[5]

Notably, the robustness difference between the rules is largest for norm-$\phi = 1$: For Plurality and STV the average 50%-winner threshold here is around 0.2, while for the other three rules it is around 0.4. This large gap is quite remarkable, as these are in some sense the elections containing the least structure and information.

When we start to reverse the sampled votes with some probability, winners become less robust (which is quite intuitive, as in case we reverse half of the votes, in expectation the first and last candidate from the central vote are equally strong). For Plurality and STV for norm-$\phi \in [0, 0.8]$, if we reverse votes with some probability, then the average 50%-winner threshold is simply shifted down by some constant value compared to the 50%-winner threshold if we do not reverse any votes. In contrast to this, for the other three voting rules, the robustness of elections for norm-$\phi = 0$ and norm-$\phi = 1$ becomes more and more similar as we increase the reversion probability (see the dotted line in Figure 2 for reversion probability 30% and the dashed line for reversion probability 50%). For reversion probability 50%, for these rules, elections with norm-$\phi = 1$ are even slightly more robust than the ones for norm-$\phi = 0$.

To explain this different behavior of the voting rules, let us focus for a moment on elections sampled from the Mallows model with norm-$\phi = 0$, central vote $c_1 \succ \cdots \succ c_m$, and 50% reversion probability: In these elections, around half of the votes, say x, are $c_1 \succ \cdots \succ c_m$ and the other half, say y, are $c_m \succ \cdots \succ c_1$. If $x > y$, then c_1 is the (strict) majority winner and thus the winner under Bucklin, Copeland, Plurality, and STV. It is also easy to see

[4] As reported by Boehmer et al. [5] real-world elections are typically "close" to some elections generated from the Mallows model. We reverse votes with some probability because the resulting elections are more interesting from a robustness perspective, as they illustrate the different behavior of voting rules.

[5] An example of such a "hidden" contender for both Bucklin and Plurality is the candidate b from the election E from the introduction (where there are 50 votes with $a \succ b \succ \cdots$ and 49 votes with $b \succ \cdots \succ a$), as in this election candidate b wins under both rules as soon as one of the first 50 voters swaps b and a. Moreover, due to the "local" nature of the rules, a cannot easily gain additional points, as a is ranked last in all votes where it is not ranked first, and for Plurality it only matters who is ranked in the first position and for Bucklin in this election it only matters who is ranked in one of the first two positions. In contrast to this, under Borda, a is also able to gain points if it is ranked in the last position (so there exist single swaps by which a can gain points which it has maybe lost by other swaps).

that c_1 is the Borda winner. However, the robustness of c_1 for the different rules varies substantially. For illustrative purposes, we just compare STV and Copeland. For STV, note that initially either c_1 or c_m is ranked in the first position in every vote (and both appear roughly the same number of times in the first position). Thus, also after some swaps are performed, it is likely that c_1 and c_m are the last two non-eliminated candidates when computing the STV winner. Accordingly, the election boils down to a pairwise comparison between c_1 and c_m. For c_m to win this comparison, it needs to be in front of c_1 in some votes where it was initially behind c_1. This requires $m-1$ specific swaps per vote, as in all such votes v_1 is initially ranked in the first and c_m is ranked in the last position. In contrast to this, for Copeland, in the initial election c_1 has score 9 and c_2 has score 7 (because c_2 wins the pairwise comparison against all candidates except c_1). Thus, for c_1 to lose the election, c_2 only needs to win the pairwise comparison against c_1, which can be achieved by performing a single swap in $x-y$ of the votes where c_1 is ranked in the first and c_2 in the second position.

Having explained the different behaviors of voting rules if we reverse votes, let us remark that from a normative perspective it seems to be more reasonable to expect that a winner in an election generated from the Mallows model with norm-$\phi = 0$ and 50% reversion probability is not too robust. In the end, these elections will always only be decided by the (small) difference between the number of reversed and non-reversed sampled votes. So from this perspective, Borda, Bucklin, and Copeland are advantageous here, despite (and in fact because) they are less robust.

5 EXPERIMENTS ON REAL WORLD DATA

We analyze the robustness of real-world election winners. For this, we not only use the original election data but also the original voting rule. We address the following four questions: **Q1.** How sensitive are election winners to equiprobable noise swaps in different types of real-world elections? **Q2.** Are there real-world elections where very few random swaps change the election outcome with high probability? **Q3.** Do the winning probabilities of candidates behave similarly in all "close" elections? **Q4.** Can the robustness of winners to random swaps be assessed via alternative (simpler) measures?

To answer these questions, we consider two types of real-world elections: sports elections and political elections. In Section 5.1 we analyze the robustness of winners of the Formula 1 World Championship. After that, in Section 5.2, we turn to high-stake political elections. As voters in large political elections typically do not reveal their full preferences, we focus on first-past-the-post elections where voters are partitioned into districts and each district sends one representative to the parliament, and we study the robustness of the party winning the most seats in such Plurality elections.

To assess candidates' winning probabilities for different perturbation levels we used the same procedure as in the previous section. As described in Section 3, for each election $E = (C, V)$ we computed $P_{E;c}(\text{norm-}\phi)$ for norm-$\phi \in \{0, 0.1, \ldots, 1\}$. For each value of norm-ϕ, we did so by sampling 500 elections where each vote $v \in V$ is replaced by a vote drawn from $\mathcal{D}_{\text{Mallows}}^{V;\text{norm-}\phi}$ and recording for each candidate the fraction of these elections where it is a winner.

years	scoring vector
2010-2018	$\mathbb{s}_{2018} = (25, 18, 15, 12, 10, 8, 6, 4, 2, 1, 0, \ldots, 0)$
2003-2009	$\mathbb{s}_{2009} = (10, 8, 6, 5, 4, 3, 2, 1, 0, \ldots, 0)$
1991-2002	$\mathbb{s}_{2002} = (10, 6, 4, 3, 2, 1, 0, \ldots, 0)$
1981-1990	$\mathbb{s}_{1990} = (9, 6, 4, 3, 2, 1, 0, \ldots, 0)^{\dagger}$

Table 1: Scoring vectors used in different editions of the Formula 1 World Championship. † : Between 1981 and 1990, computing the final score of a candidate, instead of summing up its points from all races, only the 11 highest scores were taken into account.

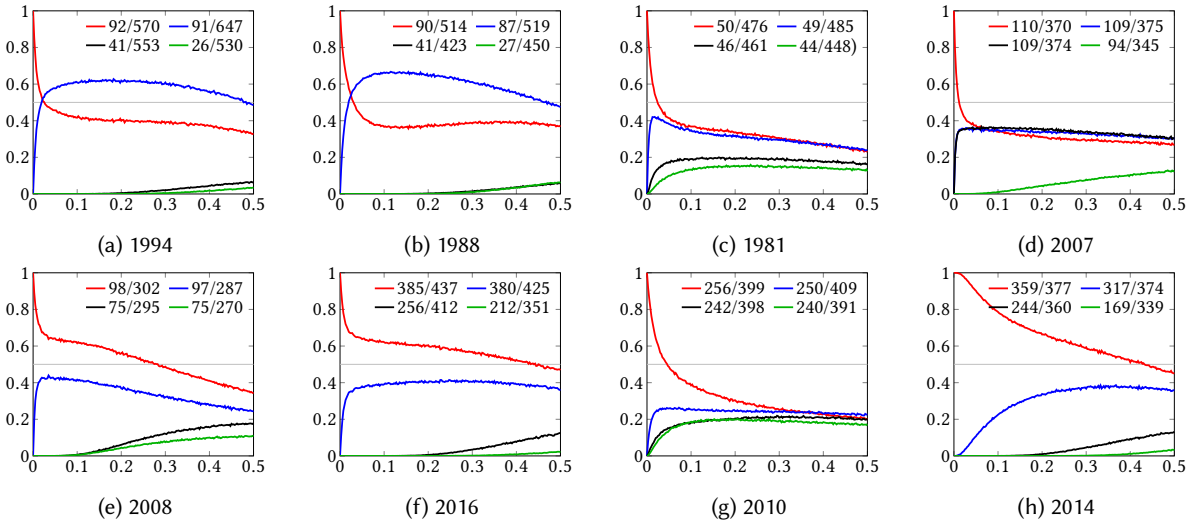
5.1 Formula 1

We consider the 38 editions of the Formula 1 World Championship between 1981 and 2018, where in each year, between 20 to 47 drivers competed in between 15 to 21 races.[6] Each driver gets a certain number of points from each race depending on its finishing position, and the candidate with the highest number of points wins (this can be interpreted as applying a positional scoring rule to the corresponding election). Over the years, different scoring vectors were used. We present them in Table 1.

*5.1.1 General overview of results (**Q1**&**Q2**).* Generally speaking, it is surprising how fragile the victory of numerous Formula 1 world champions was: The average 50%-winner threshold in our dataset is only 0.36[7], and eight elections have a 50%-winner threshold below 0.1. Going into more detail, in Figure 3, we visualize eight elections of special interest. Figures 3a to 3f all display generally quite close election: In all six the losing probability of the initial winner is already above 10% at norm-$\phi = 0.0025$. In 2007, where norm-$\phi = 0.0025$ corresponds to making on average 5 *random* swaps in the whole election, the losing probability of the original winner is even 22%. This is remarkable recalling that these elections are not artificial examples but come from the real world and recalling that we focus on random swaps, implying that there is also a very high chance that none of the top candidates are involved in a random swap, in which case the initial winner still wins. In the 2007 election, where the score of the red candidate is initially one higher than the score of the blue and the black candidate, a possible explanation for the observed general non-robustness is that the scoring vector $\mathbb{s}_{2009} = (10, 8, 6, 5, 4, 3, 2, 1, 0, \ldots, 0)$ was used. Thus, swapping down the red candidate in one vote or swapping up the black or blue candidate in one vote can suffice to make the red candidate lose the election, as a single swap can change scores by two. In fact, taking a closer look, even in 20 out of the 357 elections at swap distance 1 of this election the red candidate is not a winner. This means that even if we just make a single random swap the loosing probability of the red candidate is already 5.6%.

[6] In the corresponding election, we have one candidate for each driver and one voter for each race ranking the drivers according to the finishing position of the driver in this race. Drivers who did not participate in a race or did not finish it do not appear in the respective vote. The elections were collected by Boehmer and Schaar [8].

[7] Remarkably, this means that the average 50%-winner threshold here is lower than the average 50%-winner threshold of our considered rules on most of the Mallows elections analyzed in Section 4. This is even more remarkable recalling that the voting rules used in Formula 1 elections are rather on the robust side, as each voter awards points to many different candidates.



(a) 1994 (b) 1988 (c) 1981 (d) 2007

(e) 2008 (f) 2016 (g) 2010 (h) 2014

Figure 3: Eight Formula 1 elections. We plot $P_{E;c}(\text{norm-}\phi)$ (y-axis) as a function of norm-ϕ (x-axis) for the four most successful candidates. For each candidate, the first legend entry displays the score of the candidate in the election according to the used voting rule and the second entry displays its Borda score.

*5.1.2 Different types of close elections (**Q3**).* While all six elections from Figures 3a to 3f were really close in the sense that few random swaps can have a crucial influence on the outcome, the (non)-robustness of the winners in these six elections still comes with quite different flavors. In Figures 3a and 3b, the blue candidates overtakes the initially winning red candidate already at norm-$\phi = 0.0275$ and afterwards consistently has a higher winning probability; in such elections one could say that the winner won more by luck or accident than by merit, as in most elections close to the original election a different candidate wins. Let us focus for a moment on the 1994 Formula 1 World Championship with scoring vector $s_{2002} = (10, 6, 4, 3, 2, 1, 0, \ldots, 0)$, which consists of 16 races and was decided by one point (Figure 3a). What stands out is that the red candidate won 8 of the 16 races and came in second in 2 races, but either did not participate or did not complete the other 6 races. Thus, if we perform a single random swap involving the red candidate, then in 8 cases he loses 4 points, in 2 cases he loses 2 points and in 2 cases he gains 4 points. Thus, for 10 out of 12 swaps involving the red candidate, the blue candidate wins the election after performing the swap (note, however, that only 7.4% of all swaps involve one of the top-two candidates). Accordingly, the general non-robustness of the red candidate here is due to the fact that the red candidate is much more likely to lose points instead of gaining more if few random swaps are performed.

In contrast to Figures 3a and 3b, in Figures 3c and 3d, the red candidate starts to have roughly the same winning probability as some other candidate(s) at small norm-ϕ, however with increasing norm-ϕ the situation does not change. In such elections, it seems that the red candidate's victory was very fragile and that the top-two candidates are in fact of equal quality. Lastly, in Figures 3e and 3f, while the red candidate already starts to have a significant loosing probability at small norm-ϕ, its winning probability until norm-$\phi = 0.5$ is always clearly the highest. In such elections it seems that the red candidates victory is a bit fragile but still "justified" and grounded on solid support.

*5.1.3 Relationship between winner robustness and candidate scores (**Q4**).* Motivated by the observation that in all six considered close elections the score difference between the winner and the runner-up is between one and five and thus, in general, quite low, we now discuss the capabilities of the difference of the score of the election winner and runner-up to judge the robustness of winners. In general, there clearly is some correlation: The PCC of the score difference and the 50%-winner threshold in the Formula 1 elections is 0.66 and, in particular, with increasing score difference, on average, winners get substantially more robust. However, the correlation is not strong and there also exist elections where there is a clear difference: For instance, in 2010 (Figure 3g), the score difference is six but the 50%-winner threshold is still only 0.0475 (and thus, in particular much lower than in 2016 (Figure 3f) where the score difference is five). In 2014 (Figure 3h), the score difference is 42 and thus quite high, which is also reflected in a 50%-winner threshold of 0.4325. However, remarkably, the initial winner's loosing probability is already 1% at norm-$\phi = 0.0175$ and 10% at norm-$\phi = 0.04$, indicating that the election was much closer than what is suggested by the large score difference. Further, note that in 1988 (Figure 3b), where the red candidate seems to have won more by luck or accident than by merit, the score difference is three, whereas in 2008 (Figure 3e), where the red candidate still dominates all other candidates in terms of winning probabilities even if many swaps are performed, the score difference is only one.

A different possible approach to identify (different types of) close elections could be to consider the candidates' Borda score. The hope here is that the Borda score captures the general strength of the candidate in the election (which is not necessarily captured in the Formula 1 score, as here only points for finishing in one of the first positions are awarded). Generally speaking, the Borda score correlates with our classification of the six close elections from Figures 3a to 3f (see Section 5.1.2): In particular, if the Borda score of one candidate is significantly higher than the scores of all other

candidates, then this candidate will be the most probable winner for medium and large values of norm-ϕ. However, also the Borda score has some clear limitations: In 1988 (Figure 3b), the blue candidate has a lead of 4 Borda points, while in 1981 (Figure 3c) its lead is 9 Borda points; nevertheless, in 1988 the blue candidate quickly becomes the most probable winner, which does not happen in 1981.

5.2 Political Elections

After we have seen that sports elections regularly have non-robust winners, we now turn to a different type of election, political elections, mainly focusing on **Q1**&**Q2**. As in large political elections full preferences of voters are typically unknown, we study a specific type of political election: Several countries around the world use first-past-the-post voting in elections of different representative bodies. In these elections, the voters are typically partitioned into constituencies, with each constituency having its distinct candidates. Each constituency then sends the candidate with the highest number of votes to the representative body (and the representative body only consists of these candidates). We analyze the robustness of the strongest party in representative bodies elected by first-past-the-post elections. For this, we identify each candidate running in some constituency by its party. Then, we create one vote for each constituency ranking in the ith position the party of the candidate finishing in the ith position in this constituency (in all elections we considered, there are never two candidates from the same party running in one constituency). Now the Plurality score of a party in the constructed election is the number of seats the party gets in the representative body and the Plurality winner is the party with the highest number of seats. Which party has won the most seats in a representative body is of great practical importance, as these parties are typically responsible for leading the formation of a new government and in some elections also decide on who should fill the most important political role (e.g., in UK general elections the winning party usually decides on who should be the Prime minister). Thus, if the robustness of the winning party is low, then one might want to consider a recount of the ballots or if the winner's robustness is low in some polls, then parties have additional motivation to mobilize as many voters as possible.

We observe that such political elections seem to be very robust with the 50%-winner threshold typically being above 0.7 and are thus in particular much more robust than the Formula 1 elections considered in Section 5.1 and most of the synethetic elections analyzed in Section 4. Nevertheless, also non-robust winners regularly occur, highlighting the relevance of searching for them.

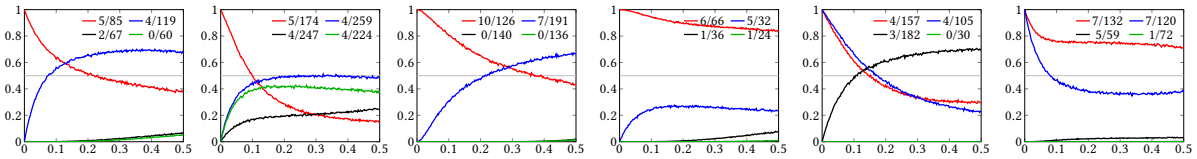
5.2.1 Polish local elections. We analyze local council elections for different Polish cities from 2014. In 2014, in all cities with up to 100 000 inhabitants a first-past-the-post system was used. For this, all cities with up to 20 000/50 000/100 000 inhabitants where divided into 15/21/23 constituencies. Our dataset consists of elections from 1317 cities (we did not include elections with an average vote length below 3) each containing on average 8.6 candidates. Notably, out of the 1317 elections 124 are tied. We generated the elections based on data provided by the Jagiellonian Center for Quantitative Research in Political Science.

Concerning the elections' robustness, in a large majority of the non-tied elections the winners are very robust; the average 50%-winner threshold is 0.78 which is very high: In fact, only seven elections have a 50%-winner threshold below 0.2 and only 98 have a 50%-winner threshold below 0.5. Overall, this is good news indicating that the considered type of political election is typically quite robust to random noise and that one does not have to worry about winners winning more by accident or luck than by merit.

Going into more details, in Figure 4, six exemplary elections of special interest are shown. The two elections from Ozorków (Figure 4a) and Wałcz (Figure 4b) were both decided by a single point (seat) and are in general quite close with a 50%-winner threshold of 0.22 and 0.1, respectively. Moreover, in both elections the initial winner has a 2% losing probability already at norm-$\phi = 0.0025$, which corresponds to making an expected number of around 0.37 swaps in the full election. While it might sound counter-intuitive that a score difference of one can be overcome by making 0.37 swaps, note that this is due to the fact that the Mallows distribution has some variance in the number of swaps that are performed. To explain why winners have a non-negligible loosing probability already for small norm-ϕ in these elections, note that, for instance, in the election from Ozorków (Figure 4a) consisting of 10 candidates and 15 voters, four voters rank the red candidate in the first position and the blue candidate in the second position. As there are overall 87 different swaps, it follows that after making a single random swap the losing probability of the red candidate is $\frac{4}{87} = 4.6\%$. This closeness highlights the importance of detecting such situations in order to be able to double check the integrity of the results.

While such situations where already very few random swaps can change the election winner with a non-negligible probability occur mostly in elections with a score difference of one, there are also some (less) extreme examples with a larger score difference. For instance, in Oświęcim (Figure 4c), the score difference is three but nevertheless, the loosing probability of the initial winner is already 1% at norm-$\phi = 0.015$ and 10% at norm-$\phi = 0.06$ (which corresponds to making an expected number of 2.3, respectively, 9 swaps in the whole election). In contrast to the former three examples, there are also elections with very robust winners, even several with just a score difference of one. The election in Księżpol (Figure 4d) is an example; together with Figures 4a and 4b this election also shows that only examining the Plurality scores is insufficient to judge the robustness of election winners. This general disconnect is also reflected in a low PCC of 0.48 between the score difference and the 50%-winner threshold on the whole dataset.

Examining the 124 tied elections, interestingly, our approach is able to identity different types of ties: On the one hand, we have numerous tied elections where the winning probabilities of the different initially winning candidates behave very similarly if more and more swaps are performed (see Figure 4e for an example; this election is also quite interesting because the initially third-ranked candidate seems to be particularly strong). On the other hand, in many of the tied elections, the winning probability of one of the winners decreases much faster than for the other, indicating that the later has a stronger general position in the election and is closer to uniquely winning the election than the other candidate (see Figure 4f for an example). This indicates that ties in elections might be of a different nature and that our approach might be a first possibility to better understand and identify them. In the analyzed



(a) Ozorków (b) Wałcz (c) Oświęcim (d) Księżpol (e) Kazimierz Biskupi (f) Jelcz-Laskowice

Figure 4: Six Poland elections. We plot $P_{E;c}(\text{norm-}\phi)$ as a function of norm-ϕ for the four most successful candidates. The legend displays the Plurality score and Borda score of each candidate in this order.

political elections, tie-breaking is of special importance because usually one party is appointed to form a government.

5.2.2 UK general elections. We now turn to national elections in the UK. In particular, we consider the twelve general elections (of the House of Parliaments) in the UK which took place between 1974 and 2019. From this, we obtained twelve elections with between 9 and 13 candidates and 635 and 659 voters. The elections were created by us based on data from the official website of the house of commons, commonslibrary.parliament.uk.

As in the Polish local elections, in general the robustness of winners is quite high in these elections; the average score difference between the winner and the runner-up is with 118 also quite high. In particular, only two out of the twelve analyzed elections have a 50%-winner threshold below 0.9; both of them are from the year 1974. In this year, there was one election in February and a reelection in October: The October election has a 50%-winner threshold of 0.5 and a score difference of 42 (where only at norm-$\phi = 0.1$ the winning probability of the initial winner drops below 99%). Thus, in this election, the winner is still pretty robust. The February election is much closer. This election consists of 635 votes over twelve candidates with an average vote length of 3.3; the Labor party won with 301 seats against the Conservative party with 297 seats. The 50%-winner threshold of this election is only 0.07, which corresponds to performing 88.7 swaps in the whole election in expectation. However, even for norm-$\phi = 0.01$, which corresponds to making 12.67 swaps in expectation, the losing probability of the Labor party is already 11.7%.[8] The general non-robustness of the Labor parties victory in this election is also reflected in the candidates' Borda scores, as the Borda score of the Conservative party is 95 points higher. To sum up, UK general elections seem to be quite robust to our noise model; however, the February 1974 election constitutes a clear outlier as the win of the Labor party in this election is fragile. In fact, the Labor party did not win the absolute majority of seats in this election and, possibly as a consequence of the non-robustness of their victory, coalition talks failed. After the Labor party governed for a short time as a minority government, a reelection was initiated. In this reelection, the Labor party won again but this times with a larger lead (also being robust to random swaps). The clear difference between the two elections which happened in the span of 8 months indicates that in political elections a significant fraction of voters can change their mind shortly after an election. This additionally motivates the study of the robustness of outcomes as an indicator for the likelihood that the outcome still reflects the voter's opinions even some time after the election, and also motivates that larger numbers of random swaps can realistically happen.

[8] Notably, even at norm-$\phi = 0{:}0025$, which corresponds to making an expected number of 3:2 swaps in the election, the losing probability is already 1% despite the fact that the initial score difference is 4 (this is due to the fact that the Mallows model has some variance in the number of swaps it actually applies).

6 CONCLUSION

In this paper, we have studied how robust election winners are to equiprobable random noise by comparing different voting rules and computing and analyzing the robustness of real-world election winners. As one of our highlights, we have identified many real-world election winners that are very sensitive to random noise, indicating the these elections were extremely close. Moreover, we have illustrated that our approach can detect a variety of different patterns and can differentiate between seemingly very similar elections. For future work it would be interesting to dive deeper into the capabilities and limitations of our approach, for instance, by further exploring the possibility to use it as a tie-breaking mechanism.

While we have already tried to make our experiments relevant to practitioners, there is certainly room for improvement: Because it is the (computationally) simpler and cleaner approach, we have considered an unweighted model where each swap has the same probability. However, this might not fully capture reality in all its facets: For instance, in the Formula 1 elections, one could argue that the probability of swapping two drivers in a race should be anti-proportional to their difference in finishing time. Moving from the unweighted to the weighted setting would also require collecting the needed weighted data, which for some type of elections is also simply not available. From a data collection point of view, it would also be beneficial to collect the full preferences of voters in political elections (and not only top-choices as it is usually done) to analyze the robustness of large scale real-world political elections. Poll stations are probably the most promising starting point here.

ACKNOWLEDGMENTS

We thank Nathan Schaar for helping us to collect the used real-world elections. NB was supported by the DFG project ComSoc-MPMS (NI 369/22). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101002854).

REFERENCES

[1] John F. Banzhaf III. 1968. One man, 3.312 votes: a mathematical analysis of the Electoral College. *Vill. L. Rev.* 13 (1968), 304.

[2] Dorothea Baumeister and Tobias Hogrebe. 2021. On the Complexity of Predicting Election Outcomes and Estimating Their Robustness. In *Proceedings of the 18th European Conference on Multi-Agent Systems (EUMAS '21)*. Springer, 228–244.

[3] Nathaniel Beck. 1975. A note on the probability of a tied election. *Public Choice* 23 (1975), 75–79.

[4] Niclas Boehmer, Robert Bredereck, Piotr Faliszewski, and Rolf Niedermeier. 2021. Winner Robustness via Swap- and Shift-Bribery: Parameterized Counting Complexity and Experiments. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI '21)*. ijcai.org, 52–58. Full version available at arxiv.org/abs/2010.09678.

[5] Niclas Boehmer, Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, and Stanislaw Szufa. 2021. Putting a Compass on the Map of Elections. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI '21)*. ijcai.org, 59–65.

[6] Niclas Boehmer, Robert Bredereck, Klaus Heeger, and Rolf Niedermeier. 2021. Bribery and Control in Stable Marriage. *J. Artif. Intell. Res.* 71 (2021), 993–1048.

[7] Niclas Boehmer, Robert Bredereck, Dusan Knop, and Junjie Luo. 2020. Fine-Grained View on Bribery for Group Identification. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI '20)*. ijcai.org, 67–73.

[8] Niclas Boehmer and Nathan Schaar. 2022. *Collecting, Classifying, Analyzing, and Utilizing Real-World Elections.* Technical Report arXiv:2204.03589 [cs.GT]. arXiv.org.

[9] Markus Brill, Ulrike Schmidt-Kraepelin, and Warut Suksompong. 2022. Margin of victory for tournament solutions. *Artif. Intell.* 302 (2022), 103600.

[10] David Cary. 2011. Estimating the Margin of Victory for Instant-Runoff Voting. In *2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX Association.

[11] Gary Chamberlain and Michael Rothschild. 1981. A note on the probability of casting a decisive vote. *J. Econ. Theory* 25, 1 (1981), 152–162.

[12] Vincent Conitzer, Matthew Rognlie, and Lirong Xia. 2009. Preference Functions that Score Rankings and Maximum Likelihood Estimation. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI '09)*. AAAI Press, 109–115.

[13] Palash Dey and Y. Narahari. 2015. Estimating the Margin of Victory of an Election Using Sampling. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI '15)*. AAAI Press, 1120–1126.

[14] Piotr Faliszewski, Edith Hemaspaandra, and Lane A. Hemaspaandra. 2009. How Hard Is Bribery in Elections? *J. Artif. Intell. Res.* 35 (2009), 485–532.

[15] Piotr Faliszewski and Jörg Rothe. 2016. Control and Bribery in Voting. In *Handbook of Computational Social Choice*. Cambridge University Press, 146–168.

[16] Raphael Gillett. 1977. Collective indecision. *Behav. Sci.* 22, 6 (1977), 383–390.

[17] Raphael Gillett. 1980. The comparative likelihood of an equivocal outcome under the Plurality, Condorcet, and Borda voting procedures. *Public Choice* (1980), 483–491.

[18] List of close election results. 2022. List of close election results — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/List_of_close_election_results [Online; accessed 18-March-2022].

[19] Thomas R. Magrino, Ronald L. Rivest, and Emily Shen. 2011. Computing the Margin of Victory in IRV Elections. In *2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX Association.

[20] Colin L Mallows. 1957. Non-null ranking models. I. *Biometrika* 44, 1/2 (1957), 114–130.

[21] Thierry Marchant. 2001. The probability of ties with scoring methods: Some results. *Soc Choice Welfare* 18, 4 (2001), 709–735.

[22] Howard Margolis. 1977. Probability of a tie election. *Public Choice* (1977), 135–138.

[23] Casey B. Mulligan and Charles G. Hunter. 2003. The empirical frequency of a pivotal vote. *Public Choice* 116, 1 (2003), 31–54.

[24] Lawrence D. Norden, Aaron Burstein, Joseph Lorenzo Hall, and Margaret Chen. 2007. *Post-election audits: Restoring trust in elections.* Brennan Center for Justice.

[25] Anand Sarwate, Stephen Checkoway, and Hovav Shacham. 2013. Risk-Limiting Audits and the Margin of Victory in Nonplurality Elections. *Statistics, Politics, and Policy* 4, 1 (2013), 29–64.

[26] Dmitry Shiryaev, Lan Yu, and Edith Elkind. 2013. On elections with robust winners. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS '13)*. IFAAMAS, 415–422.

[27] Philip B. Stark. 2008. Conservative statistical post-election audits. *Ann Appl Stat* 2, 2 (2008), 550–581.

[28] Philip B. Stark. 2008. A sharper discrepancy measure for post-election audits. *Ann Appl Stat* 2, 3 (2008), 982–985.

[29] Philip B. Stark. 2009. Efficient post-election audits of multiple contests: 2009 California tests. In *4th Annual Conference on Empirical Legal Studies Paper (CELS '09)*.

[30] Stanislaw Szufa, Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. 2020. Drawing a Map of Elections in the Space of Statistical Cultures. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '20)*. IFAAMAS, 1341–1349.

[31] Scott Wolchok, Eric Wustrow, J Alex Halderman, Hari K Prasad, Arun Kankipati, Sai Krishna Sakhamuri, Vasavya Yagati, and Rop Gonggrijp. 2010. Security analysis of India's electronic voting machines. In *Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS '10)*. ACM, 1–14.

[32] Lirong Xia. 2012. Computing the margin of victory for various voting rules. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC '12)*. ACM, 982–999.

[33] Lirong Xia. 2021. How Likely Are Large Elections Tied?. In *Proceedings of the 22nd ACM Conference on Economics and Computation (EC '21)*. ACM, 884–885.