Measuring and Controlling Divisiveness in Rank Aggregation

Rachael Colley[1] , **Umberto Grandi**[1] , **César Hidalgo**[2,3,4] , **Mariana Macedo**[2] , **Carlos Navarrete**[2]

[1]IRIT, Université Toulouse Capitole, France

[2]Center for Collective Learning, ANITI, TSE, IAST, IRIT, Université de Toulouse, France

[3]Alliance Manchester Business School, University of Manchester, UK

[4]Center for Collective Learning, CIAS, Corvinus University, Hungary

{rachael.colley,umberto.grandi}@irit.fr, {cesar.hidalgo, mariana.macedo,
carlos.navarrete}@univ-toulouse.fr

Abstract

In rank aggregation, members of a population rank issues to decide which are collectively preferred. We focus instead on identifying divisive issues that express disagreements among the preferences of individuals. We analyse the properties of our divisiveness measures and their relation to existing notions of polarisation. We also study their robustness under incomplete preferences and algorithms for control and manipulation of divisiveness. Our results advance our understanding of how to quantify disagreements in collective decision-making.

1 Introduction

Rank aggregation is the problem of ordering a set of issues according to a set of individual rankings given as input. This problem has been studied extensively in computational social choice (see, e.g., Brandt *et al.* [2016]) when the rankings are assumed to represent human preferences over, for example, candidates in a political election, projects to be funded, or more generally alternative proposals. The most common approach in this literature is to find normative desiderata for the aggregation process, including computational requirements such as the existence of tractable algorithms for its calculation and characterisations of the aggregators that satisfy them. Rank aggregation also has a wide spectrum of applications from metasearch engines [Dwork *et al.*, 2001] to bioinformatics [Kolde *et al.*, 2012], receiving attention also in the statistical ML literature [Korba *et al.*, 2017].

Previous work on rank aggregation has focused on how to best elicit which issues are the most agreed upon, without identifying the issues that divide them. Instead, a wide literature in economics and the social sciences has developed measures of social, economic, and political polarisation. Classical work analysed polarisation in the distribution of wealth, goods, and opinions [Esteban and Ray, 1994; Duclos *et al.*, 2004], showing that well-studied notions of inequality are unfit to measure polarisation as they do not consider the weight of sub-populations. Another common approach in social science uses the variance of distributions to measure polarisation (see, e.g., Musco *et al.* [2018]; Gaitonde *et al.* [2020]).

In this paper, we put forward a family of functions that starting from a collection of individual rankings are able to order issues based on their divisiveness. We compute an issue's divisiveness by aggregating the disagreement among all possible sub-populations defined by the relative preference among the other issues. Our proposal relates to the literature on three important aspects. First, a parameter in our definition allows us to move from an adaptation of the classical measure of polarisation from Esteban and Ray [1994] at one end of the spectrum to the detection of disagreements from minorities at the other end. Second, our measures are parameterised by the rank aggregation function that is used to compute the most agreed-upon issues. In this way, we can align our notions of divisiveness with the functions chosen to measure the agreement of the population.[1] Third, while existing work focused on comparing different sets of rankings based on polarisation [Can *et al.*, 2015], diversity [Hashemi and Endriss, 2014], or cohesiveness [Alcantud *et al.*, 2015], here, we aim at identifying the most divisive issues *within* a complete profile of rankings. In doing so, we do not need to assume that issues are independent, as common in the social choice literature.

Our work can further guide how to query a population towards being more inclusive and unified, e.g., through deliberative instances. This can include measures that go towards decreasing divisiveness, such as recent work suggesting the construction of recommender systems to depolarise a population [Stray, 2022], or simply take advantage of this information when steering the public debate (recent work from Ash *et al.* [2017] suggests that politicians spend more time on divisive topics than on neutral ones). Our work also contributes to social choice theory, where related notions of preference diversity have shown to have effects on the probability of paradoxes [Gehrlein and Lepelley, 2010], the competitiveness in matching markets [Hałaburda, 2010], or the computational complexity of manipulating an election [Wu *et al.*, 2022]. Moreover, our work can be useful in refining the preference analysis of applications, such as online forums or surveys, that query a population on their opinions and return aggregated information about the group as a whole.

Contribution and Paper Structure. We extend the notion of divisiveness, introduced by Navarrete *et al.* [2022], to a family of measures which take into account the size

---

[1]This is particularly important in social choice applications, where the individual preferences collected are a (possibly strategic) response to the collective decision-making rule chosen.

of a sub-population and use the well-studied scoring functions. These extensions allow us to connect divisiveness to measures of polarisation. We give a theoretical and experimental analysis of our divisiveness measures by relating them to other notions and giving bounds on their limit cases (Section 3). Importantly, we show that our measures can distinguish between key profiles which other measures cannot. We then inspect two aspects of control, first, by studying the effect of removing pairwise comparisons from the agent's rankings (Section 4.2) and second, by adding additional controlled agents (Section 4.2). All our code is available at https://github.com/CenterForCollectiveLearning/divisiveness-theoretical-IJCAI2023.

Related Work. The notion of divisiveness studied in this paper builds on the work of Navarrete *et al.* [2022], who identify the most divisive issues from proposed government programs from crowdsourced political preferences. Many papers start from a profile of rankings and compare them based on how *consensual* (equivalently, *cohesive*) they are [Bosch, 2005; García-Lapresta and Pérez-Román, 2010; Alcalde-Unzu and Vorsatz, 2008, 2013], or in the opposite direction, i.e., how *diverse* is the set of preferences [Hashemi and Endriss, 2014; Karpov, 2017]. In particular, Alcantud *et al.* [2015] measures the cohesiveness of a group by aggregating the dissimilarity of their orderings (similarly to Can *et al.* [2015] who however focus on polarisation). Most of these settings are based on pairwise comparisons, except for Alcalde-Unzu and Vorsatz [2016] and Xue *et al.* [2020], who look at patterns of varying sizes in the rank profile. One of the methods proposed by Hashemi and Endriss [2014] to measure preference diversity is to average the distance between the individual rankings and the aggregated one. This is in line with our approach, but it only provides a measure to compare different populations of rankings without going to the level of single issues. We note that also an influential theory of diversity not based on preferences but on (binary) features of not necessarily independent alternatives has been proposed by Nehring and Puppe [2002].

2 Basic Definitions

This section introduces the basics of rank aggregation and scoring rules. We put forward our definition of divisiveness, then we compare it with existing notions of polarisation.

2.1 Preliminaries

Rankings. Let $\mathcal{I} = \{a, b \dots\}$ be a finite non-empty set of m issues. A strict ranking (aka. linear order) on $\mathcal{I}$ is an asymmetric, transitive, and complete binary relation on $\mathcal{I}$. We let $a \succ b$ denote the fact that a is strictly preferred to b in the ranking $\succ$. In what follows, we will write $\succ = acdb$ for $\succ = a \succ c \succ d \succ b$, reading preferences from left to right. The set of all strict rankings over $\mathcal{I}$ will be denoted by $\mathcal{L}(\mathcal{I})$. We denote with $rank(a, \succ)$ the rank of a in $\succ$ with the first position being 1 and the last being m.

Individual Rankings. Let a finite non-empty set of $\mathcal{N} = \{1, \dots, n\}$ agents express a strict rankings over $\mathcal{I}$ (sometimes referred to as preferences). We let $\mathcal{P} = (\succ_1, \dots, \succ_n)$ denote

the resulting profile of rankings, where $\succ_i$ is agent's i ranking over $\mathcal{I}$. We let $\mathcal{N}_{a \succ b} = \{i \in \mathcal{N} \mid a \succ_i b\}$ be the set of voters in $\mathcal{N}$ who prefer a to b. The restriction of profile $\mathcal{P}$ to the agents in $\mathcal{X}$ is denoted by $\mathcal{P}_{\mathcal{X}} = \langle \succ_i \mid i \in \mathcal{X} \rangle$. When $\mathcal{X} = \mathcal{N}_{a \succ b}$ we simply write $\mathcal{P}_{a \succ b}$. We call $\mathcal{P}$ a consensual profile if for all $i, j \in \mathcal{N}$ we have that $\succ_i = \succ_j$. Every preference profile $\mathcal{P}$ can be represented as a weighted (anonymous) profile, i.e., as a set of pairs $(w_j, \succ_j)$ indicating that $w_j \in \mathbb{N}$ agents have preference $\succ_j$.

Collective Scoring of Issues. Rank aggregation functions define a collective ranking of issues based on the agreements among the individual rankings in a profile. A large number of rules have been proposed and analysed in the literature on (computational) social choice and artificial intelligence. We focus on rank aggregators defined by a scoring function, where the collective ranking over issues is obtained via a function $s : \mathcal{I} \times \mathcal{L}(\mathcal{I}) \rightarrow [0, 1]$ that assigns a score to each issue in a given profile. Notable examples are the (normalised) Borda score, which counts the number of issues strictly preferred to a given issue, $Borda(a, \mathcal{P}) = \sum_{b \in \mathcal{I} \setminus \{a\}} \frac{\#(\mathcal{N}_{a > b})}{n \cdot (m-1)}$, where $\#(\mathcal{X})$ is the cardinality of a set $\mathcal{X}$. Or the normalised Copeland score, which counts the number of majority contests won by an issue, $Cop(a, \mathcal{P}) = \frac{\#\{b \in \mathcal{I} \setminus \{a\} \mid \#(\mathcal{N}_{a > b}) > \#(\mathcal{N}_{b > a})\}}{m-1}$.

2.2 Divisiveness

For a given sub-population $\mathcal{X} \subseteq \mathcal{N}$ and issue $a \in \mathcal{I}$, we measure the divisiveness of a for $\mathcal{X}$ as the difference between the collective scoring of a in sub-population $\mathcal{X}$ and in its complement sub-population $\mathcal{N} \setminus \mathcal{X}$.

Definition 1. *[Navarrete* et al.*, 2022] The divisiveness of an issue $a \in \mathcal{I}$ with respect to a sub-population $\mathcal{X} \subseteq \mathcal{N}$ in profile $\mathcal{P}$ is defined as:*

$$\text{DIV}^s(a, \mathcal{X}, \mathcal{P}) = |s(a, \mathcal{P}_{\mathcal{X}}) - s(a, \mathcal{P}_{\mathcal{N} \setminus \mathcal{X}})|.$$

If $\mathcal{X} = \emptyset$ or $\mathcal{X} = \mathcal{N}$, we set $\text{DIV}^{\mathcal{F}}(a, \mathcal{X}, \mathcal{P}) = 0$.

Examples of a sub-population $\mathcal{X}$ can be descriptive, such as agents living in cities (thus $\mathcal{N} \setminus \mathcal{X}$ are agents living in rural areas) or agents with a given political orientation (thus, $\mathcal{N} \setminus \mathcal{X}$ would be those who do not ascribe to this orientation). We can now give a definition of divisiveness for issue a that is independent of a given sub-population by averaging over all sub-populations $\mathcal{N}_{a \succ b}$ for all other issues b. We include an additional parameter α that allows us to take into consideration the size of a sub-population allowing for a weighted average version of an issue's divisiveness.

Definition 2. *The α-divisiveness $\text{DIV}^s_\alpha(a, \mathcal{P})$ of an issue $a \in \mathcal{I}$ in profile $\mathcal{P}$, with $\alpha \in [0, 1]$ and $\ell \in \mathbb{R}^+$, is defined as:*

$$\frac{1}{m-1} \sum_{\substack{b \in \mathcal{I} \\ a \neq b}} \left(\ell \frac{\#(\mathcal{N}_{a > b}) \cdot \#(\mathcal{N}_{b > a})}{n^2} \right)^\alpha \text{DIV}^s(a, \mathcal{N}_{a \succ b}, \mathcal{P}).$$

When $\alpha = 0$ and $s = Borda$, our definition is a reformulation of the divisiveness measure from Navarrete *et al.* [2022]. When $\alpha = 1$, Definition 2 can be interpreted as one of the polarisation measures of Duclos *et al.* [2004] and Esteban and Ray [1994], calculated on the distribution of ranks

that issue a received from individuals. We refer to the multiplicative factor of the measure defined in Definition 2 in each summand as the α-factor. The α-factor is maximal when $\#(\mathcal{N}_{a \succ b}) = \#(\mathcal{N}_{b \succ a}) = n/2$, and we will often set $\ell = 4$ so that the α-factor is at most 1. ℓ is a normalising factor left open in line with Esteban and Ray [1994].

Intuitively, as α increases in Definition 2, the relevance of the size of the disagreeing sub-population increases. The following example presents the limit case of $\alpha = 0$ where the size of disagreeing sub-populations is ignored, resulting in a different divisiveness ranking than $\alpha = 1$.

Example 1. *Consider* $2k$ *agents giving their preferences in profile* $\mathcal{P}$ *over issues* $\mathcal{I} = \{a, b, c, d, e, f\}$ *as such:*

$\succ_1$ *(1 agent)*	$cadfeb$
$\succ_2$ *(k − 1 agents)*	$badfec$
$\succ_3$ *(k agents)*	$bedfac$

Assume now that $k = 5$ *and* $\ell = 4$ *(the normalising factor). Table 1 presents, for issue* a*, the table of disagreements (in terms of Borda score difference) between the 5 possible subpopulations defined by the pairwise comparisons of* a *with the remaining issues.*

Issue x	$s(\mathcal{P}_{a \succ x})$	$s(\mathcal{P}_{x \succ a})$	disagreement	α-factor
b	0.8	$0.4\overline{6}$	$0.\overline{3}$	0.36
c	$0.4\overline{6}$	0.8	$0.\overline{3}$	0.36
d	0.8	0.2	0.6	1
e	0.8	0.2	0.6	1
f	0.8	0.2	0.6	1

Table 1: Details of the calculations for divisiveness for issue a, setting $\ell = 4$ and $k = 5$ and with $s = Borda$.

From Table 1, we can compute $\text{DIV}_0^{Borda}(a, \mathcal{P})$ *by averaging the disagreements and weighting by the* α*-factors to obtain* $\text{DIV}_1^{Borda}(a, \mathcal{P})$*. Repeating this process for every issue gives us the values of divisiveness in Table 2.*

x	$Borda(x, \mathcal{P})$	$\text{DIV}_0^{Borda}(x, \mathcal{P})$	$\text{DIV}_1^{Borda}(x, \mathcal{P})$
a	0.5	$0.49\overline{3}$	0.408
b	0.9	1	0.36
c	0.1	1	0.36
d	0.6	0	0
e	0.5	$0.49\overline{3}$	0.408
f	0.4	0	0

Table 2: Normalised Borda score and divisiveness for $\alpha = 0, 1$ of issues in the profile of Example 1 with $k = 5$ and $\ell = 4$.

The most divisive issue for DIV_0^{Borda} *is* b *and* c*, who are at the opposite extreme of the ordering for the first agent with respect to the rest of the population, while the most divisive for* DIV_1^{Borda} *are* a *and* e*, who are second or second-to-last for all agents, yet this divides the population into two. Observe that this holds for any* $k \geq 5$*, showing a class of examples where the ranking of* α*-divisiveness differs significantly depending on if* $\alpha = 0$ *or* $\alpha = 1$*.*

2.3 Rank-variance

Related literature in the social sciences often measures polarisation using notions of variance, which we now adapt to rankings. Let $\mu_{\mathcal{P}}(a) = \frac{1}{n} \sum_{i \in \mathcal{N}} rank(a, \succ_i)$ be the average rank of an issue $a \in \mathcal{I}$ in a profile of rankings $\mathcal{P}$. The *rank-variance* of issue a is defined as follows:

$$Var(a, \mathcal{P}) = \frac{1}{n} \sum_{i \in \mathcal{N}} (rank(a, \succ_i) - \mu_{\mathcal{P}}(a))^2$$

The following example shows a profile in which the ranking of issues by variance differs from divisiveness (assuming $\alpha = 0$ and $s = Borda$).

Example 2. *Consider the following preference profile:*

$\succ_1$ *(10 agents)*	$abcde$
$\succ_2$ *(10 agents)*	$ebcda$
$\succ_3$ *(1 agent)*	$acbde$
$\succ_4$ *(1 agent)*	$ebdca$

The rank-variance and the divisiveness using Borda and $\alpha = 0$ *for each issue is the following:*

	a	b	c	d	e
$Var(x, \mathcal{P})$	4	$0.0\overline{45}$	$0.0\overline{90}$	$0.0\overline{45}$	4
$\text{DIV}^{Borda}(x)$	1	0.074	0.037	0.074	1

This shows that issue c *has higher variance than issues* b *and* d *but lower divisiveness (Kendall's tau correlation between the two rankings is* $\tau \approx 0.5$*).*

Unlike other notions of polarisation, we focus on the comparisons between our divisiveness measures and rank-variance as they both return information about a single issue rather than about the population as a whole.

3 Measuring Divisiveness and Polarisation

In this section, we present some basic properties of our definition of divisiveness, showing in particular that with $\alpha = 0$ it does not coincide nor is correlated with standard notions of polarisation. We also show how our definitions can be used to identify the sub-population that is most divided on an issue.

3.1 Divisiveness Bounds

We first observe that if s is a polynomially computable function, then so is DIV_α^s for any α. Moreover, if s is anonymous and neutral (as in the classical social choice terminology[2]) so is DIV_α^s for any α. A direct consequence of our definitions is that $0 \leq \text{DIV}_\alpha^s(a, \mathcal{P}) \leq 1$, for any α. We now characterise the extremes of the spectrum.

First we give the sufficient conditions for minimal divisiveness with the Borda and Copeland scorings. Let a profile $\mathcal{P}$ be rank-unanimous on a if for all $i, j \in \mathcal{N}$ we have that $rank(a, \succ_i) = rank(a, \succ_j)$. Profile $\mathcal{P}$ is unanimous if $\succ_i = \succ_j$ for all $i, j \in \mathcal{N}$. The next result shows that divisiveness is minimal when consensus is maximal, following this, we will discuss the converse of this statement in a few different ways.

---

[2] A function taking profiles of linear orders as input is anonymous if permuting the individual rankings does not change the result. It is neutral if all issues are treated equally, i.e., permuting the name of the issues results in the ranking obtained by applying the same name permutation to the previous result.

Proposition 1. *If profile $\mathcal{P}$ is rank-unanimous on a then* $\mathrm{DIV}_\alpha^{Borda}(a,\mathcal{P}) = 0$, *while not necessarily true for* $\mathrm{DIV}_\alpha^{Cop}$. *If $\mathcal{P}$ is unanimous, then* $\mathrm{DIV}_\alpha^{Cop}(a,\mathcal{P}) = 0$ *for all $a \in \mathcal{I}$.*

Proof. If $\mathcal{P}$ is rank-unanimous on a, then for all $b \in \mathcal{I}\setminus\{a\}$ we have that $Borda(a,\mathcal{P}_{a\succ b}) - Borda(a,\mathcal{P}_{b\succ a}) = 0$, and thus $\mathrm{DIV}_\alpha^s(a,\mathcal{P}) = 0$. This does not necessarily hold for DIV^{Cop}: consider 3 agents with the following rankings over issues $\mathcal{I} = \{a,b,c,d\}$, forming profile $\mathcal{P}$: $\succ_1 = abcd$, $\succ_2 = cbad$, and $\succ_3 = dbac$. Observe that for all $i \in \mathcal{N}$ that $rank(b,\succ_i) = 2$, hence $\mathcal{P}$ is rank-unanimous on b. However, $\mathrm{DIV}_0^{Cop}(b,\mathcal{P}) = \frac{1}{3}$. Finally, if $\mathcal{P}$ is unanimous, any notion of divisiveness will be equal to zero, as either $\mathcal{N}_{a\succ b}$ or $\mathcal{N}_{b\succ a}$ is empty for any pair $a,b \in \mathcal{I}$. $\square$

Profile $\mathcal{P}$ is *fully polarised* if it is split into two equally-sized sub-populations with completely opposite preferences (m is even). If n is even, $\mathcal{P}$ is fully polarised if $\succ_i = \succ_1$ for $i = 1,\ldots,n/2$, and $\succ_i = \succ_2$ for $i = n/2+1,\ldots,n$, where $rank(a,\succ_1) = m - rank(a,\succ_2)$ for all $a \in \mathcal{I}$. If n is odd, one of the sub-populations has one more agent than the other. Let a be the top-ranked issue in $\succ_1$ (hence ranked last in $\succ_2$). We show that divisiveness is maximal in such profiles:

Proposition 2. *If $\mathcal{P}$ is fully polarised and n is even, then* $\mathrm{DIV}_\alpha^{Borda}(a,\mathcal{P}) = 1$ *when $\ell = 4$, where a is the top-ranked issue in one of the two sub-populations. If $\mathcal{P}$ is fully polarised and n is odd, then* $\mathrm{DIV}_\alpha^{Cop}(a,\mathcal{P})$ *equals the α-factor.*

Proof. If $\mathcal{P}$ is fully polarised and n is even, then each summand of $\mathrm{DIV}_\alpha^{Borda}(a,\mathcal{P})$ is 1, as all agents in $\mathcal{N}_{a\succ b}$ rank a first, all in $\mathcal{N}_{b\succ a}$ rank a last, and $\#(\mathcal{N}_{a\succ b})=\#(\mathcal{N}_{b\succ a})$. Thus, $\mathrm{DIV}_\alpha^{Borda}(a,\mathcal{P})=1$. If $\mathcal{P}$ is fully polarised and n is odd, then a is a Condorcet winner in the larger sub-population and a Condorcet loser in the other. Thus, $\mathrm{DIV}^s(a,\mathcal{N}_{a>b},\mathcal{P})=1$ for each $b \in \mathcal{I}\setminus\{a\}$. As sub-populations differ by 1, $\mathrm{DIV}_\alpha^{Cop}(a,\mathcal{P}) = \left(\frac{n^2-1}{n^2}\right)^\alpha$ when $\ell = 4$. $\square$

In fully polarised profiles, two issues have maximal divisiveness: the top-ranked issues in $\succ_1$ and $\succ_2$. Moreover, in any profile, at most two issues can have maximal divisiveness.

Finally, *uniform* profiles contain each of the $m!$ possible rankings over the m issues, and each ranking is equally represented in the profile (note that n is even). They represent a fully noisy population of preferences. While the measure of polarisation proposed by Can *et al.* [2015] cannot distinguish between uniform and fully polarised profiles, we show that the ranking of divisiveness is strict in the former while in the latter all issues have the same divisiveness. The following proposition is straightforward due to the symmetry of uniform profiles:

Proposition 3. *If $\mathcal{P}$ is a uniform profile, then* $\mathrm{DIV}_\alpha^s(a,\mathcal{P}) = \mathrm{DIV}_\alpha^s(a,\mathcal{P})$ *for all $a,b \in \mathcal{I}$.*

For a uniform profile $\mathcal{P}$, $\mathrm{DIV}_\alpha^{Borda}(a,\mathcal{P}) = \frac{1}{m-1}$ when $\ell = 4$, as the average Borda score between two sub-populations $N_{a\succ b}$ and $\mathcal{N}_{b\succ a}$ differs by $1/m-1$. Whereas $\mathrm{DIV}_\alpha^{Cop}(a,\mathcal{P}) = 1$ when $\ell = 4$, as the divide of the two sub-populations always ensures that a always wins every majority contest in $\mathcal{N}_{a\succ b}$ and always loses in $\mathcal{N}_{b\succ a}$.
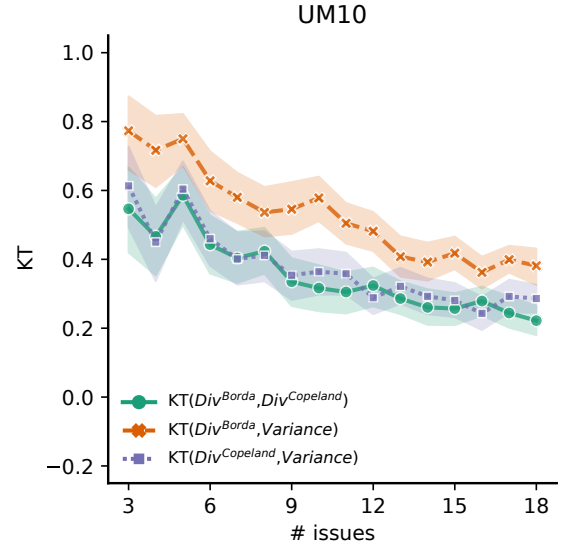

Figure 1: The average Kendall's tau correlation between each pair of divisiveness rankings using Borda and Copeland with $\alpha = 0$, and the rank variance, varying the number of issues. The average is taken from 100 profiles with 100 agents generated by UM10.

3.2 Divisiveness and Rank Variance

We conducted experiments on synthetic preference profiles to test whether divisiveness with $\alpha = 0$ correlates with the rank-variance defined in Section 2.3. We computed Kendall's tau correlation (KT) of DIV_0^{Borda} and DIV_0^{Cop} with the rank variance (cf. Section 2.3). We tested 100 profiles of rankings generated via the impartial culture (IC) and the Urn model with a correlation of 10% and 50% (named UM10 and UM50, respectively). Rankings were generated using the PrefLib library [Mattei and Walsh, 2017, 2013].

Using the Urn model with 10% correlation as an example, we plot in Figure 1 the average Kendall's tau correlation. We observe that the three measures are correlated when profiles are on a few issues but that the values of correlation decrease significantly as we increase the number of issues, and therefore increase the possible rankings for the measures to return. The correlation between the rank variance and the divisiveness computed using Borda is higher than using Copeland. The correlation between divisiveness using Borda and divisiveness using Copeland shows a similar decreasing trend. Similar results are captured for the impartial culture scenario, but the correlation decreases even more for the case of $KT(\mathrm{DIV}_0^{Borda}, \mathrm{DIV}_0^{Cop})$ and $KT(\mathrm{DIV}_0^{Cop}, Variance)$. A possible explanation of this decreasing correlation is that rank-variance only considers an issue's position in each individual ranking discarding which the rankings' structures (which are ranked above or below the issue in question).

3.3 Maximally Divided Sub-Populations

In addition to providing a ranking of issues based on their divisiveness, Definition 2 can also be used to identify the partition that maximally divides the population for an issue. This is a seemingly hard computational problem, as there is an

exponential number of sub-populations to consider, but we show that it can be solved efficiently for the Borda score.

Proposition 4. *For any profile $\mathcal{P}$ and issue $a \in \mathcal{I}$, finding the sub-population $\mathcal{X} \subseteq \mathcal{N}$ that maximises $\text{DIV}_0^{Borda}(a, \mathcal{X}, \mathcal{P})$ can be done in polynomial time.*

Proof. Consider an arbitrary preference profile $\mathcal{P}$. For an arbitrary issue $a \in \mathcal{I}$, we will use the following algorithm to find the sub-population $\mathcal{X}$ such that $\text{DIV}_0^{Borda}(a, \mathcal{X}, \mathcal{P})$ is maximal. We first order the agents with respect to their ranking of issue a. Thus, without loss of generality, we can assume that for each $i \in [1, n-1]$ we have that $rank(a, \succ_i) \geq rank(a, \succ_{i+1})$. Note that if two agents rank a at the same level, their ordering is irrelevant. We will now prove that any sub-population $\mathcal{X}$ that gives the maximum value of $\text{DIV}^{Borda}(a, \mathcal{X}, \mathcal{P})$ will partition $\mathcal{N}$ such that for some $k \in [1, n-1]$ we have that $\mathcal{X} = \{1, \cdots k\}$ and thus $\mathcal{N} \backslash \mathcal{X} = \{k+1, \cdots, n\}$. Our polynomial algorithm then tests each of the $n-1$ partitions defined by $\mathcal{X}_k = \{1, \cdots, k\}$ and returns the one that maximises $\text{DIV}_0^{Borda}(a, \mathcal{X}_k, \mathcal{P})$. Clearly, in each of these partitions $\mathcal{X}_k$, calculating $\text{DIV}_0^{Borda}(a, \mathcal{X}_k, \mathcal{P})$ can be done in polynomial time.

We now prove that any $\mathcal{X}$ that maximises $\text{DIV}_0^{Borda}(a, \mathcal{X}, \mathcal{P})$ is of the form $\mathcal{X}_k = \{1, \cdots, k\}$ for some k. To do so, we consider some $\mathcal{X}$ such that there exists some $i \in \mathcal{X}$ yet $i - z \notin \mathcal{X}$, for $z \in [1, i-1]$. We need to show that the set $\mathcal{X}' = (\mathcal{X} \backslash \{i\}) \cup \{i - z\}$ is more divisive than $\mathcal{X}$, thus $\text{DIV}_0^{Borda}(a, \mathcal{X}', \mathcal{P}) \geq \text{DIV}_0^{Borda}(a, \mathcal{X}, \mathcal{P})$. It is clear that $Borda(a, \mathcal{P}_\mathcal{X}) \leq Borda(a, \mathcal{P}_{\mathcal{X}'})$ due to our assumption on the ordering of the agents in $\mathcal{P}$ in decreasing order of rank of a. For similar reasons, we have that $Borda(a, \mathcal{P}_{\mathcal{N} \backslash \mathcal{X}}) \geq Borda(a, \mathcal{P}_{\mathcal{N} \backslash \mathcal{X}'})$. Therefore, it must be the case that $\text{DIV}_0^{Borda}(a, \mathcal{X}', \mathcal{P}) \geq \text{DIV}_0^{Borda}(a, \mathcal{X}, \mathcal{P})$. Thus, we can transform any $\mathcal{X}$ that maximises divisiveness into a sub-population of the form $\mathcal{X}_k$ by performing a finite number of swaps, as described above. $\square$

Determining the complexity of finding maximally divided sub-populations under DIV_0^{Cop} does not seem trivial, and we leave it as an open problem.

4 Divisiveness Control

Manipulation and control have been widely studied for rank aggregation procedures. Manipulation is when an individual misrepresents their reported ranking to improve the score of their favourite candidate. A classical result showed that manipulation can be performed in polynomial time for the Borda score [Bartholdi *et al.*, 1989]. Instead, control is when an external agent aims at altering the score of a designated candidate by performing actions such as removing agents or candidates from the profile of ranking. To give an example, preventing a candidate from being the Copeland winner by adding new rankings was shown to be a computationally hard problem. For an introduction to the computational complexity of these problems, we refer to the survey from Faliszewski and Rothe [2016]. In this section, we focus on two approaches to control the measure of divisiveness: (i) by removing pairwise comparisons from the agents' rankings and
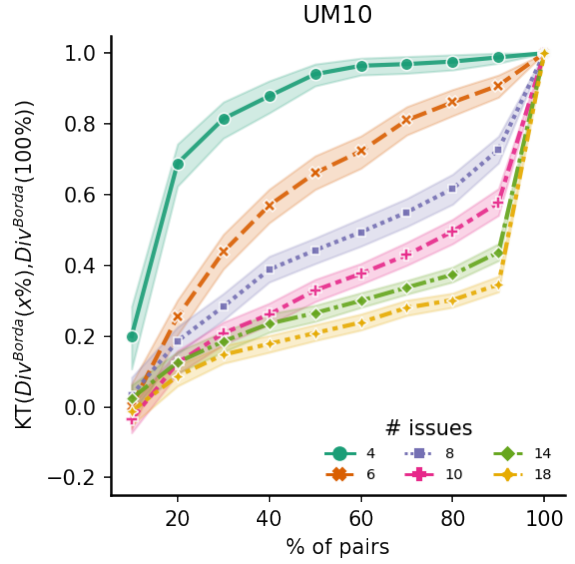

Figure 2: The average Kendall's tau correlation between the divisiveness ranking under DIV_0^{Borda} computed on the full ranking with respect to incomplete rankings. The horizontal axis shows the percentage of pairwise comparisons of the incomplete profile with respect to the complete one. The experiment is run on 100 preference profiles drawn from UM10. The different markers on the lines represent different numbers of issues in $\{4, 6, 8, 10, 14, 18\}$

(ii) by adding new agents (which could, e.g., be performed by *bots* on any platform that crowdsources individual preferences). This section focuses on the less studied divisiveness measure with $\alpha = 0$ and hence we omit α from DIV_α^s.

4.1 Removing Pairwise Comparisons

This section studies the disruption of the divisiveness measure by the deletion of pairwise comparisons from the rankings. This can be thought of as sabotage, as the control actions here do not have a clear goal, e.g., making a single issue the most divisive one. Instead, the aim of this control problem is to disrupt the divisiveness ranking such that it no longer resembles the ranking under complete preferences. To do so, we evaluate through simulations what percentage of pairwise comparisons of the agents' full rankings are required to be able to compute the divisiveness measure accurately.

As we are removing parts of the rankings given by the agents, we need to compute divisiveness on incomplete rankings as in the original definition by Navarrete *et al.* [2022]. When $s = Cop$, we see that the definition of divisiveness is well-defined on incomplete rankings. However, on incomplete rankings, we use the win rate instead of $Borda$ when calculating the divisiveness, noting that they are equivalent on complete rankings. We define the win rate of an issue a to be $\sum_{b \in \mathcal{I} \backslash \{a\}} \frac{\#(\mathcal{N}_{a > b})}{\#(\mathcal{N}_{a > b} \cup \mathcal{N}_{b > a}) \cdot (m-1)}$.

We generated 100 profiles for each of the three preference generation methods IC, UM10, and UM50, varying the number of issues $m \in [3, 18]$. We compared the average Kendall's tau correlation between the divisiveness ranking computed on the full rankings and the adapted measure of divisive-

ness computed on sub-profiles containing $X\%$ of the pairwise comparisons of the complete one ($X \in [10, 100]$ increasing by increments of 10). Here, pairwise comparisons are deleted from the profile at random.

The main message of this simulation is that disrupting divisiveness by deleting pairwise comparisons is easy. Figure 2 focuses on the case when the average KT is taken over 100 profiles with 100 agents created via the UM10 method. It shows that if there are sufficiently many issues (say more than 10), then deleting between 10 and 20% of the pairwise comparisons in the profile is sufficient to significantly decrease the accuracy of divisiveness (the correlation between complete and incomplete divisiveness is below 0.5). We also observe an inversion of the curves when the number of issues exceeds 6, tending towards an exponential shape. Our findings imply that with a large number of issues and under the assumption of moderately correlated preferences, the almost-totality of the pairwise comparisons needs to be elicited from agents to obtain an accurate measure of divisiveness.

4.2 Control by Adding Rankings

The next form of control we study is the addition of fake rankings by an external agent. Modelling this type of control is particularly realistic when the divisiveness measures are used in online forums, where attacks by bots are commonplace (such as in the experimental setting of Navarrete *et al.* [2022]). Similar problems were previously studied in the literature, such as Sybil attacks in online elections [Meir *et al.*, 2022]. We start by presenting an example in which a single agent is able to alter the divisiveness ranking.

Example 3. *Consider four agents and four issues $\mathcal{I} = \{a, b, c, d\}$. Consider that three agents have submitted their preferences, one agent with $\succ_1$, and two with $\succ_2$. The fourth agent has the truthful preference of $\succ_3$:*

$\succ_1$ *(1 agent)*	$bcad$
$\succ_2$ *(2 agents)*	$abcd$
$\succ_3$ *(1 agent)*	$acbd$

If the fourth agent submits their truthful preference, we have that: $\text{DIV}^{Borda}(a, \mathcal{P}) = {}^{12}/_{27}$, $\text{DIV}^{Borda}(b, \mathcal{P}) = {}^{7}/_{27}$, $\text{DIV}^{Borda}(c, \mathcal{P}) = {}^{8}/_{27}$, *and* $\text{DIV}^{Borda}(d, \mathcal{P}) = 0$. *Thus, a is currently the most divisive issue. As all agents agree that d is the worst of all the issues, it is the least divisive issue.*

If the agent with truthful preference $\succ_3$ wants to manipulate the divisiveness measure in order to make the issue b more divisive, then they can submit a preference $\succ'_3 = cadb$ giving profile $\mathcal{P}' = (\succ_1, (2, \succ_2), \succ'_3)$. In doing so, the measure of divisiveness changes as such: $\text{DIV}^{Borda}(a, \mathcal{P}') = {}^{19}/_{54}$, $\text{DIV}^{Borda}(b, \mathcal{P}') = {}^{38}/_{54}$, $\text{DIV}^{Borda}(c, \mathcal{P}') = {}^{19}/_{54}$, *and* $\text{DIV}^{Borda}(d, \mathcal{P}') = {}^{6}/_{54}$. *Thus, by submitting $\succ'_3$, the agent succeeded in making b the most divisive issue.*

Example 3 shows that one agent can manipulate the divisiveness measure. We now show that the problem of control by adding rankings can be solved easily by showing a simple heuristic, which we call INJECT$_s$.

INJECT$_s$ takes as input a profile $\mathcal{P}$ and an issue a which it aims to make the most divisive by adding new agents to the profile. It first computes the ranking over issues defined by $s(x, \mathcal{P})$, which we denote by $\succ_s$. This is used to create the two rankings which will be added to the profile to increase the divisiveness of a, namely $\succ_{odd}$ and $\succ_{even}$. The former modifies $\succ_s$ by putting issue a first and leaving the rest of the ranking unchanged. The latter modifies $\succ_s$ symmetrically by putting a in the last position and leaving the remaining part of the order unchanged. In this way, $\succ_{odd}$ and $\succ_{even}$ resemble the ranking of agreements computed using s, with the only difference being the position of a that alternates between first and last. INJECT$_s$ then alternates between adding $\succ_{odd}$ and $\succ_{even}$ to profile $\mathcal{P}$ until the issue a is the most divisive. We show that INJECT$_s$ always terminates and succeeds in making the target issue the most divisive.

Proposition 5. INJECT$_{Borda}$ *always terminates for $\alpha = 0$.*

Proof. Let a be the target issue. If a is already the most divisive issue in $\mathcal{P}$ then INJECT$_{Borda}$ terminates immediately. Else, we first prove that $\text{DIV}^{Borda}(a, \mathcal{P}') > \text{DIV}^{Borda}(a, \mathcal{P})$ for any profile $\mathcal{P}'$ obtained by using INJECT$_{Borda}$ on a profile $\mathcal{P}$. Take an arbitrary profile $\mathcal{P}$, issue $a \in \mathcal{I}$, and k such that $\mathcal{P}' = (\mathcal{P}, (k, \succ_{odd}), (k, \succ_{even}))$, where $\mathcal{N}'$ are the agents in $\mathcal{P}'$. Take an arbitrary $b \in \mathcal{I} \backslash \{a\}$, we have that $\text{DIV}^{Borda}(a, \mathcal{N}'_{a \succ b}, \mathcal{P}') > \text{DIV}^{Borda}(a, \mathcal{N}_{a \succ b}, \mathcal{P})$ as $Borda(a, \mathcal{P}'_{a \succ b}) \geq Borda(a, \mathcal{P}_{a \succ b})$ and $Borda(a, \mathcal{P}_{b \succ a}) \geq Borda(a, \mathcal{P}'_{b \succ a})$, with at least one of the two inequalities being strict. To see this, observe that all injected agents rank a the highest in $\mathcal{P}'_{a \succ b}$ and the lowest in $\mathcal{P}'_{b \succ a}$ (and a is not the most divisive issue). Thus, we proved that $\text{DIV}^{Borda}(a, \mathcal{P}')$ tends to 1 as k increases. To conclude, note that the rank of any issue $b \in \mathcal{I} \backslash \{a\}$ in the injected sub-profile $((k, \succ_{even}), (k, \succ_{odd}))$ varies only by one position. Thus, with k large enough the divisiveness of any issue $b \neq a$ cannot tend to 1. $\square$

Proposition 6. INJECT$_{Cop}$ *always terminates in polynomial time for $\alpha = 0$.*

Proof. Let $k = 2n + 2$ and a be the target issue, where n is the number of voters in $\mathcal{P}$. By definition, $\text{DIV}_0^{Cop}(a, \mathcal{P}') = {}^{1}/_{m-1} \sum_{b \in \mathcal{I} \backslash \{a\}} |Cop(a, \mathcal{P}'_{a \succ b}) - Cop(a, \mathcal{P}'_{a \succ b})|$. As k is sufficiently large, for all $b \in \mathcal{I} \backslash \{a\}$ we have that a is a Condorcet winner in $\mathcal{P}'_{a \succ b}$, since $n + 1$ copies of $\succ_{odd}$ were added with a as the top issue. Symmetrically, a is a Condorcet loser in $\mathcal{P}'_{b \succ a}$. Thus, we have that $Cop(a, \mathcal{P}_{\mathcal{N}_{a \succ b}}) = 1$ and $Cop(a, \mathcal{P}_{\mathcal{N}_{b \succ a}}) = 0$, which in turn implies that $\text{DIV}_0^{Cop}(a, \mathcal{P}') = 1$. By the uniqueness of a Condorcet winner, we have that no other issue $b \neq a$ can have $\text{DIV}_0^{Cop}(b, \mathcal{P}') = 1$, concluding the proof. $\square$

The previous result shows that INJECT$_s$ can manipulate the divisiveness ranking. However, it does not provide a bound on how many agents INJECT$_{Borda}$ are required and only provides a large bound for INJECT$_{Cop}$, namely $k = 2n + 2$. To complement this, we conducted simulations to estimate how many new agents INJECT$_s$ needs to alter the divisiveness ranking. For each $m \in [2, 11]$ and each of three profile generation methods (IC, UM10, UM50), we considered 100 profiles to test how many new agents INJECT$_s$ required to make
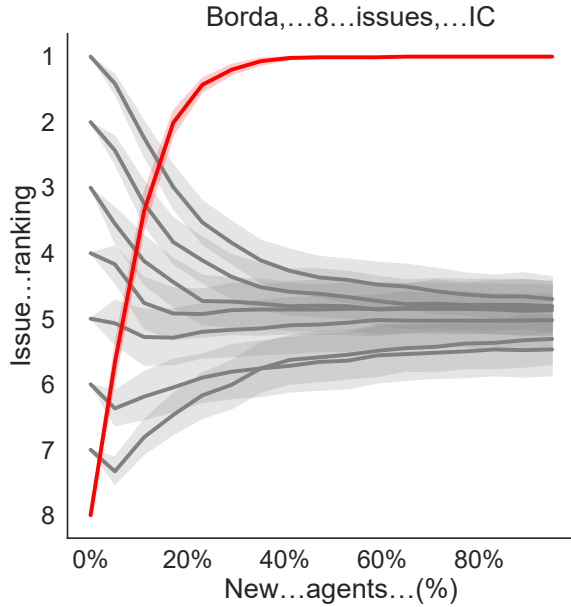
Figure 3: Percentage of new agents added by INJECT$_{Borda}$ to make the least divisive issue the most divisive. The average divisiveness of the other issues is also plotted in grey. Averages are computed over 100 IC profiles of 100 agents and 8 issues.

Figure 4: Percentage of agents added by INJECT$_{Borda}$ with 8 issues, to make the 2nd, 4th or 8th most divisive issue the most. The average is taken over 100 profiles for 100 agents generated using IC and UM50 (represented by a solid or dashed line, respectively).

the target issue the most divisive. Figure 3 focuses on IC profiles with 8 issues. It shows the divisiveness rankings of the 8 issues in the initial profile (at 0%) and their evolution when INJECT$_{Borda}$ inserts additional rankings to make the least divisive issue the most divisive (the highlighted line at the 8th position). In particular, by adding around 35% new agents we can make the least divisive issue the most via our simple algorithm. In crowdsourcing applications with wide participation this percentage implies that too many additional agents might need to be injected without being noticed. Yet, we recall that INJECT$_s$ is a simple heuristic, and this percentage could be lower with a more efficient procedure. Furthermore, we see that the average divisiveness ranking of the other issues converges to the middle of the ranking. We obtained similar results by varying the number of issues and the profile generation methods.

We also tested how many additional agents INJECT$_{Borda}$ required to reach easier targets, such as making either the second most divisive issue or an issue in the middle of the divisiveness ranking the most divisive issue. Figure 4 presents our findings for $s = Borda$ and $m = 8$, computed on 100 profiles generated using either IC or UM50. Clearly, if the task of control is harder, INJECT$_{Borda}$ needs additional agents to meet its target. More importantly, if we compare the performance of INJECT$_{Borda}$ on different preference generation methods, we see that more correlated profiles (using UM50 in our case) are harder to control no matter the target issue.

Given the results, we see that INJECT$_s$ can be an effective way of manipulating, but it simulates a static scenario. We also note that this way of manipulating requires very little information about the original profile, i.e., the controlling agent just need to know the current ranking of agreement given by
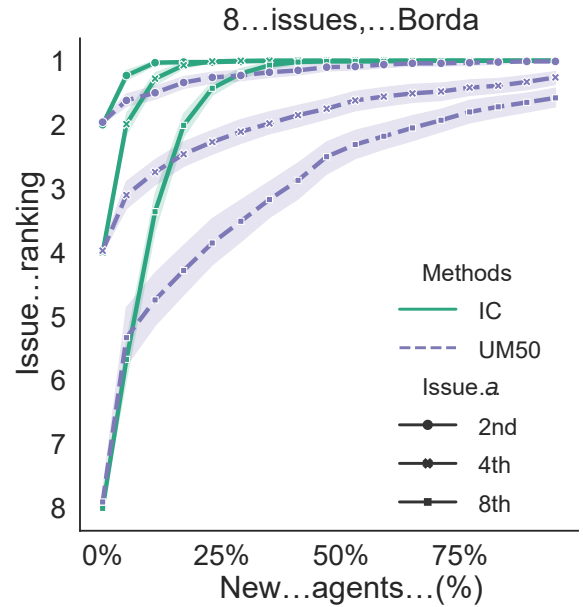
the scoring s. Furthermore, our results show that INJECT$_s$ can be used just to increase an issue's rank in the divisiveness ranking, rather than insisting that it becomes the most divisive.

5 Conclusions and Future Work

This paper extends the notion of divisiveness given by Navarrete *et al.* [2022] to a family of measures and applies them to complete rankings over issues. We ground these measures by highlighting their behaviour at limit cases and comparing them to other notions of disagreement and polarisation. We also point out how we can find a sub-population for which an issue is most divisive in polynomial time when considering the Borda score, yet, no such algorithm was found for the Copeland score. The main contribution of this paper is the study of the robustness of the divisiveness measures to external control. We showed via simulations that by randomly removing pairwise comparisons from the rankings, the correlation between the divisiveness ranking of the full vs partial rankings can drop significantly, especially when there are many issues. Furthermore, we show that a simple algorithm can affect the divisiveness ranking by inserting (a possibly large number of) controlled fake rankings.

This paper opens many directions for future work. First, how can our divisiveness measures be modified to be more robust to external control. Second, our divisiveness measures can be used to compare how divisive or polarised is a given population (instead of focusing on comparisons of single issues). Finally, following in the social choice theory tradition, we will explore axiomatic characterisations of divisiveness.

Acknowledgments

References

Jorge Alcalde-Unzu and Marc Vorsatz. The measurement of consensus: an axiomatic analysis. *FEDEA (Working Paper)*, 2008.

Jorge Alcalde-Unzu and Marc Vorsatz. Measuring the cohesiveness of preferences: an axiomatic analysis. *Social Choice and Welfare*, 41(4):965–988, 2013.

Jorge Alcalde-Unzu and Marc Vorsatz. Do we agree? measuring the cohesiveness of preferences. *Theory and Decision*, 80(2):313–339, 2016.

José Carlos Rodríguez Alcantud, R de Andrés Calle, and JM Cascón. Pairwise dichotomous cohesiveness measures. *Group Decision and Negotiation*, 24(5):833–854, 2015.

Elliott Ash, Massimo Morelli, and Richard Van Weelden. Elections and divisiveness: Theory and evidence. *The Journal of Politics*, 79(4):1268–1285, 2017.

John J Bartholdi, Craig A Tovey, and Michael A Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

Rob Bosch. *Characterizations of Voting Rules and Consensus Measures*. PhD thesis, Tilburg University, 2005.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

Burak Can, Ali Ihsan Ozkes, and Ton Storcken. Measuring polarization in preferences. *Mathematical Social Sciences*, 78:76–79, 2015.

Jean-Yves Duclos, Joan Esteban, and Debraj Ray. Polarization: concepts, measurement, estimation. *Econometrica*, 72(6):1737–1772, 2004.

Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web (WWW)*, 2001.

Joan-Maria Esteban and Debraj Ray. On the measurement of polarization. *Econometrica: Journal of the Econometric Society*, pages 819–851, 1994.

Piotr Faliszewski and Jörg Rothe. Control and bribery in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

Jason Gaitonde, Jon Kleinberg, and Éva Tardos. Adversarial perturbations of opinion dynamics in networks. In *Proceedings of the 21st ACM Conference on Economics and Computation (EC)*, 2020.

José Luis García-Lapresta and David Pérez-Román. Consensus measures generated by weighted Kemeny distances on weak orders. In *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2010.

William V. Gehrlein and Dominique Lepelley. *Voting paradoxes and group coherence: the Condorcet efficiency of voting rules*. Springer Science & Business Media, 2010.

Hanna Hałaburda. Unravelling in two-sided matching markets and similarity of preferences. *Games and Economic Behavior*, 69(2):365–393, 2010.

Vahid Hashemi and Ulle Endriss. Measuring diversity of preferences in a group. In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI)*, 2014.

Alexander Karpov. Preference diversity orderings. *Group Decision and Negotiation*, 26(4):753–774, 2017.

Raivo Kolde, Sven Laur, Priit Adler, and Jaak Vilo. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*, 28(4):573–580, 2012.

Anna Korba, Stephan Clémençon, and Eric Sibony. A learning theory of ranking aggregation. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

Nicholas Mattei and Toby Walsh. Preflib: A library of preference data HTTP://PREFLIB.ORG. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT)*, 2013.

Nicholas Mattei and Toby Walsh. A PREFLIB.ORG retrospective: Lessons learned and new directions. *Trends in Computational Social Choice. AI Access Foundation*, pages 289–309, 2017.

Reshef Meir, Nimrod Talmon, Gal Shahaf, and Ehud Shapiro. Sybil-resilient social choice with low voter turnout. In *Proceedings of the 19th European Conference on Multi-Agent Systems (EUMAS)*, 2022.

Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 world wide web conference (WWW)*, 2018.

Carlos Navarrete, Nicole Ferrada, Mariana Macedo, Rachael Colley, Jingling Zhang, Umberto Grandi, Jerome Lang, and César A Hidalgo. Understanding political agreements and disagreements: Evidence from the 2022 French presidential election. *arXiv preprint arXiv:2211.04577*, 2022.

Klaus Nehring and Clemens Puppe. A theory of diversity. *Econometrica*, 70(3):1155–1198, 2002.

Jonathan Stray. Designing recommender systems to depolarize. *First Monday*, 27(5), 2022.

Junlin Wu, Andrew Estornell, Lecheng Kong, and Yevgeniy Vorobeychik. Manipulating elections by changing voter perceptions. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, 2022.

Zhengui Xue, Zhiwei Lin, Hui Wang, and Sally McClean. Quantifying consensus of rankings based on q-support patterns. *Information Sciences*, 518:396–412, 2020.