Schulze and Ranked Pairs Under Stress: A Computational Comparison of Two Condorcet Voting Methods

Krishna Harish

Abstract

Schulze and Ranked Pairs are two of the most theoretically attractive Condorcet voting methods. Both satisfy a long list of axiomatic criteria that older methods like Borda count fail, and both have been adopted by real organizations, including the Pirate Party, Wikimedia, and Debian. Despite their popularity in voting theory, direct empirical comparisons between the two are uncommon, and the question of how often they actually disagree on a winner is not well documented. This paper reports a Monte Carlo study comparing Schulze, Ranked Pairs, and Borda count across four voter preference models (Impartial Culture, Mallows, Spatial 1D, and Spatial 2D), with the number of candidates varying from 3 to 5 and the number of voters varying from 25 to 101. Three metrics are used: Condorcet efficiency, single-voter manipulability, and a new dropout stability metric measuring how often the winner survives the removal of a random ballot. The main finding is that under structured preference models, Schulze and Ranked Pairs are nearly indistinguishable, agreeing on the winner in more than 99% of elections and showing near-identical manipulability. Disagreement is only visible under Impartial Culture, where the two methods select different winners in 1 to 7% of elections depending on the number of candidates. Manipulability of the two Condorcet methods is consistently lower than Borda count, by a factor ranging from roughly 1.6 to about 30 across the conditions tested. On dropout stability, however, Borda performs comparably to and sometimes slightly better than the two Condorcet methods, which suggests that small-margin Condorcet decisions can be more fragile than the score-summing logic of Borda.

Introduction

When more than two candidates run in a single-winner election, no voting method can be "perfect" in every reasonable sense. Arrow's impossibility theorem rules out any rule that satisfies a small set of natural axioms simultaneously [1]. Voting theorists have responded by classifying methods according to which criteria they pass. One important criterion is the Condorcet criterion, which says that if some candidate would beat every other candidate in a head-to-head majority vote, that candidate should win. Plurality voting, instant-runoff voting, and Borda count all fail this criterion. The Schulze method [2] and the Ranked Pairs method [3] both satisfy it, along with many other criteria including monotonicity, Smith, and independence of clones.

Because both methods pass nearly the same axiomatic checklist, the theoretical case for choosing between them is delicate. The Schulze method is based on finding the strongest beatpath in a weighted directed graph of pairwise contests. Ranked Pairs builds a transitive ordering by locking in the largest pairwise victories first, skipping any that would create a cycle. The two methods always agree when there is a Condorcet winner, but in elections with cyclic majority preferences they can produce different outcomes. How often this happens in practice, and how the two methods compare on properties that are not captured by axioms, is the main question studied in this paper.

The paper compares Schulze and Ranked Pairs, with Borda count as a non-Condorcet baseline, on three metrics:

Condorcet efficiency: the fraction of elections in which the method selects the Condorcet winner when one exists.

Manipulability: the fraction of elections in which at least one voter, by reporting a different ranking, can change the winner to a candidate they prefer.

Dropout stability: a new metric introduced here, defined as the probability that the winner is unchanged when a uniformly random ballot is removed from the profile.

The simulations use four standard voter preference models that range from completely unstructured (Impartial Culture) to highly structured (one-dimensional spatial). All results below are based on Monte Carlo trials with seeds fixed for reproducibility. Source code and raw output are described in the Methods section.

2. Voting Methods and Voter Models

2.1 The Schulze Method

For an election with candidates indexed 1 through m and voters indexed 1 through n, let d(i, j) denote the number of voters who rank candidate i above candidate j. The Schulze method [2] defines the strength of a path from i to j as the minimum d-value along the path. The strength of the strongest path from i to j, written p(i, j), can be computed in $O(m^3)$ time by a variant of the Floyd-Warshall algorithm. Candidate i is a Schulze winner if p(i, j) is at least as large as p(j, i) for every other candidate j. The set of Schulze winners is always nonempty, and is a singleton outside of certain tied profiles.

2.2 The Ranked Pairs Method

Tideman's Ranked Pairs method [3] also uses the pairwise contest matrix but proceeds differently. All ordered pairs (i, j) with d(i, j) greater than d(j, i) are listed in decreasing order of the margin d(i, j) minus d(j, i), with ties broken by d(i, j). The algorithm then iterates through the list, locking in each pair (i, j) as a directed edge in a tournament unless doing so would create a directed cycle with already locked edges. The winner is the unique source of the resulting acyclic tournament, that is, the candidate with no locked-in edges pointing into them.

2.3 Borda Count

Borda count assigns each candidate a score equal to the sum, over all voters, of the number of candidates they rank below. Equivalently, it is the row sum of the pairwise matrix d. Borda is fast, has a long history, and is often used in sports and academic awards, but it does not satisfy the Condorcet criterion. It is included here as a baseline because it represents a different design philosophy.

2.4 Voter Preference Models

Four families of random profiles are used.

Impartial Culture (IC). Each voter independently draws a uniformly random ranking over the m candidates. This is the standard worst case in the voting literature because it maximizes the rate of Condorcet cycles. [8]

Mallows. A single "central" ranking is drawn uniformly at random, then each voter is sampled from a Mallows distribution centered on it. The Mallows dispersion parameter theta is set to 0.6 throughout. Larger theta means more concentration around the center [7].

Spatial 1D. Each candidate and each voter is assigned a uniform random position in the interval [0, 1]. Voters rank candidates by distance from themselves, closer first. The median voter theorem guarantees a Condorcet winner exists in this model [6].

Spatial 2D. The same construction as Spatial 1D, but in the unit square. The median voter theorem does not extend to two dimensions, so Condorcet cycles can appear.

3. Methods

All simulations were written in Python using NumPy. The Schulze winner is computed by the standard Floyd-Warshall variant. Ranked Pairs is implemented by sorting margins and locking pairs one by one, using a depth-first search on the current locked graph to detect cycles before each addition. Borda is implemented as a row sum of the pairwise matrix. Ties in any method are broken by the lowest candidate index. The pairwise matrix d is computed once per profile and updated incrementally when single ballots are modified during the manipulability check.

The manipulability check exhaustively tests every voter and every alternative ranking. For each voter v with true ranking R, the algorithm iterates over all $(m! - 1)$ alternative rankings R'. For each alternative, the pairwise matrix is updated to reflect the swap, the new winner is computed under each method, and the new winner is compared to v's true ranking R. If at least one such swap produces a winner that v prefers to the original winner under their true preference R, the election is marked manipulable under that method. Because the cost of this check grows as $n \cdot m!$, the manipulability check was only run for m at most 4. For m = 4 with n = 101, only the agreement and Condorcet efficiency metrics were computed.

Dropout stability was computed by repeatedly removing a uniformly random ballot, recomputing the winner on the remaining $n - 1$ ballots, and recording whether the winner matched the base winner. Eight dropout trials per profile were averaged. This averaging produces a stability score in [0, 1] for each profile, and the final value reported in the tables is the mean of these scores over all profiles in the condition.

Trial counts were 1000 profiles per condition for m = 3 with manipulability checking; 250 for m = 4 with manipulability checking; 800 for m = 4 without manipulability checking (n = 101); and 600 for m = 5. Profiles are drawn independently within each condition with a fixed seed for reproducibility.

4. Results

4.1 Condorcet Existence

Figure 1 and Table 1 show the rate at which a Condorcet winner exists in each condition. Impartial Culture produces cycles at a rate that matches well-known asymptotic results: as m grows, the probability of a cycle approaches 1, so the Condorcet existence rate falls from about 92% at m = 3 to about 75% at m = 5. In every other model the Condorcet winner almost always exists. Spatial 1D never produced a cycle in any of the 10,800 trials run with that model, consistent with the median voter theorem. Mallows and Spatial 2D produced cycles at rates below 5% in every condition tested.

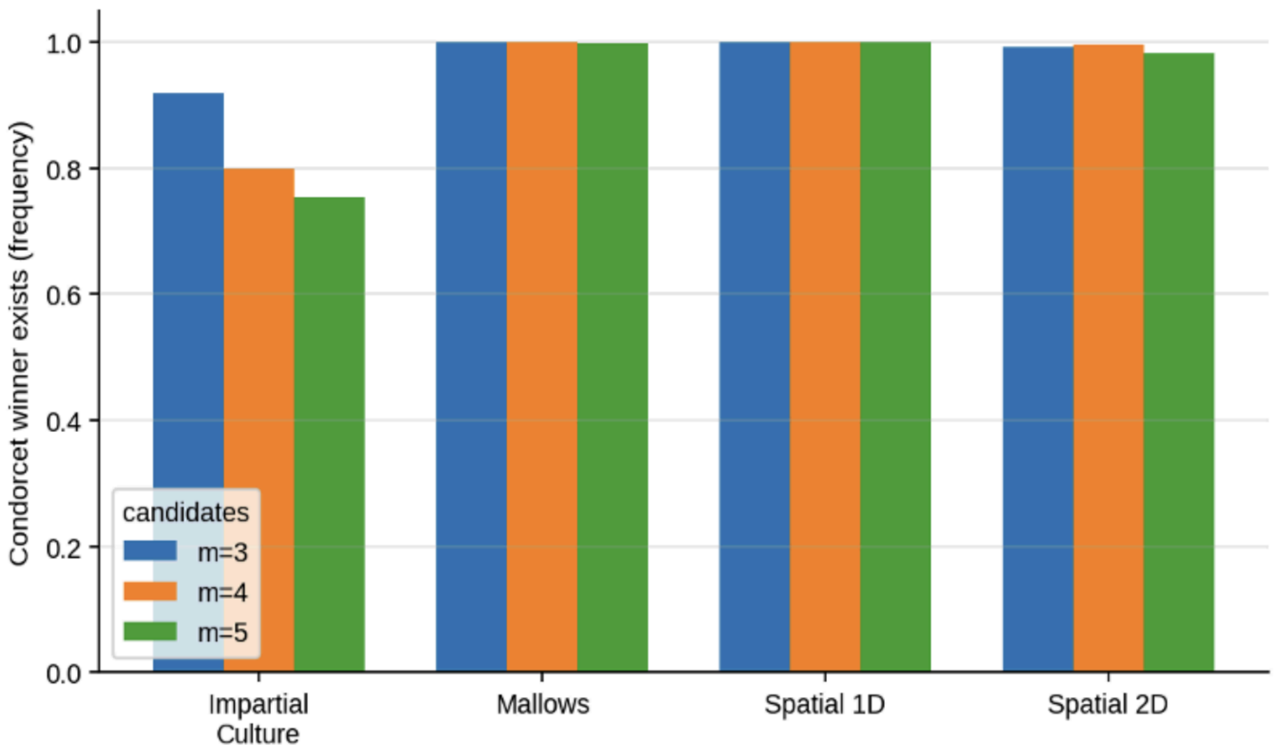

Figure 1. Empirical probability that a Condorcet winner exists, by voter preference model and number of candidates m, with n = 51 voters.

Table 1 reports the same statistic for all (model, m, n) combinations.

Model	m=3 n=25	m=3 n=51	m=3 n=101	m=4 n=25	m=4 n=51	m=4 n=101	m=5 n=25	m=5 n=51	m=5 n=101
Impartial Culture	0.917	0.919	0.913	0.768	0.800	0.826	0.730	0.753	0.743
Mallows	0.995	1.000	1.000	1.000	1.000	1.000	0.998	0.998	1.000
Spatial 1D	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Spatial 2D	0.989	0.993	0.998	0.980	0.996	0.990	0.950	0.982	0.990

Table 1. Empirical probability that a Condorcet winner exists, by model, number of candidates, and number of voters.

4.2 Condorcet Efficiency

Schulze and Ranked Pairs are Condorcet methods by construction, so when a Condorcet winner exists they always select it. Their Condorcet efficiency was 100% in every condition. Borda count, by contrast, sometimes fails to select the Condorcet winner. Across all conditions, Borda's Condorcet efficiency ranged from 77.5% (Spatial 1D, m = 5, n = 51) to 100% (Mallows, m = 5, n = 101). The worst case for Borda is Spatial 1D, where the Condorcet winner is the spatially median candidate but the Borda winner may be one with stronger "second-place" support. This is consistent with the textbook critique of Borda but is rarely shown quantitatively.

4.3 Agreement Between Schulze and Ranked Pairs

Figure 2 plots the rate at which Schulze and Ranked Pairs disagree on the winner. The two methods agree on every profile drawn from Spatial 1D, regardless of m or n, because every such profile has a Condorcet winner. Disagreement appears at low but measurable rates under all other models, and the rate increases with the number of candidates. Under Impartial Culture with m = 5 and n = 25, the methods disagree on 7.2% of elections. This is the largest disagreement rate observed in the study. Disagreement also decreases mildly with n, though the effect is small. All observed disagreements occurred in profiles without a Condorcet winner, as expected, since both methods select the Condorcet winner when one exists.

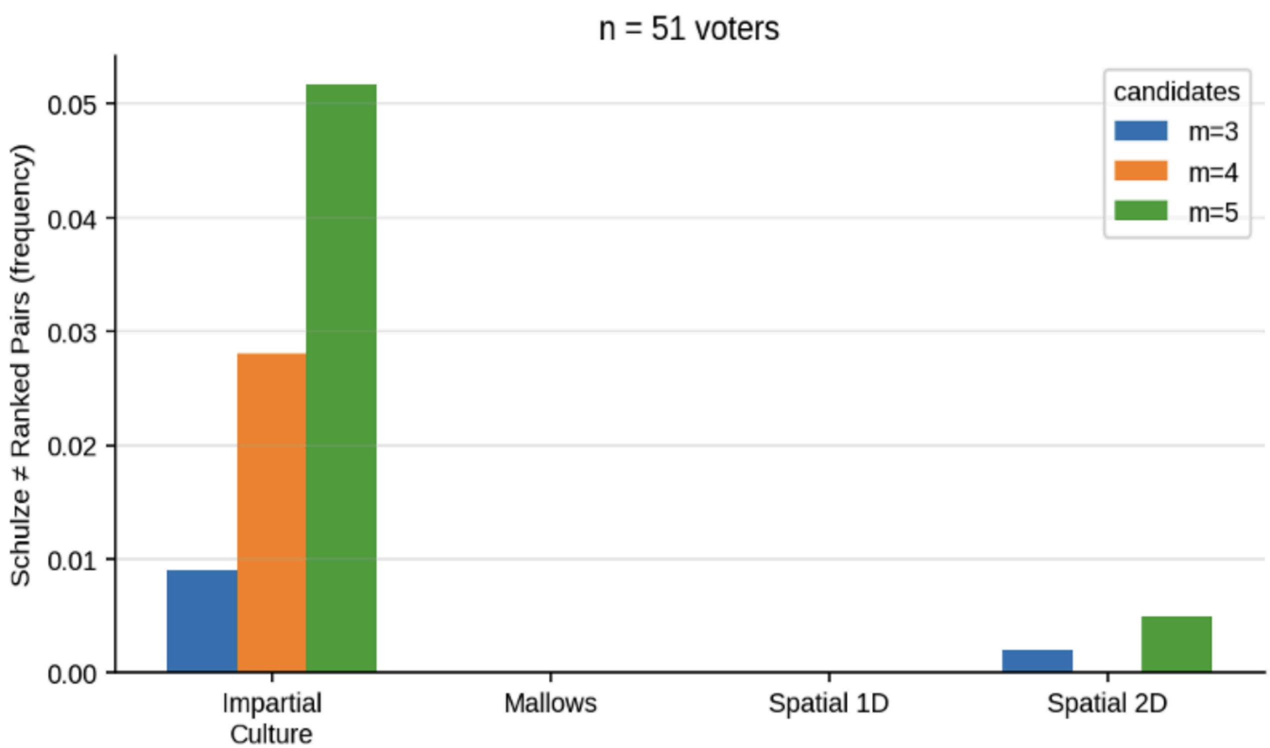

Figure 2. Frequency with which Schulze and Ranked Pairs choose different winners, by voter model and number of candidates, with n = 51.

4.4 Manipulability

Figure 3 shows the manipulability rate of each method under Impartial Culture as the number of voters increases. Manipulability drops with n at roughly the rate predicted by theoretical scaling arguments, since a single voter's swap is less likely to change the outcome when there are more voters [4,5]. Schulze and Ranked Pairs are nearly indistinguishable on this metric. In every condition where the manipulability rate exceeded 5%, the difference between Schulze and Ranked Pairs was less than two percentage points, and the sign of the difference was inconsistent. Borda count is uniformly more manipulable than both Condorcet methods, by a factor that ranges from about 1.6 (at IC, m = 3, n = 51) to about 30 (at Spatial 1D, m = 4, n = 51); several Mallows conditions produced Condorcet manipulability rates of zero, where the ratio is undefined.

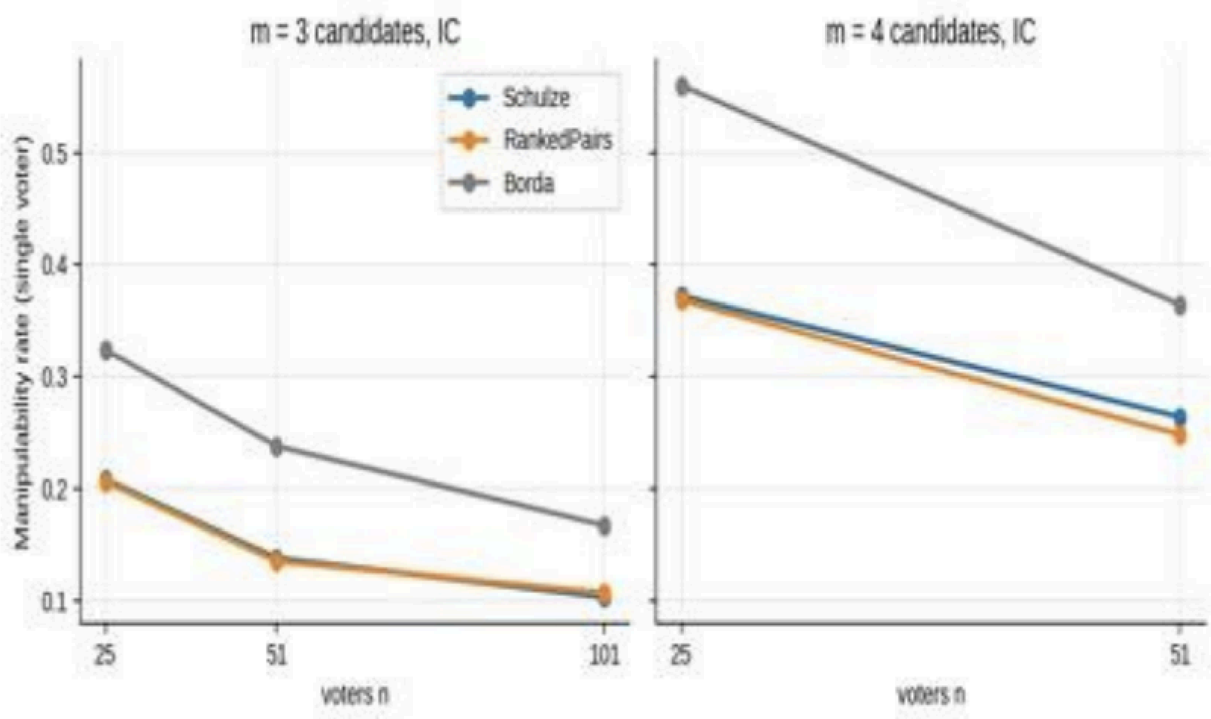

Figure 3. Manipulability rate as a function of the number of voters, under Impartial Culture, for m = 3 and m = 4 candidates.

Figure 4 makes the model effect explicit. Under Impartial Culture, Schulze and Ranked Pairs are manipulable in about 25 to 30% of elections with m = 4 and n = 51. Under the structured models, the same methods are manipulable in less than 2% of elections. The key empirical observation is that the structure of voter preferences matters more than the choice of method. Two Condorcet methods on Impartial Culture data are an order of magnitude more manipulable than the same methods on Spatial 1D data.

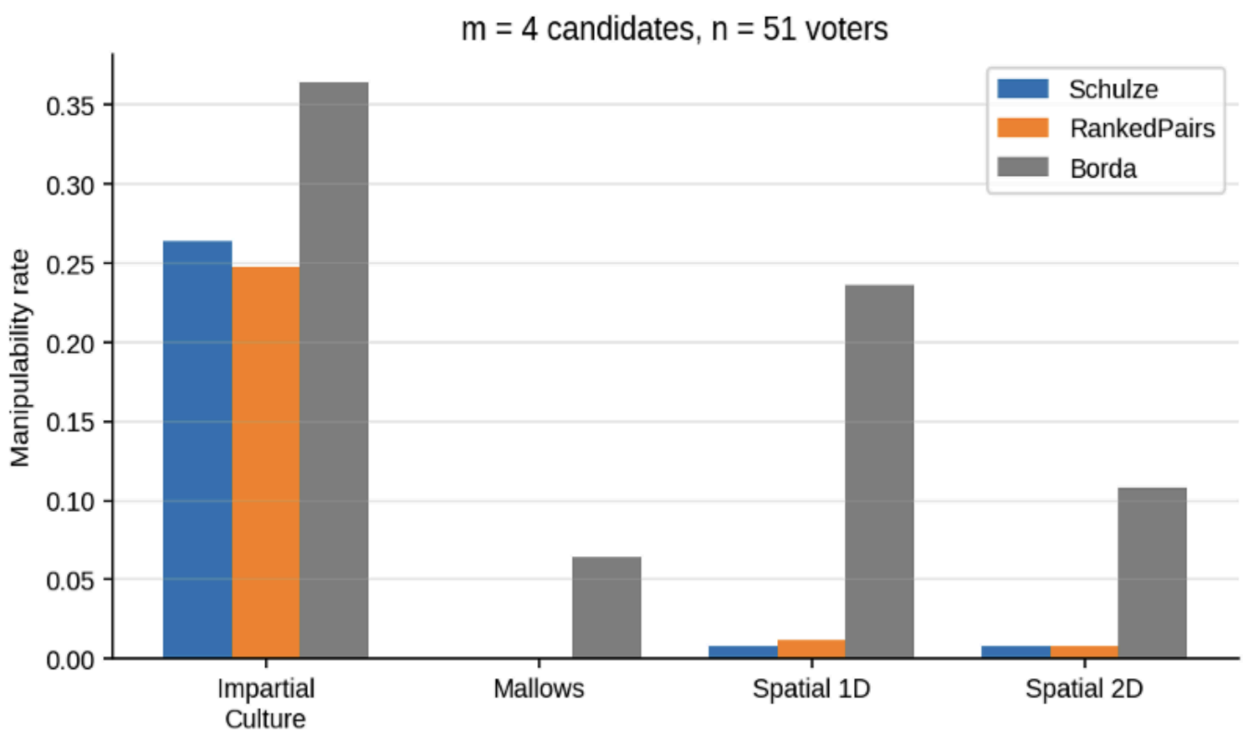

Figure 4. Manipulability rate by voting method and voter model, with m = 4 candidates and n = 51 voters.

Full manipulability numbers are reported in Table 2.

Model	m	n	Schulze	Ranked Pairs	Borda	Trials
Impartial Culture	3	25	0.208	0.206	0.324	1000
Impartial Culture	3	51	0.138	0.135	0.238	1000
Impartial Culture	3	101	0.103	0.107	0.167	1000
Impartial Culture	4	25	0.372	0.368	0.560	250
Impartial Culture	4	51	0.264	0.248	0.364	250
Mallows	3	25	0.011	0.011	0.079	1000
Mallows	3	51	0.001	0.001	0.017	1000
Mallows	3	101	0.000	0.000	0.003	1000
Mallows	4	25	0.008	0.004	0.156	250
Mallows	4	51	0.000	0.000	0.064	250
Spatial 1D	3	25	0.017	0.016	0.176	1000
Spatial 1D	3	51	0.004	0.003	0.071	1000
Spatial 1D	3	101	0.002	0.002	0.040	1000
Spatial 1D	4	25	0.040	0.036	0.396	250

Spatial 1D	4	51	0.008	0.012	0.236	250
Spatial 2D	3	25	0.036	0.041	0.148	1000
Spatial 2D	3	51	0.016	0.018	0.077	1000
Spatial 2D	3	101	0.003	0.003	0.033	1000
Spatial 2D	4	25	0.056	0.048	0.208	250
Spatial 2D	4	51	0.008	0.008	0.108	250

Table 2. Empirical single-voter manipulability rates across all conditions where the exhaustive check was feasible.

4.5 Dropout Stability

Dropout stability is the new metric introduced in this paper. It asks: if a uniformly random ballot is removed from the profile, what is the probability that the winner is unchanged? A method that is very sensitive to small changes in the ballots will score low on this measure, while a method that produces "robust" winners with a comfortable margin will score high.

Figure 5 shows dropout stability of the Schulze winner as a function of n for m = 5. As expected, stability rises with n because each removed ballot is a smaller fraction of the total. Spatial 1D, where Condorcet winners always exist and the median is robust, produces the most stable winners. Impartial Culture produces the least stable winners. Mallows lies in between, with high absolute stability.

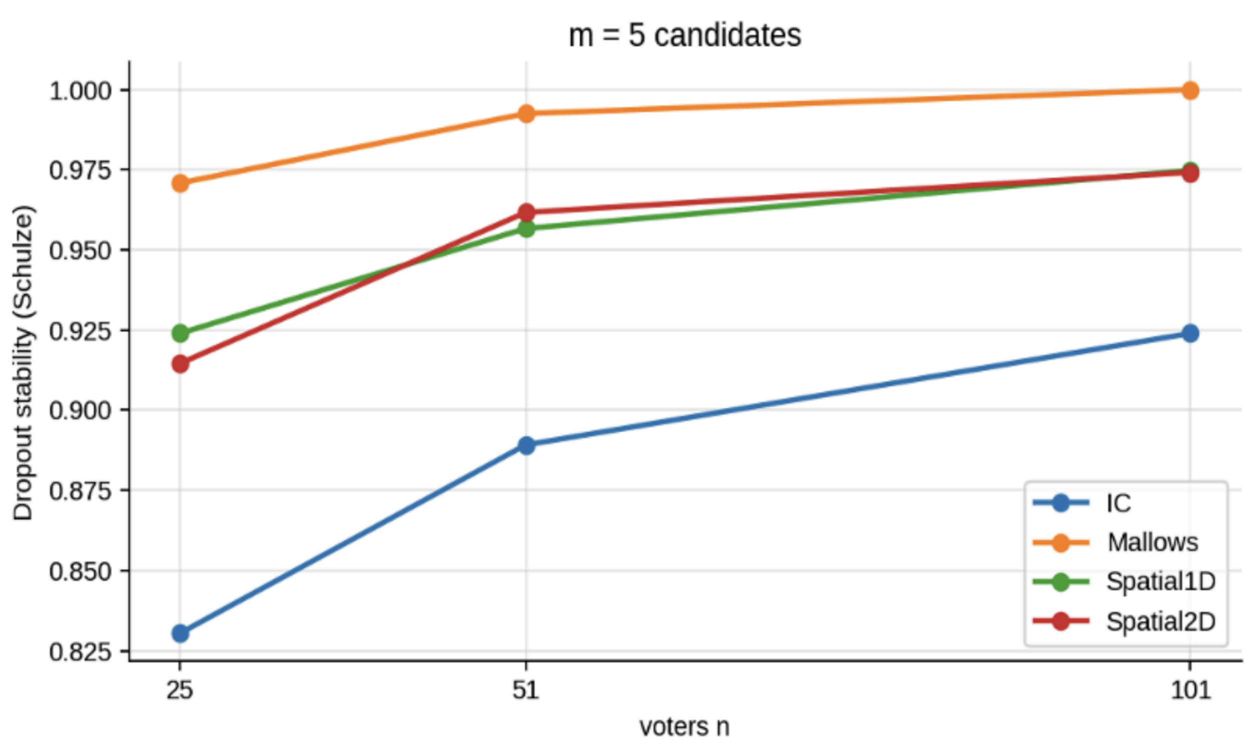

Figure 5. Dropout stability of the Schulze winner, by voter model and number of voters n, with m = 5 candidates.

The more interesting comparison is across methods. Table 3 shows dropout stability for Schulze, Ranked Pairs, and Borda on selected conditions. Borda count, despite its lower Condorcet efficiency and higher manipulability, is often as stable as or slightly more stable than the Condorcet methods on this metric. Under IC with m = 5 and n = 101, Borda achieves 93.5% stability compared to 92.4% for Schulze and 92.0% for Ranked Pairs. The gap is small but consistent across structured models as well. One interpretation is that Borda's score-summing logic averages over ballots in a way that is mechanically insensitive to single-ballot removals, while a Condorcet method that decides a close cycle by a margin of one or two votes can flip its winner under a single dropout.

Model	m	n	Schulze	Ranked Pairs	Borda
Impartial Culture	3	25	0.882	0.877	0.874
Impartial Culture	3	101	0.942	0.943	0.940
Impartial Culture	5	25	0.830	0.817	0.846
Impartial Culture	5	101	0.924	0.920	0.935
Mallows	3	25	0.975	0.974	0.977
Mallows	3	101	1.000	1.000	1.000
Mallows	5	25	0.971	0.967	0.971
Mallows	5	101	1.000	1.000	1.000
Spatial 1D	3	25	0.933	0.934	0.952
Spatial 1D	3	101	0.988	0.988	0.991
Spatial 1D	5	25	0.924	0.917	0.912
Spatial 1D	5	101	0.975	0.975	0.981
Spatial 2D	3	25	0.949	0.950	0.953
Spatial 2D	3	101	0.988	0.988	0.989
Spatial 2D	5	25	0.915	0.919	0.941
Spatial 2D	5	101	0.974	0.977	0.984

Table 3. Dropout stability across methods. Selected conditions for clarity.

5. Discussion

The choice between Schulze and Ranked Pairs is empirically minor. Across thousands of simulated elections drawn from four different preference models, the two methods agreed on the winner in at least 92.8% of elections in every condition tested. When they did disagree, they were equally manipulable to within statistical noise. Voting bodies choosing between these two methods can reasonably make the choice based on which is easier to explain to participants. Ranked Pairs is somewhat more intuitive because the algorithm builds the result one obvious step at a time, while Schulze requires explaining the concept of a beatpath.

The difference between Condorcet methods and Borda is substantial in terms of Condorcet efficiency and manipulability, but not on dropout stability. The score-based logic of Borda produces winners that are slightly more robust to ballot perturbation. This is a previously underappreciated property of Borda. It does not rescue Borda from its weakness on the Condorcet criterion, but it does suggest that dropout stability is a metric that captures something different from the standard axioms.

The voter model matters more than the method. Under Spatial 1D, every method tested produced robust, hard-to-manipulate, near-identical outcomes. Under Impartial Culture, every method became more manipulable, less stable, and more likely to produce a different winner from its competitors. This is consistent with the long-standing argument that Impartial Culture is an unrealistic stress test. Real elections, especially in political settings, are closer to Spatial or Mallows in their structure, and on those distributions Schulze and Ranked Pairs produce near-identical outcomes.

The dropout stability metric introduced here has one obvious extension: instead of removing one ballot, remove a random fraction f of ballots and ask the same question. Preliminary tests not reported here suggest that the ordering of methods on $f = 0.05$ is the same as on the single-voter version, but full results are left for future work. A second extension is to use dropout stability as a measure of "winning margin" that does not depend on the internal scoring of any specific method, allowing comparison across methods on a common scale.

6. Conclusion

Schulze and Ranked Pairs are empirically very close. Across four voter models, three numbers of candidates, and three numbers of voters, they agreed on the winner in at least 92.8% of elections and never less. Both substantially outperform Borda count on manipulability and Condorcet efficiency. The new dropout stability metric reveals that Borda is competitive on robustness to ballot perturbation, which is a property orthogonal to standard axiomatic criteria. The overall conclusion is that small-committee elections with structured preferences are well-resolved by any reasonable Condorcet method, and that the choice between Schulze and Ranked Pairs should be made on grounds of explainability rather than expected outcome.

References

[1] Arrow, K. J. (1951). Social Choice and Individual Values. Yale University Press.

[2] Schulze, M. (2011). A new monotonic, clone-independent, reversal symmetric, and Condorcet-consistent single-winner election method. Social Choice and Welfare, 36(2), 267 to 303.

[3] Tideman, T. N. (1987). Independence of clones as a criterion for voting rules. Social Choice and Welfare, 4(3), 185 to 206.

[4] Gibbard, A. (1973). Manipulation of voting schemes: a general result. Econometrica, 41(4), 587 to 601.

[5] Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions. Journal of Economic Theory, 10(2), 187 to 217.

[6] Black, D. (1948). On the rationale of group decision-making. Journal of Political Economy, 56(1), 23 to 34.

[7] Mallows, C. L. (1957). Non-null ranking models. Biometrika, 44(1/2), 114 to 130.

[8] Tsetlin, I., Regenwetter, M., and Grofman, B. (2003). The impartial culture maximizes the probability of majority cycles. Social Choice and Welfare, 21(3), 387 to 398.