The Structure of Bridging

CARTER BLAIR, Harvard University, USA
JAKOB DE RAAIJ, Harvard University, USA
ARIEL D. PROCACCIA, Harvard University, USA
MAXON RUBIN-TOLES, Harvard University, USA
NISARG SHAH, University of Toronto, Canada
MICHELLE SI, Harvard University, USA
SERENA WANG, Harvard University, USA

Influential deliberation platforms such as *Polis* and *Remesh* employ metrics for identifying *bridging* statements, which are accepted by participants who otherwise hold opposing views. A shortcoming of these metrics, however, is that they only account for inter-group connections in a fixed partition of the participants into groups. We argue that better bridging metrics must account for a richer set of possible partitions. To reason about such metrics, we develop a mathematical framework for bridging. We use it to identify two compelling metrics, *pairwise disagreement* and *p-mean bridging*, which are supported by axiomatic characterizations. Experiments on real data show that our metrics are stable, interpretable, and practical, even under the sparse observations typical of deliberation platforms.

Contents

1 Introduction

Civil discourse has been steadily eroding, marked by increasing polarization, declining trust, and a reduced capacity for constructive disagreement. In response to these trends, researchers, technologists, and practitioners have developed a wide range of online platforms designed to support deliberation.

A central principle underlying many of these platforms is the identification of statements or policy proposals that surface common ground among participants. This idea is often articulated through the notion of *bridging* [Ovadya, 2022], which refers to outcomes that are endorsed by participants who otherwise hold divergent or opposing views. Although bridging has not yet been defined in a fully rigorous manner, this core intuition has nonetheless been operationalized in a variety of ways.

An example of bridging in action is given by *Polis*, one of the best-known online deliberation platforms. In Polis, users may submit comments about the policy question at hand; they are also shown select comments submitted by others and vote on whether they agree or disagree with those comments — we refer to such votes as *approval* votes [Brams and Fishburn, 2007]. Polis then partitions the users into cohesive groups with similar opinions by completing the matrix of votes, projecting it onto the Euclidean plane through PCA, and running k-means clustering. Finally, the user-submitted comments are ranked according to a bridging metric called *group-aware consensus*, defined as the product of approval rates across the different groups in the partition [Small et al., 2021].

Another platform called *Remesh* provides a closely related but distinct view of bridging. Under Remesh, the groups are typically pre-specified, e.g., Israelis and Palestinians in a peacebuilding setting [Konya et al., 2025]. Moreover, the bridging score of a comment is the *minimum* approval rate across those groups rather than the product [Konya et al., 2023].

Polis and Remesh have both been influential; for example, Polis informed national ridesharing policy in Taiwan, while Remesh was employed by the United Nations for peacebuilding efforts in Libya. Nevertheless, we believe their approaches to bridging can be refined. At the heart of our criticism is the reliance on a single partition into groups, which may fail to capture important cleavages in the population. A challenge in exploring more nuanced notions, however, is that we currently lack a framework for reasoning about bridging. Therefore, our goal is to

> *... develop a formalism for bridging and, within it, single out attractive bridging metrics that take into account many plausible group structures.*

1.1 Our approach and results

A key question that immediately arises is which groups should be taken into account when designing *bridging functions* that measure the degree to which comments are bridging. Our answer builds on the idea that participants are *endogenously* split into two groups based on whether they approve or disapprove a comment y. A high approval rate for a comment x from *both* groups induced by another comment y — namely approvers and disapprovers of y — is a sign that x bridges a specific fault line defined by y. We measure the overall bridging score of x by averaging over all of these cleavages, that is, all of the partitions induced by comments y other than x itself. We call these measures *mean split functions*.

While the focus on these functions may seem like a contestable design choice, we argue that it is a restriction that arises naturally. Specifically, in Section 3, we prove that mean split bridging functions are *characterized* by three basic axioms: (election) additivity, anonymity, and homogeneity.

Using the basic structure of bridging provided by mean split functions, the next task is to decide how to quantify bridging with respect to a given partition into two groups (induced by

a comment y) — which we call a *split evaluator*. In Section 4, we observe that Polis and Remesh implicitly advance two different views of this question. Remesh favors *proportionality*, meaning that a comment is more bridging the more proportionally its approving participants are split between the groups. By contrast, Polis endorses *connectivity*, in the sense that a comment is more bridging the more pairs of approving participants belong to two different groups. Taking each of these two perspectives helps us identify appealing and practical bridging functions.

Starting with the connectivity principle, we develop a bridging function that we call *pairwise disagreement*. It has an especially intuitive interpretation: the bridging score of x is proportional to the sum of disagreement between pairs of participants approving x, where disagreement is defined as the fraction of comments approved by one member of the pair but not by the other. In Section 5, we prove that the split evaluator associated with pairwise disagreement is the only split evaluator that is consistent with Polis and satisfies a few other simple axioms. We further justify pairwise disagreement by establishing an independent characterization from first principles.

In Section 6, we adopt the proportionality viewpoint. It leads to a subfamily of split evaluators — parameterized by p — that we call *p-mean*. We show that this family, too, lends itself to an axiomatic characterization.

A practical challenge in deploying our bridging functions is that online deliberation platforms only obtain partial votes; that is, each participant only votes on a small subset of comments. In Section 7, we provide sample complexity bounds for estimating the aforementioned bridging functions under partial votes, showing that it suffices to have a number of participants that is logarithmic in the number of comments.

Finally, in Section 8, we implement our bridging functions and test them on two datasets, the French presidential elections (where we have access to complete votes) and Polis comments (an instance of sparse vote data). We visually demonstrate how pairwise disagreement, p-mean, and Polis bridging functions induce different bridging rankings depending on the distribution of approvals and disapprovals. We also show that our bridging functions are significantly more robust to missing votes than the approach used by Polis, which relies on a single partition into groups.

1.2 Related work

We discuss loosely related work in Appendix A, and focus here on the most closely related papers: those exploring how to measure (dis)agreement in elections with ordinal preferences. Two approaches prevail: Averaging over disagreement between pairs of voters, and averaging over disagreement of the population on pairs of alternatives. Both approaches have methodological similarities to our metrics in this paper. We give a high-level overview and refer to Karpov [2017] for a more extensive discussion. We note that neither approach has yet been extended specifically to approval preferences, to the best of our knowledge.

Disagreement between pairs of voters: Esteban and Ray [1994] measure how *polarized* a population is by, among others, summing over the *antagonism* between any pair of individuals in the population. They define antagonism as (a monotone function of) the individuals' distance in a 1-dimensional metric space. Ozkes [2013], Hashemi and Ulle [2014], and Karpov [2017] apply this polarization measure to ordinal elections, using the Kemeny distance and Spearman's footrule to measure the antagonism between pairs of voters. As we explain in more detail later, in spirit, this approach resembles our pairwise disagreement function.

Disagreement on pairs of alternatives: Alcalde-Unzu and Vorsatz [2013], Can et al. [2015], and Hashemi and Ulle [2014] consider for any pair of alternatives (x, y) the two groups consisting of the voters preferring x to y and of the voters preferring y to x. They define the agreement on (x, y) as the difference in the size of the groups, normalized by the number of voters, and let the disagreement be 1 minus the agreement. They measure how contentious an election is by averaging

(a monotone function of) the disagreement over all pairs of alternatives (x, y). Navarrete et al. [2023] extend this approach to measuring how *divisive* a single alternative is, still in an election with ordinal preferences. To calculate the divisiveness of an alternative x, they average over (a monotone function of) the absolute difference of the 'score' (e.g. the normalized Borda score) of x between the two groups given by (x, y) for all alternatives $y \neq x$. Colley et al. [2023] consider upweighting pairs (x, y) with high disagreement (in the definition from this paragraph); Endriss [2025] proposes also considering splits into two groups that are not necessarily due to a pair (x, y). These approaches resemble our approach of averaging a score for an alternative over group splits induced by the election. Notably, however, Navarrete et al. [2023], Colley et al. [2023], and Endriss [2025] measure divisiveness by averaging over a *different* set of group splits for every x, while we use the same set of group splits for all alternatives.

2 Model

An *approval election* $\mathcal{R} = (N, C)$ consists of a finite collection of *voters* N and a finite collection of *alternatives* (which can be *candidates* in a political election or *comments* in a deliberation platform) C. Since each voter approves a subset of alternatives in approval voting, we equivalently represent each alternative $x \in C$ as a subset of N consisting of the voters who approve x; that is, voter $i \in N$ approves alternative x if $i \in x$. We denote $\overline{x} = N \setminus x$ and say that voter $i \in N$ disapproves alternative x if $i \in \overline{x}$. We denote the number of voters as $n = |N|$ and the number of alternatives as $m = |C|$.

A *group* is a subset of voters, denoted $G \subseteq N$. Its *relative size* is $w_G = |G|/n$. For alternative x and group G, the *approval fraction of x in G* is $a_{x|G} = \frac{|x \cap G|}{|G|}$. For an alternative $y \in C$, we define the *(partition into) groups induced by y* as $\mathcal{G}_y = (y, \overline{y})$, which consists of the group of voters approving y and the group of voters disapproving y, with relative sizes w_y and $w_{\overline{y}}$, respectively.

We are interested in metrics that measure how bridging an alternative x is in an approval election $\mathcal{R}$. A *bridging function* $\mathcal{B}(x; \mathcal{R})$ maps an alternative $x \subseteq N$ and an approval election $\mathcal{R} = (N, C)$ to a *bridging score* in $[0, 1]$. Note that it is not necessary that $x \in C$. When N is clear from the context and C consists of a single alternative y, we will use $\mathcal{B}(x; y)$ as a shorthand for $\mathcal{B}(x; \mathcal{R})$.

Our results in Section 3 show that, subject to mild axioms, bridging functions are composed of a more fundamental building block, which we term *split evaluators*. Given a partition $\mathcal{G} = (G_h)_{h \in [r]}$ of the set of voters N and an alternative $x \in C$, a *split evaluator* b returns the score $b(a_{x|G_1}, \ldots, a_{x|G_r}; w_{G_1}, \ldots, w_{G_r}) \in [0, 1]$; notice that it depends only on the approval fractions of x in the groups and the relative sizes of the groups. The *Polis split evaluator* is

$$b^{\text{Polis}}(a_{x|G_1}, \ldots, a_{x|G_r}; w_{G_1}, \ldots, w_{G_r}) = \textstyle\prod_{h \in [r]} a_{x|G_h},$$

and the *Remesh split evaluator* is

$$b^{\text{Remesh}}(a_{x|G_1}, \ldots, a_{x|G_r}; w_{G_1}, \ldots, w_{G_r}) = \min_{h \in [r]} a_{x|G_h}.$$

Polis and Remesh use b^{Polis} and b^{Remesh}, respectively, for a given partition into groups $\mathcal{G}$ as their bridging functions $\mathcal{B}$. By contrast, we work with groups defined endogenously. Informally, we will use a split evaluator b to measure how bridging alternative x is among the approvers and disapprovers of another alternative y — using the partition $\mathcal{G}_y = (y, \overline{y})$ — and then aggregate the resulting score across all y to calculate the overall bridging score $\mathcal{B}(x; \mathcal{R})$. The following example instantiates these concepts.

Example 1. Consider an approval election $\mathcal{R} = (N, C)$ with the set of voters $N = [6]$ and the set of alternatives $C = \{x, y\}$ given by $x = \{1, 2, 3, 4\}$ and $y = \{1, 5\}$; that is, voter 1 approves both x and y; voters 2, 3, and 4 approve only x; voter 5 approves only y; and voter 6 approves neither alternative.

Let us evaluate how well each alternative bridges the groups of approvers and disapprovers of the other alternative.

Evaluating x under $\mathcal{G}_y = (y, \overline{y})$. Alternative y induces the groups $\mathcal{G}_y = (y, \overline{y}) = (\{1, 5\}, \{2, 3, 4, 6\})$ with relative sizes $w_y = \frac{|y|}{n} = \frac{1}{3}$ and $w_{\overline{y}} = \frac{|\overline{y}|}{n} = \frac{2}{3}$. The approval rates of x in these groups are

$$a_{x|y} = \frac{|x \cap y|}{|y|} = \frac{|\{1\}|}{2} = \frac{1}{2}, \qquad a_{x|\overline{y}} = \frac{|x \cap \overline{y}|}{|\overline{y}|} = \frac{|\{2, 3, 4\}|}{4} = \frac{3}{4}.$$

Hence,

$$b^{\text{Polis}}(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}}) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}, \qquad b^{\text{Remesh}}(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}}) = \min\left\{\frac{1}{2}, \frac{3}{4}\right\} = \frac{1}{2}.$$

Evaluating y under $\mathcal{G}_x = (x, \overline{x})$. Alternative x induces the groups $\mathcal{G}_x = (x, \overline{x}) = (\{1, 2, 3, 4\}, \{5, 6\})$ with relative sizes $w_x = \frac{|x|}{n} = \frac{2}{3}$ and $w_{\overline{x}} = \frac{|\overline{x}|}{n} = \frac{1}{3}$. The approval rates of y in these groups are

$$a_{y|x} = \frac{|y \cap x|}{|x|} = \frac{|\{1\}|}{4} = \frac{1}{4}, \qquad a_{y|\overline{x}} = \frac{|y \cap \overline{x}|}{|\overline{x}|} = \frac{|\{5\}|}{2} = \frac{1}{2}.$$

Hence,

$$b^{\text{Polis}}(a_{y|x}, a_{y|\overline{x}}; w_x, w_{\overline{x}}) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}, \qquad b^{\text{Remesh}}(a_{y|x}, a_{y|\overline{x}}; w_x, w_{\overline{x}}) = \min\left\{\frac{1}{4}, \frac{1}{2}\right\} = \frac{1}{4}.$$

3 Mean Split Bridging Functions

So far, we have imposed no structure on the bridging function $\mathcal{B}$; indeed, such a function can assign an arbitrary score to every alternative x in every approval election $\mathcal{R}$, which may not reflect how bridging the alternative is in any intuitive sense. Let us consider a natural structure for bridging functions.

Definition 1. The *mean split bridging function* $\mathcal{B}$ corresponding to split evaluator b is given by

$$\mathcal{B}(x; \mathcal{R}) = \frac{1}{m} \sum_{y \in C} b(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}}),$$

for all alternatives x and approval elections $\mathcal{R}$.

The key idea is to use the split evaluator $b(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}})$ to measure how bridging alternative x for the two groups induced by another alternative y, voters in y who approve it and voters in $\overline{y}$ who disapprove it. Generally speaking, x is a bridging alternative if it has a high approval rate from both these groups. Then, an overall bridging score for x for the entire election $\mathcal{R}$ is obtained by averaging its bridging score with respect to all alternatives y.

It turns out that mean split bridging functions from Definition 1 can be characterized using three standard axioms from social choice theory.

A1. Election Additivity. Let $\mathcal{R}' = (N, C')$ and $\mathcal{R}'' = (N, C'')$ be any two approval election instances with the same set of voters N, and with $|C'| = m'$ and $|C''| = m''$ alternatives, respectively. Let $\mathcal{R} = (N, C' \cup C'')$ be the combined election with $m = m' + m''$ alternatives. Then,

$$\mathcal{B}(x; \mathcal{R}) = \frac{m'}{m} \cdot \mathcal{B}(x; \mathcal{R}') + \frac{m''}{m} \cdot \mathcal{B}(x; \mathcal{R}'').$$

A2. Anonymity. For a permutation $\pi : N \to N$, define $\pi \circ y = \{\pi(i) \in N : i \in y\}$ and $\pi \circ C = \{\pi \circ y : y \in C\}$. For any permutation $\pi : N \to N$ and elections $\mathcal{R} = (N, C)$ and $\mathcal{R}' = (N, \pi \circ C)$, it holds that $\mathcal{B}(x; \mathcal{R}) = \mathcal{B}(x; \mathcal{R}')$.

A3. Homogeneity. Take any $k \in \mathbb{Z}_{\geqslant 1}$. For a set S, let $S^{(k)} = \cup_{a \in S} \left\{a^{(1)}, \ldots, a^{(k)}\right\}$ be the set that contains k duplicates of every element of S. For an election $\mathcal{R} = (N, C)$, the election $\mathcal{R}^{(k)} = (N^{(k)}, \left\{x^{(k)} : x \in C\right\})$ contains k duplicates of each voter $i \in N$ with each duplicate approving the same subset of alternatives as i. Then, it holds that $\mathcal{B}(x; \mathcal{R}) = \mathcal{B}(x^{(k)}; \mathcal{R}^{(k)})$ for each $x \in C$.

Election additivity (Axiom A1) resembles the linearity axiom that plays a key role in characterizing the Shapley value [Shapley, 1953]. Anonymity (Axiom A2) and homogeneity (Axiom A3) are extremely mild axioms in voting theory [Brandt et al., 2016], satisfied by most common voting rules. Anonymity (Axiom A2) embodies the "one person, one vote" principle, demanding no discrimination among voters based on their identities, while homogeneity (Axiom A3) normalizes an election by the number of voters, making bridging scores comparable across elections with different numbers of voters.

Theorem 1. *A bridging function satisfies Axioms A1, A2 and A3 if and only if it is a mean split bridging function.*

Proof. The "if" direction follows trivially: Axiom A1 from Definition 1, and Axioms A2 and A3 from the definition of split evaluators. We therefore focus on showing that a bridging function satisfies Axioms A1, A2 and A3 only if it is a mean split bridging function.

Applying election additivity (Axiom A1) repeatedly, we get

$$\mathcal{B}(x; \mathcal{R}) = \frac{1}{m} \sum_{y \in C} \mathcal{B}(x; y). \tag{1}$$

This allows decomposing an overall bridging score of an alternative x into the average of its bridging scores with respect to other alternatives y. It remains to show that Axioms A2 and A3 can reduce its bridging score with respect to alternative y, $\mathcal{B}(x; y)$, to a split evaluator.

We next show that anonymity (Axiom A2) implies that the pairwise bridging score $\mathcal{B}(x; y)$ depends only on four numbers: the number of voters who approve/disapprove x while approving/disapproving y. That is,

$$\mathcal{B}(x; y) = f(|x \cap y|, |\overline{x} \cap y|, |x \cap \overline{y}|, |\overline{x} \cap \overline{y}|) \tag{2}$$

for some function $f : \mathbb{Z}^4_{\geqslant 0} \to [0, 1]$ for all x, y. Consider any elections $\mathcal{R} = (N, C)$ and $\mathcal{R}' = (N, C')$, and alternatives $x, y \in C$ and $x', y' \in C'$. Suppose we have

$$s = |x \cap y| = |x' \cap y'|, \quad t = |x \cap \overline{y}| = \left|x' \cap \overline{y'}\right|, \quad u = |\overline{x} \cap y| = \left|\overline{x'} \cap y'\right|, \quad v = |\overline{x} \cap \overline{y}| = \left|\overline{x'} \cap \overline{y'}\right|.$$

We want to prove that $\mathcal{B}(x; y) = \mathcal{B}(x'; y')$. Recall that this is shorthand for $\mathcal{B}(x, (N, \{y\})) = \mathcal{B}(x', (N, \{y'\}))$. Consider a permutation $\pi : N \to N$ such that

$$\pi(x \cap y) = \{1, \ldots, s\}, \qquad \pi(x \cap \overline{y}) = \{s + 1, \ldots, s + t\},$$
$$\pi(\overline{x} \cap y) = \{s + t + 1, \ldots, s + t + u\}, \qquad \pi(\overline{x} \cap \overline{y}) = \{s + t + u + 1, \ldots, s + t + u + v = n\}.$$

Then, anonymity (Axiom A2) implies that

$$\mathcal{B}(x; y) = \mathcal{B}(\{1, \ldots, s + t\}, ([n], \{\{1, s\} \cup \{s + t + 1, s + t + u\}\})).$$

An analogous argument maps $\mathcal{B}(x'; y')$ to the same value, yielding $\mathcal{B}(x; y) = \mathcal{B}(x', y')$, as desired.

Next, we show that homogeneity (Axiom A3) implies that the dependence on the four numbers above can be reduced to the dependence on the corresponding fractions, normalized by the total

number of voters in the instance. That is, there exists a function $g : \Delta^3 \cap \mathbb{Q} \to [0,1]$ (with Δ^3 being the standard 3-simplex) such that

$$f(|x \cap y|, |\overline{x} \cap y|, |x \cap \overline{y}|, |\overline{x} \cap \overline{y}|) = g\left(\frac{|x \cap y|}{n}, \frac{|\overline{x} \cap y|}{n}, \frac{|x \cap \overline{y}|}{n}, \frac{|\overline{x} \cap \overline{y}|}{n}\right)$$

for all x, y. Consider any choices of $n, n' \in \mathbb{Z}_{\geqslant 1}$, $x, y \subseteq [n]$, $\mathcal{R} = ([n], \{y\})$, $x', y' \subseteq [n']$, and $\mathcal{R}' = ([n'], \{y'\})$ such that

$$\frac{|x \cap y|}{n} = \frac{|x' \cap y'|}{n'}, \quad \frac{|\overline{x} \cap y|}{n} = \frac{\left|\overline{x}' \cap y'\right|}{n'}, \quad \frac{|x \cap \overline{y}|}{n} = \frac{\left|x' \cap \overline{y}'\right|}{n'}, \quad \text{and } \frac{|\overline{x} \cap \overline{y}|}{n} = \frac{\left|\overline{x}' \cap \overline{y}'\right|}{n'}.$$

Then, we want to show that $\mathcal{B}(x; \mathcal{R}) = \mathcal{B}(x'; \mathcal{R}')$. We use homogeneity (Axiom A3) on $\mathcal{R}$ duplicated n' times and $\mathcal{R}'$ duplicated n times. Specifically, we get

$$\begin{aligned}\mathcal{B}(x; \mathcal{R}) = \mathcal{B}(x^{(n')}; \mathcal{R}^{(n')}) = f(n' \cdot |x \cap y|, n' \cdot |\overline{x} \cap y|, n' \cdot |x \cap \overline{y}|, n' \cdot |\overline{x} \cap \overline{y}|) = \\ = f(n \cdot |x' \cap y'|, n \cdot \left|\overline{x}' \cap y'\right|, n \cdot \left|x' \cap \overline{y}'\right|, n \cdot \left|\overline{x}' \cap \overline{y}'\right|) = \mathcal{B}((x')^{(n)}; (\mathcal{R}')^{(n)}) = \mathcal{B}(x'; \mathcal{R}'),\end{aligned}$$

where the translation from $\mathcal{B}$ to f and back uses Equation (2). This yields the desired equation $\mathcal{B}(x; \mathcal{R}) = \mathcal{B}(x'; \mathcal{R}')$.

It only remains to note that

$$\frac{|x \cap y|}{n} = a_{x|y} w_y, \quad \frac{|\overline{x} \cap y|}{n} = (1 - a_{x|y}) w_y, \quad \frac{|x \cap \overline{y}|}{n} = a_{x|\overline{y}} w_{\overline{y}}, \quad \text{and } \frac{|\overline{x} \cap \overline{y}|}{n} = (1 - a_{x|\overline{y}}) w_{\overline{y}},$$

so we can define our split evaluator b as

$$b(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}}) = g\left(\frac{|x \cap y|}{n}, \frac{|\overline{x} \cap y|}{n}, \frac{|x \cap \overline{y}|}{n}, \frac{|\overline{x} \cap \overline{y}|}{n}\right),$$

yielding $\mathcal{B}(x; y) = b(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}})$ as desired. □

Complementing Theorem 1, we remark that Axioms A1, A2 and A3 used in the characterization are independent of each other; every pair of them is satisfied by a bridging function that does not average split evaluations.

4 Two Perspectives on Bridging: Proportionality and Connectivity

The definition and axiomatic justification of mean split bridging functions reduces the problem of designing a bridging function $\mathcal{B}$ to the simpler problem of designing a split evaluator b. In particular, in mean split bridging functions, the split evaluator b is applied only to induced group partitions $\mathcal{G}_y$. To simplify notation moving forward, we will drop the subscripts for x and G whenever they are clear from the context and abbreviate the notation to $b(a_1, a_2; w)$, where $w_1 = w$ and $w_2 = 1 - w$. For b^{Polis} and b^{Remesh} that do not depend on w, we sometimes write only $b(a_1, a_2)$.

Let us recall our intuitive understanding of bridging: an alternative is bridging if it is endorsed by *many participants* who otherwise hold *very opposing views*. This identifies two dimensions in which an alternative can be bridging:

Approval. An alternative that is more approved is more bridging. In particular, if the approvers of one alternative are a subset of the approvers another alternative, the latter is more bridging.

Pluralism. An alternative with a more pluralistic set of approvers, i.e., having more disagreement among them on other alternatives, is more bridging. In particular, if two alternatives have the same number of approvers, the alternative where the approvers are spread *more evenly* across groups is more bridging.

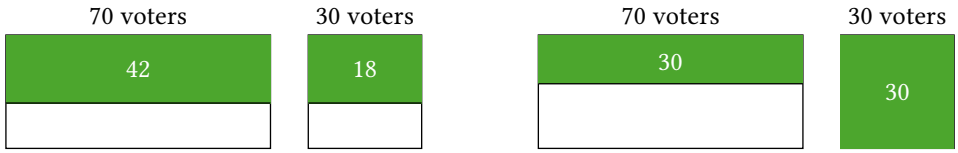
While the approval criterion is rather straightforward, the pluralism criterion is not: even b^{Polis} and b^{Remesh} disagree on what a *more even* split of the approvers into groups is.

Example 2. Consider two groups of relative sizes $w_1 = w = 0.7$ and $w_2 = 1 - w = 0.3$, and an alternative with a total approval rate of $a = 0.6$. We know that $a = a_1 w + a_2(1 - w)$, but the split of the total approval rate a into group-approval rates (a_1, a_2) can vary between $(0.6/0.7, 0)$ and $(0.3/0.7, 1)$.

Along this axis, $b^{\text{Remesh}}(a_1, a_2) = \min\{a_1, a_2\}$ is maximized at $a_1 = a_2 = 0.6$; this split for 100 voters is shown in Figure 1a.

Perhaps surprisingly, this is not the split that maximizes $b^{\text{Polis}}(a_1, a_2) = a_1 \cdot a_2$. For example, $b^{\text{Polis}}(0.6, 0.6) = 0.36$ whereas $b^{\text{Polis}}(3/7, 1) = 3/7 \approx 0.43 > 0.36$. Indeed, this is the maximizer: b^{Polis} is uniquely maximized when $a_1 w_1 = a_2 w_2 = 0.6/2$, which yields $(a_1, a_2) = (3/7, 1)$. This split for 100 voters is shown in Figure 1b.

Note that Polis aims to equalize the number of approvers from the two groups, while Remesh aims to equalize the approval rates from the two groups.



(a) A *proportional* approval split (b) A *connective* approval split

Fig. 1. Two different ways of distributing 60 approving voters across two groups of unequal size.

Based on the example, we further classify the pluralism dimension of bridging into two distinct paradigms:

Proportionality. An alternative with a fixed number of approvers a is most bridging if the number of approvers in each group is proportional to the size of the group, i.e., if the approval rates a_h are the same across all groups.

Connectivity. We say an alternative creates a *connection* between every pair of its approvers who come from different groups. An alternative with a fixed number of approvers a is the most bridging if it creates the highest number of connections. That is the case if the number of approving voters $n \cdot a_h \cdot w_h$ per group is (as) equal (as possible) across all groups.

These observations, including the "as equal as possible" comment in the last sentence, are formalized in the following result. The simple proof of the generalization to an arbitrary number of groups appears in Section B.1.

Theorem 2. *Fixing the total approval rate $a = a_1 w_1 + a_2 w_2$, b^{Remesh} is uniquely maximized when $a_1 = a_2 = a$, while b^{Polis} is uniquely maximized when $a_1 w = \min\{w, c\}$ and $a_2(1 - w) = \min\{1 - w, c\}$, where c is the solution to $\min\{w, c\} + \min\{1 - w, c\} = a$.*

In the next two sections, we consider which other split evaluators b are consistent with connectivity and proportionality, and provide axiomatic derivations for certain split evaluators.

5 Supporting Connectivity: Pairwise Disagreement

As demonstrated by Example 3 and formalized in Theorem 2, an obvious choice for a connective split evaluator is the Polis split evaluator. However, $b^{\text{Polis}}(a_1, a_2; w) = a_1 \cdot a_2$ makes the peculiar choice of disregarding the group sizes, since they consider only a single split into groups, so the fixed group sizes do not influence the ranking. But our mean split approach averages a split evaluator across many pairs of groups, which demands that the value of the split evaluator be comparable

across group splits with different values of w. How should the group sizes be taken into account, then? The following example is instructive.

Example 3. Consider a population of $n = 100$ voters with two groups, and an alternative that has $na_1w = na_2(1-w) = 30$ approvers in either group. In the first scenario, the groups are of uneven sizes with $w = 0.7$ (and $1 - w = 0.3$), so that $b^{\text{Polis}}(a_1, a_2) = 3/7 \cdot 1 \approx 0.43$. In the second scenario, the groups are of the same size $w_1 = w_2 = 0.5$, so that $b^{\text{Polis}}(a_1, a_2) = 3/5 \cdot 3/5 = 0.36$. The two scenarios are shown in Figure 2.

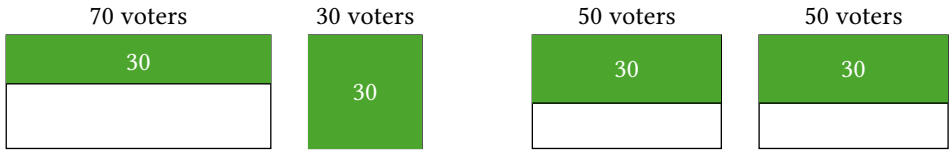


Fig. 2. Two possible splits of the population into two groups with 30 approvers each.

Observe that the only difference between the two scenarios is how the disapproving voters are distributed between the groups. We observe that b^{Polis} assigns a higher score to the alternative in the case where the disapproving voters are less pluralistic (all from the same group). This may be argued for by extending our intuitive definition of bridging: If we want to value approvals more highly if they come from a pluralistic group of approvers, we may also be more concerned about disapprovals from a pluralistic group, as is the case in the (50,50) split. By contrast, a line of work in social choice on veto power [Halpern et al., 2025, Moulin, 1981] argues that large, homogeneous groups of dissatisfied voters are undesirable, and that consequently homogeneous groups should have power to 'veto' proposals.

In the absence of a decisive resolution to the above debate, we propose to consider split evaluators that are agnostic to the distribution of the disapproving voters.

A4. Independence of disapproving voters. For any $w, a_1, a_2 \in [0,1]$ and $w', a_1', a_2' \in [0,1]$ such that the number of approving voters in each group is the same, i.e., $a_1w = a_1'w'$ and $a_2(1-w) = a_2'(1-w')$, it holds that $b(a_1, a_2; w) = b(a_1', a_2'; w')$.

In particular, we propose the following split evaluator. Recall that a *connection* is created between any two voters that are in different groups but are both approving; thus, the number of bridges in an instance is $(na_1w_1) \cdot (na_2w_2)$. A natural, connective approach to measure how bridging an alternative is, with no regard to the disapproving voters, is to simply count the number of connections, normalized to lie in $[0,1]$ so it can be compared across different values of n:

Definition 2. The *pairwise disagreement* split evaluator is

$$b^{\text{PD}}(a_1, a_2; w) = 4 \cdot a_1w \cdot a_2(1-w) = 4w(1-w) \cdot a_1a_2.$$

The respective *pairwise disagreement* mean split bridging function is

$$\mathcal{B}^{\text{PD}}(x; \mathcal{R}) = \tfrac{1}{m} \textstyle\sum_{y \in C} b^{\text{PD}}(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}}).$$

Our arguments in favor of b^{PD} are threefold: First, we show that (up to normalization), it is the only split evaluator that resembles b^{Polis} in two key features, while being agnostic to disapproving voters. Then, we provide a different set of intuitive axioms with no reference to Polis and show that they also uniquely characterize b^{PD} (up to normalization). Finally, we introduce a natural bridging function based on ideological distances between voters and prove that it is equivalent to b^{PD} — thereby explaining the choice of the name *pairwise disagreement*. Based on this, we highlight connections between b^{PD} and existing notions of measuring disagreement among voters.

5.1 Resembling Polis

We want our split evaluator to be b^{Polis}-like by inducing the same ranking as b^{Polis} for any fixed group sizes $w, 1-w$.

A5. Polis Ranking. For any $w \in [0,1]$, $b(a_1, a_2; w)$ induces the same ordering over (a_1, a_2) as $b^{\text{Polis}}(a_1, a_2; w)$. That is,

$$b(a_1, a_2; w) \geqslant b(a_1', a_2'; w) \Leftrightarrow b^{\text{Polis}}(a_1, a_2; w) \geqslant b^{\text{Polis}}(a_1', a_2'; w).$$

Additionally, we enforce that b conserves the linearity of b^{Polis}, at least when both groups are of equal size.

A6. Diagonal Linearity. For any $a, w \in [0,1]$ and $\lambda \in [0, 1/\alpha]$, $b(\lambda a, \lambda a; w) = \lambda^2 b(a, a; w)$.

This already suffices to uniquely characterize b^{PD} up to normalization, and in this class, b^{PD} is the natural choice that has range $[0,1]$ — see the proof in Section B.2.

Theorem 3. *Up to multiplication by a constant $c > 0$ independent of a_1, a_2, w, Axioms A4, A5 and A6 uniquely characterize b^{PD}.*

5.2 An independent axiomatic characterization of b^{PD}

Given the success of Polis, generalizing b^{Polis} is arguably already a desirable property in practice. Still, we can further place b^{PD} on a theoretically sound foundation by giving a different axiomatization, independent of b^{Polis}.

To that end, recall that one of the two factors determining how bridging an alternative is how evenly split the approvers are among the groups. We quantify this as an axiom by demanding that the effect of moving towards a more even group split should be proportional to the current difference in the number of approvers in the two groups. This is a strengthened version of assuming that b is maximizing connectivity.

A7. Linear Pigou-Dalton Principle. The marginal effect of one additional voter in the first group being approving and one less voter in the second group being approving should depend linearly on the difference in approving voters across the two groups. That is,

$$\left[\frac{\partial}{\partial \varepsilon} b\left(\frac{a_1 w + \varepsilon}{w}, \frac{a_2(1-w) - \varepsilon}{1-w}; w\right)\right]_{\varepsilon=0} \propto (a_2(1-w) - a_1 w),$$

with a positive proportionality constant.

We furthermore impose for normalization that an alternative that has no approval in either group is not bridging.

A8. Single-Group Approval. For any $a_1, a_2, w \in [0,1]$, $b(a_1, 0, w) = b(0, a_2, w) = 0$.

Together with the diagonal linearity axiom from before, Axiom A6, this again pins down b^{PD}, as shown in the following theorem, whose proof is relegated to Section B.3.

Theorem 4. *Up to multiplication by a constant $c > 0$ independent of a_1, a_2, w, Axioms A4, A6, A7 and A8 uniquely characterize b^{PD}.*

The reader may have noticed that so far, we have not argued for diagonal linearity (Axiom A6) beyond it being a property of b^{Polis}. While Axiom A6 is *one* natural choice to make for the behavior of a split evaluator b along the diagonal of $a_1 = a_2$, we point out that there is a second, equally natural, choice.

A9. Diagonal Linearity 2. For any $a, w \in [0,1]$ and $\lambda \in [0, 1/\alpha]$, $b(\lambda a, \lambda a; w) = \lambda b(a, a; w)$.

This change to the diagonal linearity axiom leads to a slightly different split evaluator.

Definition 3. The *harmonic pairwise disagreement* split evaluator is

$$b^{\mathrm{HPD}}(a_1, a_2; w) = \frac{4a_1 w a_2(1-w)}{a_1 w + a_2(1-w)}.$$

The respective *harmonic pairwise disagreement* mean split bridging function is

$$\mathcal{B}^{\mathrm{HPD}}(x; \mathcal{R}) = \tfrac{1}{m} \textstyle\sum_{y \in C} b^{\mathrm{HPD}}(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}}).$$

Theorem 5. *Up to multiplication by a constant $c > 0$ independent of a_1, a_2, w, Axioms A4 and A7 to A9 uniquely characterize b^{HPD}.*

We give a proof in Section B.4. Note that the term 'harmonic' indicates two properties: First, $b^{\mathrm{HPD}}(a_1, a_2; w)$ is the harmonic mean of $a_1 w$ and $a_2(1-w)$, multiplied by 2. Second, since b^{HPD} is equal to b^{PD} divided by the approval rate of the alternative, $a_1 w + a_2(1-w)$, it holds that b^{HPD} (and $\mathcal{B}^{\mathrm{HPD}}$) value pluralism ('harmony') more and approval rate less than b^{PD} (and $\mathcal{B}^{\mathrm{PD}}$). We revisit harmonic pairwise disagreement in more detail, together with (normal) pairwise disagreement, in the next subsection.

5.3 Bridging voter disagreement

To conclude the discussion of connective bridging functions, we briefly consider an approach to bridging different from the mean split approach. We prove that $\mathcal{B}^{\mathrm{PD}}$ and $\mathcal{B}^{\mathrm{HPD}}$ are equivalent to variants of this approach and have similarities to existing notions of voter disagreement.

Let us once again recall our intuitive definition of an alternative being bridging: It is approved by voters that usually disagree. Thus, given a *disagreement* metric $(d_{i,j}(\mathcal{R}))_{i,j \in N}$ capturing how divergent the views of voters i and j are in election $\mathcal{R}$, we may want to measure the bridging score of an alternative x by summing over all the disagreements of the voters that approve x, to get (with a normalizing factor) the bridging function

$$\mathcal{B}^d(x; \mathcal{R}) = \frac{1}{n^2} \sum_{i,j \in x} d_{i,j}(\mathcal{R}). \tag{3}$$

It is not surprising that for most distance metrics, $\mathcal{B}^d$ will not be a mean split bridging function. However, for one very natural choice, we actually recover pairwise disagreement, up to multiplication by a normalizing constant — proof in Section B.5.

Theorem 6. *Let the* Hamming disagreement *of two voters $i, j \in N$ be the fraction of alternatives for which their vote differs so that*

$$d_{i,j}(\mathcal{R}) = \tfrac{1}{m} \textstyle\sum_{y \in C} \mathbb{1}\left[(i \in y \wedge j \notin y) \vee (i \notin y \wedge j \in y)\right].$$

For this d,

$$\mathcal{B}^d(x; \mathcal{R}) = \mathcal{B}^{PD}(x; \mathcal{R}).$$

Using a disagreement metric d to measure how much disagreement is present in an election as

$$\Delta^d(\mathcal{R}) = \frac{1}{n^2} \sum_{i,j \in N} d_{i,j}(\mathcal{R})$$

has previously been used for elections $\mathcal{R}$ with ordinal preferences [Alcalde-Unzu and Vorsatz, 2013, Can et al., 2015, Hashemi and Ulle, 2014], In particular, Hashemi and Ulle [2014] propose using Spearman's footrule as a disagreement metric for strict ordinal preferences, which reduces to our notion of Hamming disagreement for approval preferences.

Hence, it becomes apparent that our notion of how bridging an alternative x is based on a disagreement metric as put forth in Equation (3) is equivalent to measuring the disagreement

among the approvers of x using Δ^d, then upweighting the result by the approval rate of x — see Section B.6 for the immediate proof.

Corollary 1. Let $\mathcal{R}|_x = (N \cap x, \{y \cap x : y \in C\})$ be the election $\mathcal{R} = (N, C)$ restricted to the voters approving x. Let d be Hamming disagreement and recall that $w_x = |x|/n$ is the fraction of voters that approve x. It holds that

$$\mathcal{B}^{\text{PD}} = w_x^2 \cdot \Delta^d(\mathcal{R}|_x), \qquad \mathcal{B}^{\text{HPD}} = w_x \cdot \Delta^d(\mathcal{R}|_x).$$

We believe that bridging functions of the structure described in this section, with other disagreement metrics $d_{i,j}$ to measure disagreement between pairs of voters or other functions Δ to measure disagreement among entire groups of voters, are an interesting subject for future work. We describe some potential extensions in Section 9.

6 Supporting Proportionality: p-Mean Bridging

We now focus on bridging functions that satisfy proportionality: for a fixed number of approving voters in the population, an alternative is most bridging if these approving voters are split across groups evenly, in proportion to the groups sizes. Thus, supplementing the pairwise disagreement functions proposed above, we now introduce a different family of split evaluators to better capture this notion of proportionality, and give the axioms that uniquely characterize it.

Definition 4. The *p-mean* with parameter $p \in \mathbb{R} \cup \{\pm\infty\}$, is

$$M_p(a_1, a_2; w) = \begin{cases} (w(a_1)^p + (1-w)(a_2)^p)^{1/p} & \text{for } p \in (-\infty, 0) \cup (0, \infty) \\ a_1^w a_2^{1-w} & \text{for } p = 0 \\ \min\{a_1, a_2\} & \text{for } p = -\infty \\ \max\{a_1, a_2\} & \text{for } p = \infty \end{cases}.$$

The *p-mean* split evaluator, with parameter $p \in \{-\infty\} \cup (-\infty, 1]$, is

$$b^{p\text{-mean}}(a_1, a_2; w) = M_p(a_1, a_2; w).$$

The respective bridging function is $\mathcal{B}^{p\text{-mean}}(x; \mathcal{R}) = \frac{1}{m} \sum_{y \in C} b^{p\text{-mean}}(a_{x|y}, a_{x|\overline{y}}; w_y)$.

The p-mean split evaluator can be interpreted as assuming that a group derives utility from approving voters equal to the approval fraction in this group, then using a p-mean welfare function on the groups' 'utilities' as the split evaluator. The parameter p is interpolating between the two dimensions of bridging, approval and pluralism. At $p = 1$, $b^{1\text{-mean}} = a_1 w + a_2(1-w)$ is the fraction of approving voters — the value of the split evaluator is independent of how they are distributed across the groups. At the other extreme, $p = -\infty$, where $b^{(-\infty)\text{-mean}} \equiv b^{\text{Remesh}}$, only the smaller of the two approval ratings matters; gaining approval in the other group does not increase the split evaluator's value — only pluralistic approval matters. Values of p between $-\infty$ to 1 interpolate between these extremes. Note that we do not consider $p > 1$ as this assigns higher bridging scores to alternatives that have a more uneven, thus less pluralistic, split of approvers.

The p-mean split evaluator is proportional, as shown in the following result, whose trivial proof appears in Section B.7.

Theorem 7. *Let $a, w \in [0, 1]$. On the line $a = a_1 w + a_2(1-w)$ for $a_1, a_2, w \in [0, 1]$, it holds that $b^{p\text{-mean}}(a_1, a_2; w)$ is maximized when $a_1 = a_2 = a$.*

Similarly to pairwise disagreement in the previous section, we provide two axiomatic justifications for p-mean split evaluators: First, we show that $b^{(-\infty)\text{-mean}}$ is only split evaluator that resembles b^{Remesh} in two key features. Then, we provide a different set of axioms with no reference to Remesh and show that they uniquely characterize the family of p-mean split evaluators.

6.1 Resembling Remesh

We want our split evaluator to be b^{Remesh}-like by inducing the same ranking as b^{Remesh} for any fixed group sizes $w, 1-w$.

A10. Remesh Ranking. For any $w \in [0,1]$, $b(a_1, a_2; w)$ induces the same ordering over (a_1, a_2) as $b^{\text{Remesh}}(a_1, a_2; w)$. That is,

$$b(a_1, a_2; w) \geqslant b(a_1', a_2'; w) \Leftrightarrow b^{\text{Remesh}}(a_1, a_2; w) \geqslant b^{\text{Remesh}}(a_1', a_2'; w).$$

Additionally, in the spirit of proportionality, we enforce that whenever the approval rate is homogeneous across groups, b assigns this approval rate as the split evaluation score.

A11. Diagonal Idempotency. For any $a, w \in [0,1]$, $b(a, a; w) = a$.

These axioms yield a characterization of $b^{(-\infty)\text{-mean}}$, as shown in Section B.8.

Theorem 8. *Axioms A10 and A11 uniquely characterize* $b^{(-\infty)\text{-mean}}$.

6.2 An independent axiomatic characterization of the $b^{p\text{-mean}}$ family

The family of weighted p-means for more than two elements (that is, groups) has been shown to be uniquely characterized by a set of intuitive axioms, including work by Cousins [2021] building on Debreu [1959]. In a similar spirit, we give a set of axioms that uniquely identify $b^{p\text{-mean}}$.

A12. Additivity. There exist continuous functions g and h so that $b(a_1, a_2; w) = h(wg(a_1) + (1-w)g(a_2))$.

A13. Linearity. For any $a_1, a_2, w \in [0,1]$ and $\lambda \in [0, \min\{1/a_1, 1/a_2\}]$, $b(\lambda a_1, \lambda a_2; w) = \lambda b(a_1, a_2; w)$.

A14. Weak Proportionality. For a fixed total approval share $a = a_1 w + a_2(1-w)$, it holds that $b(a_1, a_2; w)$ is maximized when $a_1 = a_2$.

We show (in Section B.9) that up to multiplication by a positive constant, $b^{p\text{-mean}}$ is the class of all split evaluators that are additive, linear, and (weakly) proportional.

Theorem 9. *Up to multiplication by a constant* $c \geqslant 0$*, Axioms A12 to A14 uniquely characterize* $b^{p\text{-mean}}$.

7 Estimating Bridging Functions With Partial Votes

Now we address the practicality of deploying our bridging functions on online platforms such as Polis and Remesh by considering the constraint of partially observed preferences. On digital democracy platforms, comments are shown to voters through a comment routing algorithm (described in Section 8), so not all voters will vote on every comment, and voters can choose to skip a comment if shown. Thus, the vote matrix A from which we compute the bridging scores — where rows denote voters, columns denote comments, and a 1 or -1 in A_{ij} means voter i approved/disapproved comment j — will often be sparse. Polis handles this through simple matrix completion by imputing the column-wise mean for unknown votes. We, however, work with partial votes and derive consistent estimators for the pairwise disagreement and p-mean bridging scores.

For simplicity, we assume each voter i independently votes on comment x with probability q_x, which depends on the comment but not the voter. For each voter i and comment x, define $V_{(i,x)} \sim \text{Ber}(q_x)$ to track whether i votes on x, where the $V_{(i,x)}$ are mutually independent for all $i \in N$ and $x \in C$ by assumption. Using these partial votes from all voters, our goal is to estimate bridging functions applied to the full election.

To estimate w_x, we let s_x, $s_{\overline{x}}$ denote the counts of approvers, disapprovers of comment x, and to estimate $a_{x|y}, a_{x|\overline{y}}$ for comment pairs x, y with $x \neq y$, we additionally define $s_{(a,b)}$ for $(a,b) \in$

$\{x, \overline{x}\} \times \{y, \overline{y}\}$:

$$s_x \sim \sum_{i \in x} V_{(i,x)}, \qquad s_{\overline{x}} \sim \sum_{i \in \overline{x}} V_{(i,x)}, \qquad s_{(a,b)} \sim \sum_{i \in a \cap b} V_{(i,x)}.$$

We can use these random variables to define empirical proxies of the inputs to the bridging functions:

$$\widehat{w}_y := \frac{s_y}{s_y + s_{\overline{y}}}, \qquad \widehat{a}_{x|y} := \frac{s_{(x,y)}}{s_{(x,y)} + s_{(\overline{x},y)}}, \qquad \widehat{a}_{x|\overline{y}} := \frac{s_{x,\overline{y}}}{s_{(x,\overline{y})} + s_{(\overline{x},\overline{y})}}.^{1}$$

To estimate bridging functions given incomplete information about voters' preferences, a natural approach is to substitute the inputs $w_y, a_{x|y}, a_{x|\overline{y}}$ with their empirical proxies. In fact, this approach is sufficient for bridging estimators with good asymptotic behavior. We will show that naive estimators for $\mathcal{B}^{\mathrm{PD}}(x; \mathcal{R})$ and $\mathcal{B}^{p-\mathrm{mean}}(x; \mathcal{R})$, which simply replace input quantities with their empirical estimates, converge to the true quantities with high probability for large values of n. Specifically, we consider the following estimators for $\mathcal{B}^{\mathrm{PD}}$ and $\mathcal{B}^{p-\mathrm{mean}}$, respectively:

$$\widehat{\mathcal{B}}^{\mathrm{PD}}(x; \mathcal{R}) := \frac{1}{m} \sum_{y \in C} b^{\mathrm{PD}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) = \frac{4}{m} \sum_{y \in C} \widehat{a}_{x|y} \widehat{a}_{x|\overline{y}} \widehat{w}_y (1 - \widehat{w}_y), \tag{4}$$

$$\widehat{\mathcal{B}}^{p\text{-mean}}(x; \mathcal{R}) := \frac{1}{m} \sum_{y \in C} b^{p\text{-mean}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) = \frac{1}{m} \sum_{y \in C} \left(\widehat{w}_y (\widehat{a}_{x|y})^p + (1 - \widehat{w}_y)(\widehat{a}_{x|\overline{y}})^p \right)^{\frac{1}{p}}. \tag{5}$$

The latter estimator is for the case $p \in (-\infty, 0) \cup (0, 1]$. Conditional on seeing approvals and disapprovals sufficiently often, we can bound the absolute difference between the estimators and true statistics. The following conditions make the problem of estimation tractable:

C1. There exists a $q_{\min} > 0$ such that $q_{\min} := \min_{x \in C} q_x$ and a $w_{\min} \in (0, \frac{1}{2}]$ such that for all $x \in C$, $w_x \in [w_{\min}, 1 - w_{\min}]$.[2]

C2. There exists $a_{\min} \in (0, 1]$ such that for all $x, y \in C$ with $x \neq y$, $a_{x|y}, a_{x|\overline{y}} \geqslant a_{\min}$.

C1 is quite mild and requires only that comments are shown to voters often enough and that approval is not close to 0 or 1; if a comment has approval $o(n)$, we can safely disregard it, and if a comment has approval $1 - o(n)$, we can say it is most bridging. While C2 is stronger, it is used only for the estimation bound of $\mathcal{B}^{p-\mathrm{mean}}$ when $p \in (-\infty, 0) \cup (0, 1]$, and without it, estimation of $\mathcal{B}^{\mathrm{PD}}$ and $\mathcal{B}^{p-\mathrm{mean}}$ for $p \in \{-\infty, 0\}$ is still possible. The following theorem formally states how the assumptions are used to derive sample complexity bounds.

Theorem 10 (Estimator consistency for partial votes setting.). *Fix instance $\mathcal{R}$, $x \in C$, and some $\epsilon, \delta \in (0, 1)$.*

(i) Under condition C1, let $\gamma := \min\{q_{min}, w_{min}\}$. Then, taking a particular[3]

$$n \in O\left(\frac{\ln(\frac{m}{\delta})}{\epsilon^2 \gamma^3} \right) \tag{6}$$

suffices to ensure the following:

$$\mathbb{P}[|\widehat{\mathcal{B}}^{PD}(x; \mathcal{R}) - \mathcal{B}^{PD}(x; \mathcal{R})| \geqslant \epsilon] \leqslant \delta. \tag{7}$$

[1] For the purpose of theoretical guarantees, each of these empirical estimates may take arbitrary value when its denominator is 0 since our high-probability bounds condition on a positive denominator for each term.

[2] Importantly, this imposes an upper bound on approval so that both "approvers" and "disapprovers" of x are $\Omega(n)$.

[3] Setting $n = \frac{9000 \ln(\frac{m}{\delta})}{\epsilon^2 \gamma^3}$ is sufficient.

(ii) Fix some $p \in (-\infty, 0) \cup (0, 1]$. Under conditions C1 and C2, let $\eta := \min\{a_{min}/2, q_{min}, w_{min}\}$. Then, taking a particular[4]

$$n \in O\left(\frac{\ln(\frac{m}{\delta})\eta^{2p-7}(p^2+1)}{\epsilon^2 p^2}\right) \tag{8}$$

suffices to ensure the following:

$$\mathbb{P}[|\widehat{\mathcal{B}}^{p-mean}(x;\mathcal{R}) - \mathcal{B}^{p-mean}(x;\mathcal{R})| \geqslant \epsilon] \leqslant \delta. \tag{9}$$

The theorem's proof is technical but mostly employs standard concentration inequalities; it is relegated to Section B.10. We further provide tighter special cases for $p = -\infty$ and $p = 0$.

While one might worry that these worst-case bounds require values of n that are infeasible for online platforms like Polis at their current scale, they give a qualitative dependence on the parameters. The experiments in the next section show that these estimators work well in practice.

8 Experiments

Sections 5 and 6 develop two bridging metrics, pairwise disagreement (PD) and the p-mean, with axiomatic characterizations. A natural question is how theoretical differences translate to applications with real data.

We compare PD and the p-mean against the group-aware consensus metric (GAC) deployed by Polis [Small et al., 2021]. We do not compare with the Remesh bridging metric [Konya et al., 2025] for two reasons. First, Remesh requires pre-specified demographic groups unavailable in our datasets. Second, the Remesh metric $b^{\text{Remesh}} = \min\{a_1, a_2\}$ coincides with the $p \to -\infty$ limit of the p-mean family, which we already include in our experiments, albeit applied to different groups.

We aim to answer three questions. First, on fully observed approval elections, do PD, GAC, and p-mean favor qualitatively different candidates? Second, how stable are the rankings returned by each metric as observations become sparser? Third, on naturally sparse approval data from Polis conversations, what kind of comments does each metric find most bridging?

8.1 Data

Our primary analysis focuses on two datasets which allow for evaluation under complete and naturally sparse approval data. We include three additional complete approval datasets in Section C.4.

French approval elections (complete data). We use two approval voting datasets collected during the French presidential elections in 2002 and 2007 from PrefLib [Mattei and Walsh, 2013]. The 2002 dataset has 2597 voters with 16 candidates, and the 2007 data has 2836 voters with 12 candidates.

Polis conversation (naturally sparse data). We also use approval data from the Seattle $15 minimum wage Polis discussion, sourced from PrefLib [Mattei and Walsh, 2013]. This dataset has 54 comments, 339 voters, and an observation rate q, defined as the fraction of possible alternative–voter approval entries that are observed, of 15.7%.

8.2 Mean Split Bridging Function Computation

We briefly describe how each function handles missing data. On complete data, PD and p-mean reduce to the formulas given in Section 5 and 6 respectively.

[4] Setting $n = \frac{288\ln(\frac{m}{\delta})\eta^{2p-7}(p^2+1)}{\epsilon^2 p^2}$ is sufficient.

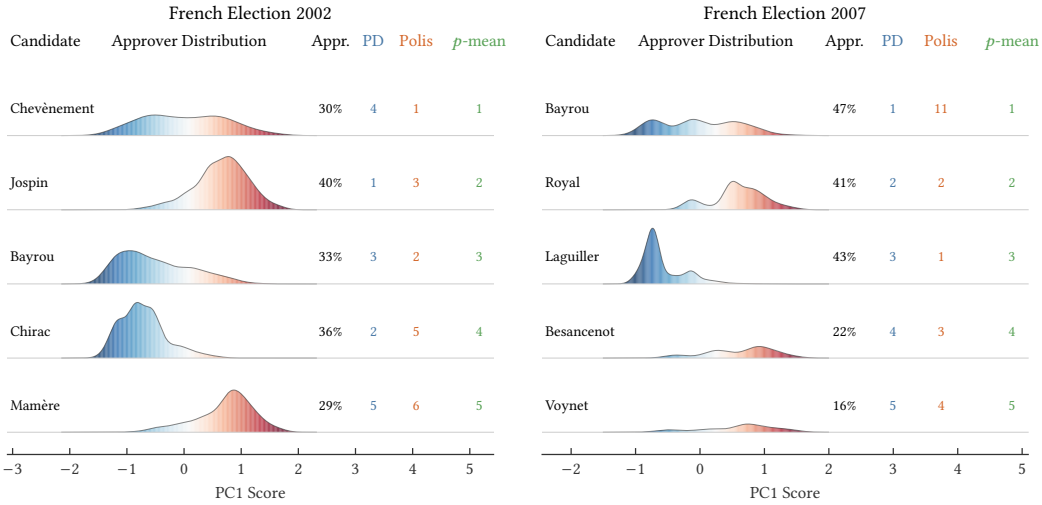


Fig. 3. Top 5 candidates by p-mean bridging score in the French presidential elections of 2002 (left) and 2007 (right). Ridgeline plots show the distribution of approving voters along the first principal component (PC1). Columns report overall approval rate and each candidate's rank under PD, Polis GAC, and p-mean.

Pairwise Disagreement and p-Mean. To calculate the bridging scores for these functions we first estimate $\widehat{w}_y, \widehat{a}_{x|y}$ and $\widehat{a}_{x|\overline{y}}$ from the data. We then use Equation (4) and Equation (5) to estimate PD and p-mean respectively. Unless otherwise stated, all experiments use $p = -\infty$ in p-mean.

Group-Aware Consensus. GAC follows the Polis pipeline [Small et al., 2021]: after filtering out voters with fewer than $\min\{7, m\}$ responses and imputing missing entries with per-comment mean approval, we compute a 2D PCA embedding of the vote matrix and cluster voters via k-means, selecting $k \in \{2, 3, 4, 5\}$ by silhouette score. The GAC score of item x is the product of Laplace-smoothed per-cluster approval rates, $\mathrm{GAC}(x) = \prod_{g=1}^{K} \frac{A_g(x)+1}{R_g(x)+2}$, where $A_g(x)$ and $R_g(x)$ are the observed approval count and total response count in cluster g. One difference from the Polis implementation is that Polis updates the clustering incrementally as additional voters are observed, using the previous clustering to inform the next update, whereas we perform a one-shot clustering on the final dataset. Section C.1 validates our GAC implementation against official Polis scores.

8.3 Experiment 1: A Qualitative Comparison on Fully Observed Elections

On complete data, all three metrics can be computed without estimation error, so any ranking differences reflect genuine disagreements about what constitutes bridging.

Setup. For each French election dataset, we compute all three metrics and display the top five candidates by p-mean with each candidate's overall approval rate and rank under each metric. To visualize each candidate's support distribution, we project voters onto the first principal component (PC1) of the centered vote matrix and plot a kernel density estimate of each candidate's approvers along this axis. We also plot each candidate in approval-heterogeneity space (Figure 4), where the y-axis measures approval and the x-axis measures mean Hamming distance among approvers.

Results. The results appear in Figures 3 and 4. In both elections, the candidate ranked first by p-mean has an approver distribution that spans the PC1 axis evenly, appearing visually bridging. In the 2002 election, no method selects a Pareto-dominated candidate in the approval-heterogeneity

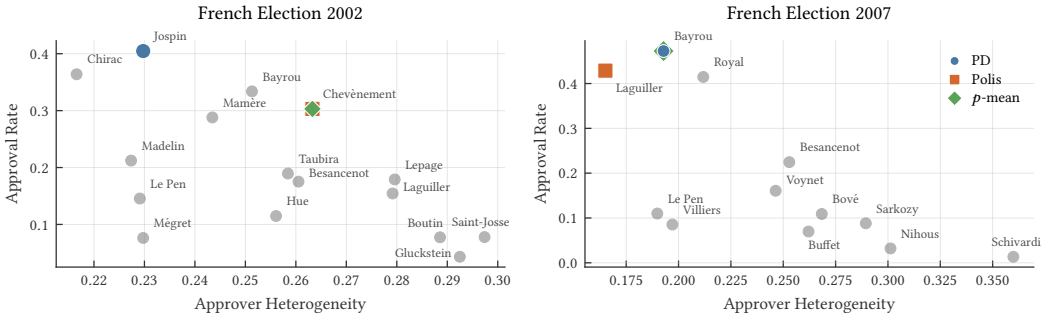


Fig. 4. Candidates plotted in approval-heterogeneity space for the 2002 (left) and 2007 (right) French presidential elections. Heterogeneity is measured as the mean Hamming distance among a candidate's approvers. The candidate with the highest score is highlighted for each bridging function.

plot. PD selects Jospin, who has higher approval but lower approver heterogeneity, while p-mean and GAC select Chèvenement, who has lower approval but whose approvers are more diverse. In the 2007 election, the candidate selected by PD and p-mean (Bayrou, a centrist) Pareto-dominates the candidate selected by GAC (Laguiller, a candidate with concentrated support). GAC ranks Bayrou 11th out of 12, highlighting the sensitivity of Polis GAC to the choice of partition: despite Bayrou's broad support, the particular clustering induces a low GAC score.

8.4 Experiment 2: Robustness Under Missing Data

In practice, Polis conversations are sparse: most voters see only a subset of comments. On the complete French election data, we test how stable each metric's ranking is as the observation rate decreases under two missingness models. For each election and metric, we treat the ranking induced by the complete matrix as ground truth and report Kendall's τ between this and the ranking estimated from a partially observed matrix. We use the estimation procedure described in Section 8.2. We run 20 independent trials per condition across observation rates $q \in \{0.05, 0.10, \ldots, 0.95\}$.

Missing completely at random (MCAR). We retain each matrix entry independently with probability q and compute all three metrics on the masked matrix.

Simulated Polis routing. Polis uses a routing algorithm to prioritize comments. We simulate this process to evaluate our metrics under non-uniform missingness. Each comment c receives a priority score following the production formula

$$\pi(c) = \Big((1 - p(c))\,(1 + E(c))\,a(c) \;\cdot\; \big(1 + 8 \cdot 2^{-S(c)/5}\big)\Big)^2,$$

where $a(c)$ and $p(c)$ are approval and pass rates, and $S(c)$ is the number of responses collected so far. $E(c)$ measures the distance from the center of the embedding to a theoretical participant who disapproved[5] comment c (and voted on no other comments), scaled to account for the sparsity of that single vote. For each voter, comments are sampled without replacement with probabilities proportional to $\pi(c)$.

Results. Figure 5 shows the results. Under MCAR, PD and p-mean both maintain high rank correlation with the ground truth, while GAC degrades noticeably when the observation rate is

[5]The Polis white paper [Small et al., 2021] states they create a hypothetical voter that agrees with comment c while the Polis source code creates a hypothetical voter that disagrees. We follow the source code in our implementation.

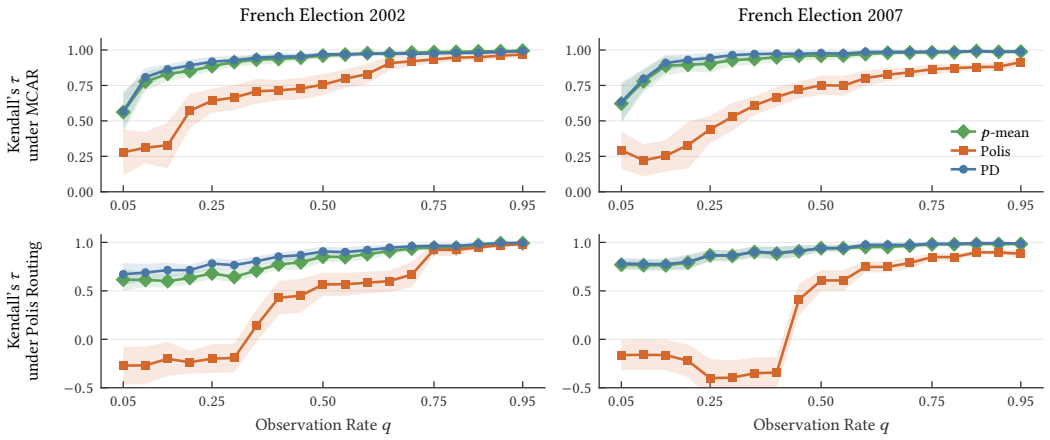


Fig. 5. Kendall's τ between estimated and ground-truth rankings as a function of observation rate. Lines show means over 20 trials. Shaded bands show ±1 standard deviation.

less than 0.5. Under simulated Polis routing, the gap widens: PD and p-mean remain stable, but GAC shows negative correlation at low observation rates in both elections. A likely explanation is that GAC concentrates its sensitivity into a single clustering step, which becomes unreliable at low observation rates. It is worth noting that the median observation rate across the Polis datasets on PrefLib is 15.5%, squarely in the range where GAC degrades in this experiment. In addition to the analysis of the full ranking here, we report top-1 accuracy in Section C.2. We also vary p in Section C.3 and confirm that robustness is largely insensitive to the choice of p.

8.5 Experiment 3: Polis Case Study

We apply all metrics to a Polis conversation from the Seattle $15 minimum wage discussion. Because missingness here is endogenous to both the routing algorithm and participants' choices, this experiment is descriptive. PCA requires a complete vote matrix, so we instead embed voters into one dimension using metric multidimensional scaling (MDS) on a voter-voter dissimilarity matrix $\delta(i, j)$, defined as the fraction of co-observed items on which voters i and j disagree.

Results. The results are shown in Figure 6. Arguably, p-mean picks the comment that appears the most bridging from the perspective of the approver distribution. Further, the comment chosen by p-mean Pareto-dominates the comment chosen by the Polis GAC in approval-heterogeneity space.

9 Discussion

At a conceptual level, our approach treats bridging as a property of an alternative's *approver structure*: an alternative is considered more bridging when it is approved by voters who otherwise tend to disagree. This interpretation is deliberately agnostic to semantic content and instead focuses on observable approval behavior. As a result, our metrics are well suited for platforms like Polis or Remesh, where approval votes are the primary signal available. At the same time, this abstraction means that our notion of bridging does not attempt to distinguish between different sources of disagreement — such as ideological distance, issue salience, or strategic voting — and instead collapses them into a single disagreement signal induced by the vote matrix.

Our approach hinges on the alternatives being a meaningful proxy for inferring opinion splits in the population. We believe that this generally holds true, especially in online deliberation

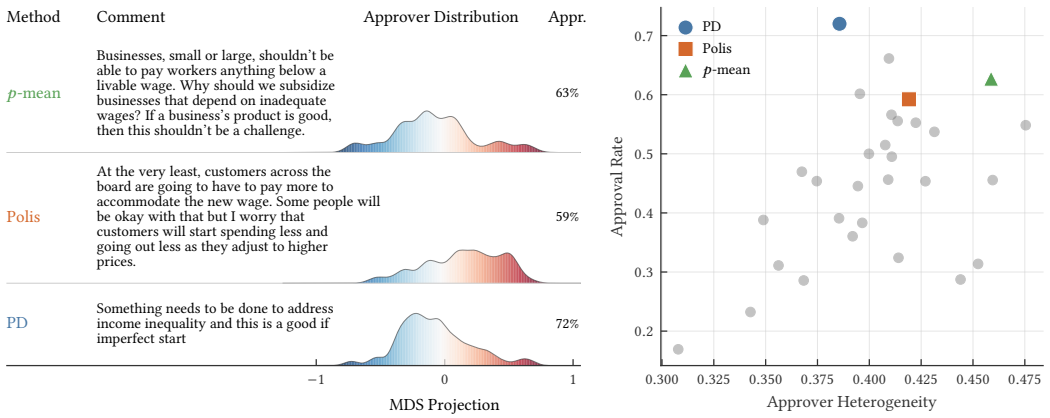


Fig. 6. (Left) Top-ranked comment under each metric for the Seattle $15 minimum wage Polis conversation. Each row shows the approver distribution along the MDS axis and approval rate. (Right) Approval-heterogeneity plot showing the comment ranked highest by each function.

platforms such as Polis: More contentious issues will be the subject of more comments while off-topic comments, leading to irrelevant opinion splits, will be filtered out with moderation. Automating this process is a natural direction for future work. We suggest a mean-split approach where the split evaluators for the splits induced by the comments are not evenly weighted. Through human moderation or LLM support, one may rate the substantiality of a comment, to downweight comments for being platitudes or off-topic and upweight comments with clear and relevant ideas.

The metrics we propose measure how much an alternative bridges *disagreement*. Faliszewski et al. [2023] point out that disagreement in a population can stem both from diversity — a large variety of distinct opinions — or polarization/fragmentation — a small number of groups with homogeneous opinions within but substantial disagreement across groups. Our mean-split approach is not suited to differentiate between these two settings, since it only considers pairs of alternatives. One may account for this by changing the pairwise voter disagreement metric $d_{i,j}$ or the election disagreement metric Δ, as discussed in Section 5.3, to place higher value on polarization or diversity. For example, using Euclidean distance (or any L^p norm for $p > 1$) instead of the Hamming distance will value alternatives that bridge a polarized set of voters more highly, while using a L^p 'norm' for $p < 1$ (where it is no longer a norm) will emphasize diversity.

Throughout our paper we treat the two induced groups in the split due to some comment, the approvers and the disapprovers, as identical. Switching the approvers and disapprovers of an alternative will not change the bridging score of any other alternative. In the case of online deliberation, this is rooted in the observation that a rational voter would flip their vote if a comment was changed to its logical negation. However, in other use cases such as elections (or even in online deliberation with irrational voters), one may argue that if two voters approve the same candidate, this makes them more similar than if they disapprove the candidate. To account for this, one may consider using the Jaccard distance instead of the Hamming distance as the distance metric $d_{i,j}$.

While there is much room to continue refining our framework for bridging, we believe it sets the stage for a more nuanced and rigorous understanding of the concept. Given the destructive role polarization plays in shaping discourse, developing ways to identify bridging outcomes is not just a technical challenge but a foundational one for democratic deliberation.

References

J. Alcalde-Unzu and M. Vorsatz. 2013. Measuring the Cohesiveness of Preferences: an Axiomatic Analysis. *Social Choice and Welfare* 41, 4 (2013), 965–988.

S. J. Brams and P. C. Fishburn. 2007. *Approval Voting* (2nd ed.). Springer.

F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice*. Cambridge University Press.

B. Can, A. I. Ozkes, and T. Storcken. 2015. Measuring Polarization in Preferences. *Mathematical Social Sciences* 78 (2015), 76–79.

R. Colley, U. Grandi, C. A. Hidalgo, M. Macedo, and C. Navarrete. 2023. Measuring and Controlling Divisiveness in Rank Aggregation. arXiv:2306.08511.

C. Cousins. 2021. An Axiomatic Theory of Provably-Fair Welfare-Centric Machine Learning. In *Proceedings of the 35th Annual Conference on Neural Information Processing Systems (NeurIPS)*. 16610–16621.

G. Debreu. 1959. *Topological Methods in Cardinal Utility Theory*. Technical Report 76. Cowles Foundation for Research in Economics, Yale University.

C. Dong, M. Bullinger, T. Wąs, L. Birnbaum, and E. Elkind. 2025. Selecting Interlacing Committees. In *Proceedings of the 24th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. 630–638.

U. Endriss. 2025. On the Difficulty of Measuring Divisiveness of Proposals under Ranked Preferences. arXiv:2512.24467.

J.-M. Esteban and D. Ray. 1994. On the Measurement of Polarization. *Econometrica* 62, 4 (1994), 819–851.

P. Faliszewski, A. Kaczmarczyk, K. Sornat, S. Szufa, and T. Wąs. 2023. Diversity, Agreement, and Polarization in Elections. arXiv:2305.09780.

D. Halpern, A. D. Procaccia, and W. Suksompong. 2025. The Proportional Veto Principle for Approval Ballots. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI)*. 3900–3907.

G. H. Hardy, J. E. Littlewood, and G. Pólya. 1934. *Inequalities*. Cambridge University Press.

V. Hashemi and E. Ulle. 2014. Measuring Diversity of Preferences in a Group. In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI)*. 423–428.

A. Karpov. 2017. Preference Diversity Orderings. *Group Decision and Negotiation* 26 (2017).

A. Konya, L. Schirch, C. Irwin, and A. Ovadya. 2023. Democratic Policy Development Using Collective Dialogues and AI. arXiv:2311.02242.

A. Konya, L. Thorburn, W. Almasri, O. Adomi Leshem, A. D. Procaccia, L. Schirch, and M. A. Bakker. 2025. Using Collective Dialogues and AI to Find Common Ground Between Israeli and Palestinian Peacebuilders. In *Proceedings of the 8th ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 312–333.

N. Mattei and T. Walsh. 2013. PrefLib: A Library for Preferences. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT)*. 259–270.

H. Moulin. 1981. The Proportional Veto Principle. *The Review of Economic Studies* 48, 3 (1981), 407–416.

C. Navarrete, M. Macedo, R. Colley, J. Zhang, N. Ferrada, M. E. Mello, R. Lira, C. Bastos-Filho, U. Grandi, J. Lang, and C. A. Hidalgo. 2023. Understanding Political Divisiveness Using Online Participation Data from the 2022 French and Brazilian Presidential Elections. *Nature Human Behaviour* 8, 1 (2023), 137–148.

A. Ovadya. 2022. Bridging-Based Ranking. Belfer Center for Science and International Affairs report.

A. I. Ozkes. 2013. Preferential Polarization Measures. HAL working paper.

L. S. Shapley. 1953. A Value for n-person Games. In *Contributions to the Theory of Games*, H.W. Kuhn and A. W. Tucker (Eds.). Annals of Mathematical Studies, Vol. 2. Princeton University Press, 307–317.

C. Small, M. Bjorkegren, T. Erkkilä, L. Shaw, and C. Megill. 2021. Polis: Scaling Deliberation by Mapping High Dimensional Opinion Spaces. *Revista De Pensament I Anàlisi* 26, 2 (2021).

M. H. Tessler, M. A. Bakker, D. Jarrett, H. Shehan, M. J. Chadwick, R. Koster, G. Evans, L. Campbell-Gillingham, T. Collins, D. C. Parkes, M. Botvinick, and C. Summerfield. 2024. AI Can Help Humans Find Common Ground in Democratic Deliberation. *Science* 386, 6719 (2024).

S. Wojcik, S. Hilgard, N. Judd, D. Mocanu, S. Ragain, M. B. F. Hunzaker, K. Coleman, and J. Baxter. 2022. Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation. arXiv:2210.15723.

A Additional Related Work

Faliszewski et al. [2023] point out that disagreement in the voting body can come from multiple sources. They argue that an election is *polarized* (or *fragmented*) if there are two (or more) groups where voters within a group hold very similar viewpoints but voters between groups disagree heavily; an example being half the voters having one preference ranking and the other half having its inverse. They argue that an election is *diverse* if there are many different viewpoints present; an example being *every* possible preference ranking being equally likely in the population. They present metrics to measure whether an election is polarized or diverse.

Dong et al. [2025] explore a related problem of selecting *interlacing committees* based on approval votes. Their Pairs objective counts the number of pairs of voters who approve a common alternative in the committee, and their Cons objective counts the number of pairs of voters that are connected by a sequence of alternatives in the committee. In our case of a singleton "committee," both objectives reduce to $\binom{n_a}{2}$, where n_a is the number of voters approving alternative a.

Halpern et al. [2025] propose an adaptation of the *proportional veto core* [Moulin, 1981] to approval votes. The key idea behind their notion is that the more "flexible" voters are (the more alternatives they approve), the more power they have. It can be argued that this notion provides another formal way to reason about bridging, and indeed it was the starting point for our work. However, we found that the formulation of Halpern et al. [2025] is incompatible with online deliberation platforms. For example, in the reasonable case where the set of alternatives is considered to be closed under negation, all (rational) voters are equally flexible regardless of their preferences.

There are several approaches to bridging that strongly rely on machine learning. One example is the Community Notes system developed by X [Wojcik et al., 2022]. Under this system, users propose notes providing context or corrections on posts; other contributors then rate these notes on helpfulness. The system uses a machine-learning model based on matrix factorization to infer both note quality and rater viewpoints from historical rating patterns. A note is identified as bridging when it is rated as helpful by contributors whose inferred viewpoints are meaningfully different from one another, and only notes that achieve this broad, cross-group helpfulness threshold are shown publicly. Another recent example is the Habermas Machine [Tessler et al., 2024], a system that uses large language models and voting to generate new statements that find common ground among participants. The latter uses a voting rule (specifically, the Schulze method) to select the most bridging statement.

B Missing Proofs

B.1 Proof of Theorem 2

To establish the claim about Remesh, note that if $a_1 = a_2 = a$, $b^{\text{Remesh}}(a_1, a_2) = a$, but if this is not the case, then due to $a_1 w + a_2(1-w) = a$, we have either $a_1 < a$ or $a_2 < a$, which yields $b^{\text{Remesh}}(a_1, a_2) < a$.

To establish the claim about Polis, observe that since w is fixed, maximizing $b^{\text{Polis}} = a_1 \cdot a_2$ is equivalent to maximizing $(a_1 w \cdot a_2(1-w))^{1/2}$. Thus, we are maximizing the geometric mean of two terms, $a_1 w$ and $a_2(1-w)$, with individual upper bound (w and $1-w$, respectively) and a fixed sum (a), which has a well-known closed form obtained by water-filling (which can be easily verified via the KKT conditions). Specifically, we find a unique value c such that both terms $a_1 w$ and $a_2(1-w)$ are made equal to c, except when their upper bounds are less than c (so, we set $a_1 w = w$ if $w < c$ and $a_2(1-w) = 1-w$ if $1-w < c$). □

B.2 Proof of Theorem 3

First, we'll show that we can write $b(a_1, a_2; w) = g_0(a_1 a_2; w)$ for some function g_0. Assume towards a contradiction that for some $w \in [0,1]$ there exist $a_1, a_2, a_1', a_2' \in [0,1]$ so that $a_1 a_2 = a_1' a_2'$ but $b(a_1, a_2; w) \neq b(a_1, a_2; w)$. We know that in this case $b^{\text{Polis}}(a_1, a_2; w) = a_1 a_2 = a_1' a_2' = b^{\text{Polis}}(a_1', a_2'; w)$, so Axiom A5 implies that $b(a_1, a_2; w) = b(a_1', a_2'; w)$; the desired contradiction. By Axiom A5 it furthermore follows that g_0 is strictly monotonically increasing in $a_1 a_2$.

By Axiom A6, $g_0(a_1 a_2; w) = b(\sqrt{a_1 a_2}, \sqrt{a_1 a_2}; w) = a_1 a_2 b(1, 1; w) = a_1 a_2 g_0(1; w)$, so we can write $b(a_1, a_2; w) = g(w) a_1 a_2$ for $g(w) = g_0(1; w)$. Axiom A4 implies that for fixed s_1, s_2, the function $b(\frac{s_1}{w}, \frac{s_2}{1-w}, w)$ is constant for all w with $s_1 \leqslant w \leqslant 1 - s_2$. Since $b(\frac{s_1}{w}, \frac{s_2}{1-w}, w) = \frac{g(w)}{w(1-w)} s_1 s_2$, this implies that $g(w) = c \cdot w(1-w)$ for all $0 < s_1 \leqslant w \leqslant 1 - s_2 < 1$, thus for all $w \in (0,1)$. For $w = 0$, note that by Axiom A4, $b(a_1, a_2; 0) = b(a_1', a_2; 0)$ for any $a_1, a_1', a_2 \in [0,1]$. In particular, we get $g(0) a_1 a_2 = b(a_1, a_2; 0) = b(a_1', a_2; 0) = g(0) a_1' a_2$, which for any for any $a_2 > 0$ and $a_1 \neq a_1'$ implies that $g(0) = 0$. Analogously, for $w = 1$, note that by Axiom A4, $b(a_1, a_2; 1) = b(a_1, a_2'; 1)$ for any $a_1, a_2, a_2' \in [0,1]$. In particular, we get $g(1) a_1 a_2 = b(a_1, a_2; 1) = b(a_1, a_2'; 1) = g(1) a_1 a_2'$, which for any for any $a_1 > 0$ and $a_2 \neq a_2'$ implies that $g(1) = 0$. Thus, $b(a_1, a_2; w) = cw(1-w) a_1 a_2$ for some constant c; Axiom A5 implies $c > 0$.

It follows that no function other than $c \cdot b^{\text{PD}}$ for $c > 0$ satisfies all three axioms. Verifying that all such functions indeed satisfies the three axioms is trivial. □

B.3 Proof of Theorem 4

First, we will show that we can write $b(a_1, a_2; w) = h(a_1 w, a_2(1-w))$ for some function h. Assume that there exist $a_1, a_2, w, a_1', a_2', w' \in [0,1]$ so that $a_1 w = a_1' w'$ and $a_2(1-w) = a_2'(1-w')$ but $b(a_1, a_2; w) \neq b(a_1, a_2; w)$. This is a contradiction to Axiom A4, the claim follows.

Now, consider a fixed $a = a_1 w + a_2(1-w) \in [0,1]$ and define $g_a(x) = h(x, a - x)$. We get that $\frac{d}{dx} g_a(x) = \frac{d}{dx} h(x, a - x) = \left[\frac{d}{d\varepsilon} b(\frac{a_1 w + \varepsilon}{w}, \frac{a_2(1-w) - \varepsilon}{1-w}; w) \right]_{\varepsilon = 0}$, so Axiom A7 implies that $\frac{d}{dx} g_a(x) = c_a((a - x) - x)$ for some constant $c_a > 0$ for every a. Integrating, we get that $g_a(x) = c_a(ax - x^2) + c_a' = c_a(a - x)x + c_a'$ for some constant c_a'. Thus, for any fixed a, we can write $b(a_1, a_2; w) = c_a(a_1 w)(a_2(1-w)) + c_a'$.

By Axiom A8 we know that $b(1, 0, w) = c_a' = 0$ for all $a \in [0,1]$. Axiom A6 now implies that $c_{\lambda a}(\lambda a w)(\lambda a(1-w)) = b(\lambda a, \lambda a; w) = \lambda^2 b(a, a; w) = \lambda^2 c_a(aw)(a(1-w))$. Thus, for any $a, a' \in (0,1]$, $c_a = c_{a'}$. Since $b \equiv 0$ whenever $a = 0$, we may w.l.o.g. also assume that $c_0 = c_a$ for all $a \in [0,1]$.

It follows that we can write $b(a_1, a_2; w) = c(a_1 w)(a_2(1-w)) = c b^{\text{PD}}(a_1, a_2; w)$ for some $c > 0$, so no function other than $c \cdot b^{\text{PD}}$ for $c > 0$ satisfies all four axioms. Verifying that all such functions indeed satisfy the four axioms is trivial. □

B.4 Proof of Theorem 5

Analogously to the proof of Theorem 4 up to invoking Axiom A6, we obtain that we can write $b(a_1, a_2; w) = c_a(a_1 w)(a_2(1-w))$ for some $c_a > 0$ where $a = a_1 w + a_2(1-w)$. Axiom A9 now implies that $c_{\lambda a}(\lambda a w)(\lambda a(1-w)) = b(\lambda a, \lambda a; w) = \lambda b(a, a; w) = \lambda c_a(aw)(a(1-w))$. Thus, for any $a \in (0,1]$, $c_{\lambda a} = \frac{1}{\lambda} c_a$. Since $b \equiv 0$ whenever $a = 0$, we may w.l.o.g. assume that $c_a = \frac{1}{a} c_1$ for all $a \in [0,1]$. It follows that we can write

$$b(a_1, a_2; w) = c \frac{(a_1 w)(a_2(1-w))}{a_1 w + a_2(1-w)} = c b^{\text{HPD}}(a_1, a_2; w)$$

for some $c > 0$, so no function other than $c \cdot b^{\text{HPD}}$ for $c > 0$ satisfies all four axioms. Verifying that all such functions indeed satisfy the four axioms is trivial. □

B.5 Proof of Theorem 6

By plugging in the definition of $d_{i,j}(\mathcal{R})$ and changing the order of summation, we get

$$\mathcal{B}^d(x;\mathcal{R}) = \frac{1}{n^2}\sum_{i,j\in x} d_{i,j}(\mathcal{R}) = \frac{1}{m}\sum_{y\in C}\frac{1}{n^2}\sum_{i,j\in x}\mathbb{1}\left[(i\in y \wedge j\notin y)\vee(i\notin y\wedge j\in y)\right].$$

The second summation is the number of pairs of voters that disagree on a given alternative y while both approving alternative x, this is $|x\cap y||x\cap\overline{y}| = n^2 a_{x|y}w_y a_{x|\overline{y}}w_{\overline{y}}$. We obtain

$$\mathcal{B}^d(x;\mathcal{R}) = \frac{1}{m}\sum_{x\in C} a_{x|y}w_y a_{x|\overline{y}}w_{\overline{y}} = \frac{1}{m}\sum_{x\in C} b^{\mathrm{PD}}(a_{x|y}, a_{x|\overline{y}}; w_y, w_{\overline{y}}) = \mathcal{B}^{\mathrm{PD}}(x;\mathcal{R})$$

□

B.6 Proof of Corollary 1

Plugging in the definitions and noting that $d_{i,j}(\mathcal{R}\mid_x) = d_{i,j}(\mathcal{R})$ for all $i,j\in x$, we get

$$w_x^2\cdot\Delta^d(\mathcal{R}|_x) = \left(\frac{|x|}{n}\right)^2\frac{1}{|x|^2}\sum_{i,j\in x} d_{i,j}(\mathcal{R}) = \mathcal{B}^d(x;\mathcal{R}) = \mathcal{B}^{\mathrm{PD}}(x;\mathcal{R}),$$

where the last equality follows from Theorem 6. The second part of the corollary follows since by definition, $\mathcal{B}^{\mathrm{HPD}}(x;\mathcal{R}) = \frac{1}{w_c}\mathcal{B}^{\mathrm{PD}}(x;\mathcal{R})$. □

B.7 Proof of Theorem 7

By the generalized mean inequality, $M_p(a_1,a_2;w)\leqslant M_q(a_1,a_2;w)$ whenever $p<q$. We know that $M_p(a,a;w)=a$. For any a_1,a_2 such that $a=a_1w+a_2(1-w)$, it thus holds that

$$M_p(a_1,a_2;w)\leqslant M_1(a_1,a_2;w) = wa_1+(1-w)a_2 = a = M_p(a,a;w).$$

□

B.8 Proof of Theorem 8

First, we'll show that we can write $b(a_1,a_2;w) = g(\min\{a_1,a_2\};w)$ for some function g. Assume towards a contradiction that for some $w\in[0,1]$ there exist $a_1,a_2,a_1',a_2'\in[0,1]$ so that $\min\{a_1,a_2\} = \min\{a_1',a_2'\}$ but $b(a_1,a_2;w)\neq b(a_1,a_2;w)$. We know that in this case $b^{\mathrm{Remesh}}(a_1,a_2;w) = \min\{a_1,a_2\} = \min\{a_1',a_2'\} = b^{\mathrm{Remesh}}(a_1',a_2';w)$, so Axiom A10 implies that $b(a_1,a_2;w) = b(a_1',a_2';w)$; the desired contradiction.

By Axiom A11, it follows that for any $a,w\in[0,1]$, $g(a;w) = g(\min\{a,a\};w) = b(a,a;w) = a$. Thus, no function other than $b^{(-\infty)\text{-mean}}$ satisfies both axioms. Verifying that it indeed satisfies the two axioms is trivial. □

B.9 Proof of Theorem 9

Combining Axioms A12 and A13, we get that for any $a\in[0,1]$ it holds that $ab(1,1;w) = b(a,a;w) = h(wg(a)+(1-w)g(a)) = h(g(a))$, to get that $h(g(a)) = c\cdot a$ for all $a\in[0,1]$. Thus, we can write $b(a_1,a_2;w) = cg^{-1}(wg(a_1)+(1-w)g(a_2)) = cM_g(a_1,a_2;w)$, the generalized mean generated by g, M_g, multiplied by a constant. It is well known that all generalized means generated by a continuous function that satisfy linearity, Axiom A13, are p-means [Hardy et al., 1934].

Now, assume towards a contradiction that there exists a function $f\equiv cM_g$ that is c times a linear, generalized mean generated by a continuous function g, but not c times a p-mean, i.e. $f\not\equiv cM_p$. Since linearity is unaffected by multiplication by a constant, dividing this f by c gives a linear, generalized mean generated by a continuous function g, M_g, that is not equivalent to a p-mean, a

contradiction. Thus, we can conclude that $b(a_1, a_2; w) = cM_p(a_1, a_2; w)$ for some constant c and $p \in \mathbb{R} \cup \{\pm\infty\}$.

Since $M_p(a_1, a_2; w) \geqslant 0$ and $b(a_1, a_2; w) \in [0, 1]$, we know that $c \geqslant 0$. For any $p \in (1, \infty)$,

$$b(1/2, 1/2; 1/2) = M_p(1/2, 1/2; 1/2) = 1/2 < (1/2)^{1/p} = M_p(1, 0; 1/2) = b(1, 0; 1/2),$$

and for $p = \infty$ it holds that

$$b(1/2, 1/2; 1/2) = M_p(1/2, 1/2; 1/2) = 1/2 < 1 = M_p(1, 0; 1/2) = b(1, 0; 1/2),$$

which all are in violation of Axiom A14.

Thus, all functions abiding by all three axioms are p-means for $p \leqslant 1$, multiplied by a constant $c \geqslant 0$. It is trivial to check that all p-means satisfy Axioms A12 and A13. By Theorem 7, we know that these functions satisfy Axiom A14. □

B.10 Proof of Theorem 10

First, we prove a concentration bound for empirical estimates of means given incomplete information, which allows us to bound the atomic quantities $\widehat{w}_y$, $\widehat{a}_{x|y}$, and $\widehat{a}_{x|\overline{y}}$:

Lemma 1. *Fix a finite population of size L with exactly K "successes," and let $\theta := K/L$. Include each population element independently with probability $\pi \in (0, 1]$. Let*

(1) $D :=$ number included,
(2) $X :=$ number of included successes,
(3) $\hat{\theta} := X/D$ when $D > 0$.

Let $\mu := \mathbb{E}[D] = \pi L$. Then for any $\epsilon \in (0, 1)$,

$$\mathbb{P}[|\hat{\theta} - \theta| \geqslant \epsilon] \leqslant 3\exp(-\epsilon^2\mu/8). \tag{10}$$

Proof. Since $D \sim \text{Bin}(L, \pi)$ and $\mathbb{E}[D] = \mu$, Chernoff's bound gives

$$\mathbb{P}[D \leqslant \frac{\mu}{2}] \leqslant \exp(-\frac{\mu}{8}). \tag{11}$$

Then, conditional on $D = d$, we want to bound the absolute difference $|\hat{\theta} - \theta|$. To do so, we define the random variables $X_1, \ldots, X_d$, where $X_i := \mathbb{I}[\text{draw } i \text{ is a success}]$, and so

$$\hat{\theta} = \frac{X}{d} = \frac{\sum_{i=1}^{d} X_i}{d}.$$

Note that $\mathbb{E}[\hat{\theta}] = \frac{K}{L} = \theta$,[6] where the randomness is now only over the particular elements of the population included, and the realization $d \sim D$ is fixed.

Hoeffding's inequality applies to sampling without replacement, giving the bound:

$$\mathbb{P}[|\hat{\theta} - \theta| \geqslant \epsilon] = \mathbb{P}[|X - \mathbb{E}[X]| \geqslant \epsilon d] \leqslant 2\exp(-2\epsilon^2 d). \tag{12}$$

Taking a union bound over (11) and (12) for the case that $d \geqslant \frac{\mu}{2}$ yields the following inequality:

$$\mathbb{P}[|\hat{\theta} - \theta| \geqslant \epsilon] \leqslant \exp(-\mu/8) + 2\exp(-\epsilon^2\mu). \tag{13}$$

The RHS $\leqslant 3\exp(-\epsilon^2\mu/8)$, giving the result. □

Corollary 2 (Error Bounds for Atomic Quantities)**.** Fix some $x, y \in C$ and some $\epsilon \in (0, 1)$. Under assumption $C1$, the following hold:

[6] To see this, note that $X \sim \text{Hypergeometric}(N, K, d)$, whose expectation is $d\frac{K}{N}$.

(i)
$$\mathbb{P}[|\widehat{w}_y - w_y| \geqslant \epsilon] \leqslant 3\exp(-\epsilon^2 n q_{\min}/8). \tag{14}$$

(ii)
$$\mathbb{P}[|\widehat{a}_{x|y} - a_{x|y}| \geqslant \epsilon] \leqslant 3\exp(-\epsilon^2 n w_{\min} q_{\min}^2/8). \tag{15}$$

(iii)
$$\mathbb{P}[|\widehat{a}_{x|\overline{y}} - a_{x|\overline{y}}| \geqslant \epsilon] \leqslant 3\exp(-\epsilon^2 n w_{\min} q_{\min}^2/8). \tag{16}$$

Proof. Each result follows from a straightforward application of Lemma 1:

(i) Applying Lemma 1 with $L = n$, $\pi = q_y$, $\theta = w_y$, $\widehat{\theta} = \widehat{w}_y$, and $\mu = \mathbb{E}[s_y + s_{\overline{y}}] = nq_y \geqslant nq_{\min}$ gives (14).

(ii) To estimate $\widehat{a}_{x|y}$ for $x \neq y$[7], we work within the subpopulation of voters y, which satisfies $|y| = nw_y$. In the language of Lemma 1, a voter i in this subpopulation is a "success" if he additionally approves x. We say voter j is "drawn" if she votes on x and y (so $V_{(i,x)} = V_{(i,y)} = 1$), which occurs with probability $q_x q_y$. Thus, we can apply Lemma 1 with $L = nw_y \geqslant nw_{\min}$, $\pi = q_x q_y$, $\theta = a_{x|y}$, $\widehat{\theta} = \widehat{a}_{x|y}$, and $\mu = nw_y q_x q_y \geqslant nw_{\min} q_{\min}^2$, giving (15).

(iii) An analogous argument for $a_{x|\overline{y}}$ using the upper bound on w_y implicit in C1 gives (16).

□

Next, we will compose these inequalities to bound the deviations $|b^{\mathrm{PD}}(\widehat{a}_{x,y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{\mathrm{PD}}(a_{x,y}, a_{x|\overline{y}}; w_y)|$, $|b^{p-\mathrm{mean}}(\widehat{a}_{x,y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{p-\mathrm{mean}}(a_{x,y}, a_{x|\overline{y}}; w_y)|$ for pairs of comments x, y when $x \neq y$.

Lemma 2 (Composition of Error Bounds: PD). *Suppose* $w, \widehat{w}, a_1, \widehat{a}_1, a_2, \widehat{a}_2 \in [0,1]$. *Define* $\epsilon := \max\{|\widehat{a}_1 - a_1|, |\widehat{a}_2 - a_2|, |\widehat{w} - w|\}$. *Then,*
$$|b^{PD}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b^{PD}(a_1, a_2; w)| \leqslant 15\epsilon. \tag{17}$$

Proof. First, write

$$|b^{\mathrm{PD}}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b^{\mathrm{PD}}(a_1, a_2; w)| = \max\{b^{\mathrm{PD}}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b^{\mathrm{PD}}(a_1, a_2; w), b^{\mathrm{PD}}(a_1, a_2; w) - b^{\mathrm{PD}}(\widehat{a}_1, \widehat{a}_2; \widehat{w})\}.$$

First, we bound $b^{\mathrm{PD}}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b^{\mathrm{PD}}(a_1, a_2; w)$:

$$\begin{aligned} b^{\mathrm{PD}}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b^{\mathrm{PD}}(a_1, a_2; w) &\leqslant (a_1 + \epsilon)(a_2 + \epsilon)(w + \epsilon)((1 - w) + \epsilon) - a_1 a_2 w(1 - w) \\ &= (a_1 a_2 + a_1\epsilon + a_2\epsilon + \epsilon^2)(w(1 - w) + w\epsilon + (1 - w)\epsilon + \epsilon^2) - a_1 a_2 w(1 - w) \\ &\leqslant 15\epsilon. \end{aligned}$$

The final inequality holds because there are 15 terms remaining after subtracting $a_1 a_2 w(1 - w)$, and each includes ϵ multiplied by a term no greater than 1.

By writing
$$b^{\mathrm{PD}}(a_1, a_2; w) - b^{\mathrm{PD}}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) \leqslant a_1 a_2 w(1 - w) - (a_1 - \epsilon)(a_2 - \epsilon)(w - \epsilon)((1 - w) - \epsilon),$$
one can perform a similar computation to verify that $b^{\mathrm{PD}}(a_1, a_2; w) - b^{\mathrm{PD}}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) \leqslant 15\epsilon$ as well, giving the result. □

Lemma 3 (Composition of Error Bounds: p-mean). *We bound* $|b^{p-mean}(\widehat{a}_1, \widehat{a}_2; w) - b(a_1, a_1; w)|$ *separately for the cases* $p \in (-\infty, 0) \cup (0, 1]$; $p = -\infty$; *and* $p = 0$:

[7]Note that when $x = y$, $a_{x|y} = 1$ and $a_{x|\overline{y}} = 0$, so deterministically, $b^{\mathrm{PD}}(a_{x|y}, a_{x|\overline{y}}; w_y) = b^{p-\mathrm{mean}}(a_{x|y}, a_{x|\overline{y}}; w_y) = 0$ for $p \in \{-\infty, 0\}$, and there is no need for estimation. When $p \in (-\infty, 0) \cup (0, 1]$, $b^{p-\mathrm{mean}}(a_{x|y}, a_{x|\overline{y}}; w_y) = (w_y)^{1/p}$, and so we have the bound from with $|\widehat{a}_1 - a_1| = |\widehat{a}_2 - a_2| = 0$ despite $a_{\min}$ not applying here. Thus, we don't consider the case $x = y$ separately. *Technically, $a_{x|\overline{x}}$ would be undefined in the case that $x = \emptyset$ or $\overline{x} = \emptyset$, but these cases are ruled out by* C2.

(i) ($p \in (-\infty, 0) \cup (0, 1]$):
Fix $p \in (-\infty, 0) \cup (0, 1]$. Suppose $w, \widehat{w} \in [0, 1]$; $a_1, a_2 \in [a_{min}, 1]$; and $\widehat{a}_1, \widehat{a}_2 \in [\alpha, 1]$, where $\alpha = a_{min}/2$. For ease of notation, define

$$b := wa_1^p + (1-w)a_2^p,$$
$$\widehat{b} := \widehat{w}\widehat{a}_1^p + (1-\widehat{w})\widehat{a}_2^p,$$

and so $(b)^{1/p} = b^{p-mean}(a_1, a_2; w)$, $(\widehat{b})^{1/p} = b^{p-mean}(\widehat{a}_1, \widehat{a}_2; \widehat{w})$.
Then

$$\left|(\widehat{b})^{\frac{1}{p}} - (b)^{\frac{1}{p}}\right| \leqslant ((a_{min}/2)^{p-1})(\frac{1}{|p|}|\hat{w} - w| + \max\{|\widehat{a}_1 - a_1|, |\widehat{a}_2 - a_2|\}). \tag{18}$$

(ii) ($p = -\infty$):
Assume $p = -\infty$. Suppose $a_1, \widehat{a}_1, a_2, \widehat{a}_2 \in [0, 1]$. Define $\epsilon := \max\{|\widehat{a}_1 - a_1|, |\widehat{a}_2 - a_2|\}$.
Then

$$\left|b^{p-mean}(\widehat{a}_1, \widehat{a}_2; w) - b^{p-mean}(a_1, a_2; w)\right| \leqslant \epsilon \tag{19}$$

(iii) ($p = 0$): Assume $p = 0$. From our definitions, we know that $b^{0-mean}(a_1, a_2; w) = a_1^w a_2^{1-w}$ and $b^{0-mean}(\widehat{a}_1, \widehat{a}_2; \widehat{w}) = \widehat{a}_1^{\widehat{w}}\widehat{a}_2^{1-\widehat{w}}$. Then

$$|\widehat{b}^{p-mean} - b^{p-mean}| \leqslant \ln\left(\frac{1}{\alpha}\right)|\widehat{w} - w| + \frac{1}{\alpha}\max\{|\widehat{a}_1 - a_1|,\ |\widehat{a}_2 - a_2|\} \tag{20}$$

Proof. We present proofs for the cases $p \in (-\infty, 0) \cup (0, 1]$; $p = -\infty$; and $p = 0$:

(i) First, we bound the quantity $|\widehat{b} - b|$. We add and subtract $w\widehat{a}_1^p + (1-w)\widehat{a}_2^p$:

$$\widehat{b} - b = (\widehat{w} - w)(\widehat{a}_1^p - \widehat{a}_2^p) + w(\widehat{a}_1^p - a_1^p) + (1-w)(\widehat{a}_2^p - a_2^p).$$

Hence

$$|\widehat{b} - b| \leqslant |\widehat{w} - w| \cdot |\widehat{a}_1^p - \widehat{a}_2^p| + \max\{|\widehat{a}_1^p - a_1^p|, |\widehat{a}_2^p - a_2^p|\}. \tag{21}$$

Next, we bound $|\widehat{a}_1^p - \widehat{a}_2^p|$.
If $0 < p \leqslant 1$, then $\widehat{a}_j^p \in [0, 1]$, so $|\widehat{a}_1^p - \widehat{a}_2^p| \leqslant 1$.
If $p < 0$, then $\widehat{a}_j \in [\alpha, 1]$ implies $\widehat{a}_j^p \in [1, \alpha^p]$, so $|\widehat{a}_1^p - \widehat{a}_2^p| \leqslant \alpha^p - 1$.
Thus, in any case,

$$|\widehat{a}_1^p - \widehat{a}_2^p| \leqslant \alpha^{p-1}. \tag{22}$$

Next, we apply a Lipschitz bound for $x \mapsto x^p$ on $[\alpha, 1]$. For any $p \neq 0$, the derivative satisfies $\frac{d}{dx}x^p = px^{p-1}$, and on $[\alpha, 1]$, $|px^{p-1}| \leqslant |p|\alpha^{p-1}$. We can apply mean value theorem with $x, y \in [\alpha, 1]$:

$$|x^p - y^p| \leqslant |p|\alpha^{p-1}|x - y|.$$

Applying this to $(\widehat{a}_1, a_1)$ or $(\widehat{a}_2, a_2)$, we obtain

$$|\widehat{a}_j^p - a_j^p| \leqslant |p|\alpha^{p-1}|\hat{a}_j - a_j|. \tag{23}$$

Plugging (22), (23) into (21) gives

$$|\widehat{b} - b| \leqslant \alpha^{p-1}(|\widehat{w} - w| + |p|\max\{|\widehat{a}_1 - a_1|, |\widehat{a}_2 - a_2|\}). \tag{24}$$

Finally, we give a Lipschitz bound for the outer map $t \mapsto t^{1/p}$. First, let $f(t) = t^{1/p}$.
If $0 < p \leqslant 1 : t \in [0, 1]$ since $b, \widehat{b} \in [0, 1]$, and

$$f'(t) = \frac{1}{p}t^{1/p-1}$$

Since $1/p - 1 \geqslant 0$ and $t \leqslant 1$, we have $t^{1/p-1} \leqslant 1$, hence $|f'(t)| \leqslant 1/p = 1/|p|$.

If $p < 0 : b, \widehat{b} \geqslant 1$, so $t \in [1, \infty)$. Also

$$|f'(t)| = \left|\frac{1}{p} t^{1/p-1}\right| = \frac{1}{|p|} t^{1/p-1}.$$

Since $1/p - 1 < 0$ and $t \geqslant 1$, we have $t^{1/p-1} \leqslant 1$, hence $|f'(t)| \leqslant 1/|p|$.
In either case,

$$|f(\widehat{b}) - f(b)| \leqslant \frac{1}{|p|} |\widehat{b} - b|. \tag{25}$$

Combining (24) and (25) gives (18).

(ii) Without loss of generality, assume $b^{p-\text{mean}}(a_1, a_2; w) = \min\{a_1, a_2\} = a_1$.
Case 1: $\widehat{a}_1 \leqslant \widehat{a}_2$.
Here, $b(\widehat{a}_1, \widehat{a}_2; \widehat{w}) = \widehat{a}_1$, so $|b(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b(a_1, a_2; w)| = |\widehat{a}_1 - a_1| \leqslant \max\{|\widehat{a}_1 - a_1|, |\widehat{a}_2 - a_2|\}$.
Case 2: $\widehat{a}_1 > \widehat{a}_2$.
We will split this case further:
Case 2a: $|\widehat{a}_2 - a_1| \leqslant |\widehat{a}_1 - a_1|$.
When this holds, $|b(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b(a_1, a_2; w)| = |\widehat{a}_2 - a_1| \leqslant |\widehat{a}_1 - a_1| \leqslant \max\{|\widehat{a}_1 - a_1|, |\widehat{a}_2 - a_2|\}$.
Case 2b: $|\widehat{a}_2 - a_1| > |\widehat{a}_1 - a_1|$. This is only possible if $\widehat{a}_2 < a_1$, but this implies $|\widehat{a}_2 - a_2| \geqslant |\widehat{a}_2 - a_1| = |b(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b(a_1, a_2; w)|$.
In any case, $|b(\widehat{a}_1, \widehat{a}_2; \widehat{w}) - b(a_1, a_2; w)| \leqslant \max\{|\widehat{a}_1 - a_1|, |\widehat{a}_2 - a_2|\} = \epsilon$, concluding the proof.

(iii) For $p = 0$ we use logarithms and start by taking the logs of the true value and estimated value: $\ln b = w \ln a_1 + (1 - w) \ln a_2$ and $\ln \widehat{b} = \widehat{w} \ln \widehat{a}_1 + (1 - \widehat{w}) \ln \widehat{a}_2$. Add and subtract $w \ln \widehat{a}_1 + (1 - w) \ln \widehat{a}_2$ to obtain

$$\ln \widehat{b} - \ln b = (\widehat{w} - w)(\ln \widehat{a}_1 - \ln \widehat{a}_2) + w(\ln \widehat{a}_1 - \ln a_1) + (1 - w)(\ln \widehat{a}_2 - \ln a_2).$$

And we get that

$$|\ln \widehat{b} - \ln b| \leqslant |\widehat{w} - w| \cdot |\ln \widehat{a}_1 - \ln \widehat{a}_2| + \max\{|\ln \widehat{a}_1 - \ln a_1|,\ |\ln \widehat{a}_2 - \ln a_2|\}. \tag{26}$$

Since $\widehat{a}_1, \widehat{a}_2 \in [\alpha, 1]$, we have $|\ln \widehat{a}_1 - \ln \widehat{a}_2| \leqslant \ln(1) - \ln(\alpha) = \ln\left(\frac{1}{\alpha}\right)$. Also, $\ln(\cdot)$ is $1/\alpha$-Lipschitz on $[\alpha, 1]$ because $(\ln x)' = 1/x \leqslant 1/\alpha$ for $x \in [\alpha, 1]$, so by the mean value theorem, $|\ln \widehat{a}_j - \ln a_j| \leqslant \frac{1}{\alpha} |\widehat{a}_j - a_j|$ for $j \in \{1, 2\}$.
Combining these bounds gives

$$|\ln \widehat{b} - \ln b| \leqslant \ln\left(\frac{1}{\alpha}\right) |\widehat{w} - w| + \frac{1}{\alpha} \max\{|\widehat{a}_1 - a_1|,\ |\widehat{a}_2 - a_2|\}. \tag{27}$$

Finally, since $a_1, a_2, \widehat{a}_1, \widehat{a}_2 \in (0, 1]$, we have $b, \widehat{b} \in (0, 1]$ and thus $\ln b, \ln \widehat{b} \leqslant 0$. By the mean value theorem applied to e^t on $(-\infty, 0]$, there exists ξ between $\ln b$ and $\ln \widehat{b}$ such that $|\widehat{b} - b| = |e^{\ln \widehat{b}} - e^{\ln b}| = e^{\xi} |\ln \widehat{b} - \ln b| \leqslant |\ln \widehat{b} - \ln b|$. Combining that with (26) gives the desired bound.

□

We can now combine the previous results to bound the deviations $|\widehat{\mathcal{B}} - \mathcal{B}|$ with high probability.

Theorem 11 (Estimator consistency for partial votes setting (Complete)). *Fix instance $\mathcal{R}$, $x \in C$, and some $\epsilon, \delta \in (0, 1)$.*

(i) Under condition C1, let $\gamma := \min\{q_{min}, w_{min}\}$.
Then, taking a particular

$$n \in O\left(\frac{\ln(\frac{m}{\delta})}{\epsilon^2 \gamma^3}\right) \tag{28}$$

suffices to ensure the following:

$$\mathbb{P}[|\widehat{\mathcal{B}}^{PD}(x;\mathcal{R}) - \mathcal{B}^{PD}(x;\mathcal{R})| \geqslant \epsilon] \leqslant \delta. \tag{29}$$

(ii) Fix some $p \in (-\infty, 0) \cup (0, 1]$. Under conditions C1 and C2, let $\eta := \min\{a_{min}/2, q_{min}, w_{min}\}$. Then, taking a particular

$$n \in O\left(\frac{\ln(\frac{m}{\delta})\eta^{2p-7}(p^2+1)}{\epsilon^2 p^2}\right) \tag{30}$$

suffices to ensure the following:

$$\mathbb{P}[|\widehat{\mathcal{B}}^{p-mean}(x;\mathcal{R}) - \mathcal{B}^{p-mean}(x;\mathcal{R})| \geqslant \epsilon] \leqslant \delta. \tag{31}$$

(iii) Assume $p = -\infty$. Under condition C1, let $\gamma := \min\{q_{min}, w_{min}\}$. Then, taking a particular

$$n \in O\left(\frac{\ln(\frac{m}{\delta})}{\epsilon^2\gamma^3}\right) \tag{32}$$

suffices to ensure the following:

$$\mathbb{P}[|\widehat{\mathcal{B}}^{p-mean}(x;\mathcal{R}) - \mathcal{B}^{p-mean}(x;\mathcal{R})| \geqslant \epsilon] \leqslant \delta. \tag{33}$$

(iv) Assume $p = 0$. Under conditions C1 and C2, define $\alpha := \frac{a_{\min}}{2}$ and let $\eta := \min\{\alpha, q_{\min}, w_{\min}\}$. Then, taking a particular

$$n \in O\left(\frac{\ln(\frac{m}{\delta})\left(\ln^2(\frac{1}{\eta})+1\right)}{\epsilon^2\,\eta^5}\right) \tag{34}$$

suffices to ensure the following:

$$\mathbb{P}\left[\left|\widehat{\mathcal{B}}^{p-mean}(x;\mathcal{R}) - \mathcal{B}^{p-mean}(x;\mathcal{R})\right| \geqslant \epsilon\right] \leqslant \delta. \tag{35}$$

Proof. We repeat a similar technique to show the final sample complexity bounds for each case:

(i) Take $n = \frac{9000\ln(\frac{m}{\delta})}{\epsilon^2\gamma^3}$.
We can take $\epsilon_w = \epsilon_a = \frac{\epsilon}{15}$. For the selected value of n, (14),(15), and (16) hold for (ϵ, δ) pairs $(\epsilon_w, \frac{\delta}{3m})$, $(\epsilon_a, \frac{\delta}{3m})$, and $(\epsilon_a, \frac{\delta}{3m})$, respectively. By union bound, all three bounds hold simultaneously with probability at least $\frac{\delta}{m}$.
Thus, with probability at least $1 - \frac{\delta}{m}$, (17) applies, giving

$$|b^{\mathrm{PD}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{\mathrm{PD}}(a_{x|y}, a_{x|\overline{y}}; w_y)| \leqslant \epsilon. \tag{36}$$

By union bound, the above holds $\forall y \in C$ with probability at least $1-\delta$. Conditional on the bound (36) holding for all $y \in C$, we can write

$$\begin{aligned}\left|\widehat{\mathcal{B}}^{\mathrm{PD}}(x;\mathcal{R}) - \mathcal{B}^{p-\mathrm{PD}}(x;\mathcal{R})\right| &= \left|\frac{1}{m}\sum_{y\in C} b^{\mathrm{PD}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - \frac{1}{m}\sum_{y\in C} b^{\mathrm{PD}}(a_{x|y}, a_{x|\overline{y}}; w_y)\right| \\ &= \left|\frac{1}{m}\sum_{y\in C}\left[b^{\mathrm{PD}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{\mathrm{PD}}(a_{x|y}, a_{x|\overline{y}}; w_y)\right]\right| \\ &\leqslant \frac{1}{m}\sum_{y\in C}\left|b^{\mathrm{PD}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{\mathrm{PD}}(a_{x|y}, a_{x|\overline{y}}; w_y)\right| \\ &\leqslant \frac{1}{m}\sum_{y\in C}\epsilon = \epsilon,\end{aligned}$$

giving the result.

(ii) Take $n = \frac{288\ln(\frac{m}{\delta})\eta^{2p-7}(p^2+1)}{\epsilon^2 p^2}$.

We can take $\epsilon_w = \frac{\epsilon|p|}{2\alpha^{p-1}}, \epsilon_a = \frac{\epsilon}{2\alpha^{p-2}}$. For the selected value of n, (14), (15), and (16) hold for (ϵ,δ) pairs $(\epsilon_w, \frac{\delta}{3m})$, $(\epsilon_a, \frac{\delta}{3m})$, and $(\epsilon_a, \frac{\delta}{3m})$, respectively. By union bound, all three bounds hold simultaneously with probability at least $\frac{\delta}{m}$.

Note that we have $\epsilon_a = \frac{\epsilon}{2\alpha^{p-2}} \leqslant \alpha$, ensuring that $\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}} \in [\alpha, 1]$ conditional on the atomic bounds holding. Thus, with probability at least $1-\frac{\delta}{m}$, (18) applies, giving

$$|b^{p-\mathrm{mean}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{p-\mathrm{mean}}(a_{x|y}, a_{x|\overline{y}}; w_y)| \leqslant \epsilon. \tag{37}$$

By union bound, the above holds $\forall y \in C$ with probability at least $1-\delta$. A similar argument as in the PD case gives the result.

(iii) Take $n = \frac{32\ln(\frac{m}{\delta})}{\epsilon^2\gamma^3}$.

We can take $\epsilon_a = \epsilon$. For the selected value of n, (15) and (16) hold for (ϵ,δ) pairs $(\epsilon_a, \frac{\delta}{2m})$ and $(\epsilon_a, \frac{\delta}{2m})$, respectively. By union bound, both bounds hold simultaneously with probability at least $\frac{\delta}{m}$.

Thus, with probability at least $1-\frac{\delta}{m}$, (19) applies, giving

$$|b^{p-\mathrm{mean}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{p-\mathrm{mean}}(a_{x|y}, a_{x|\overline{y}}; w_y)| \leqslant \epsilon. \tag{38}$$

By union bound, the above holds $\forall y \in C$ with probability at least $1-\delta$. The standard argument gives the result.

(iv) Take $n = \frac{1152\ln(\frac{m}{\delta})\left(\ln^2(\frac{1}{\eta})+1\right)}{\epsilon^2\eta^5}$, where $\alpha := \frac{a_{\min}}{2}$ and $\eta := \min\{\alpha, q_{\min}, w_{\min}\}$.

We can take $\epsilon_w = \frac{\epsilon}{2\ln(1/\alpha)}$ and $\epsilon_a = \frac{\alpha\epsilon}{2}$. Note that $\epsilon_a \leqslant \alpha$ since $\epsilon \in (0,1)$, ensuring that $\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}} \in [\alpha, 1]$ conditional on the atomic bounds holding.

For the selected value of n, (14), (15), and (16) hold for (ϵ,δ) pairs $(\epsilon_w, \frac{\delta}{3m})$, $(\epsilon_a, \frac{\delta}{3m})$, and $(\epsilon_a, \frac{\delta}{3m})$, respectively. By union bound, all three bounds hold simultaneously with probability at least $1-\frac{\delta}{m}$.

Thus, with probability at least $1-\frac{\delta}{m}$, (20) (the $p=0$ case) applies, giving

$$\left|b^{p-\mathrm{mean}}(\widehat{a}_{x|y}, \widehat{a}_{x|\overline{y}}; \widehat{w}_y) - b^{p-\mathrm{mean}}(a_{x|y}, a_{x|\overline{y}}; w_y)\right| \leqslant \ln\left(\frac{1}{\alpha}\right)\epsilon_w + \frac{1}{\alpha}\epsilon_a = \epsilon. \tag{39}$$

We apply the union bound the same way as all the parts above, which gives the result.

□

C Additional Experiments

C.1 Polis GAC Replication

To ensure a fair comparison between bridging metrics, we validate our implementation of Polis's Group-Aware Consensus (GAC) against the official Polis scores.

We use the BG2050 Volunteers dataset, which contains both raw votes and GAC scores exported from Polis. When we compute GAC using Polis's original cluster assignments, our scores agree with the exported scores up to numerical tolerance: Spearman $\rho = 1.000$, Kendall $\tau = 1.000$, and maximum absolute difference $< 10^{-10}$. This confirms that our implementation matches Polis's scoring given identical group assignments.

When we instead compute GAC using our own k-means clustering, which was implemented with the aim of being as close to the Polis method as possible, agreement with the exported Polis scores decreases (Spearman $\rho = 0.926$, Kendall $\tau = 0.773$), and the maximum absolute difference increases to 0.565. This divergence arises from differences in clustering methodology. Polis uses temporal seeding from incremental vote arrivals, whereas we apply single-shot clustering. While it was not the intention of the experiment, this again highlights the sensitiveity induced by relying on a single partition of voters into groups. Table 1 summarizes these results.

Table 1. Validation of GAC implementation against Polis (BG2050 dataset, $n = 371$ comments)

Metric	Polis groups	Our k-means
Spearman ρ	1.000	0.926
Kendall τ	1.000	0.773
Max. absolute difference	$< 10^{-10}$	0.565

C.2 Additional Robustness Results

In Section 8.4 we analyzed how correlated the estimated ranking and the true ranking were at different observation rates. However, one may care more about the accuracy of the top of the ranking compared to the bottom of the ranking. For this reason, we also look at stability from the perspective of how well each function recovers the highest ranked candidate. We measure this through top-1 accuracy which is the fraction of trials where the estimated top candidate matches the true top candidate. We aggregate over 20 trials under both the MCAR missingness and the Polis routing induced missingness.

Results. The results are plotted in Figure 7. Overall, PD appears to recover the highest ranked candidate most often. In the 2002 French election dataset p-mean and GAC perform comparably. However, in the 2007 French election, PD and p-mean perform comparably and significantly outperform GAC.

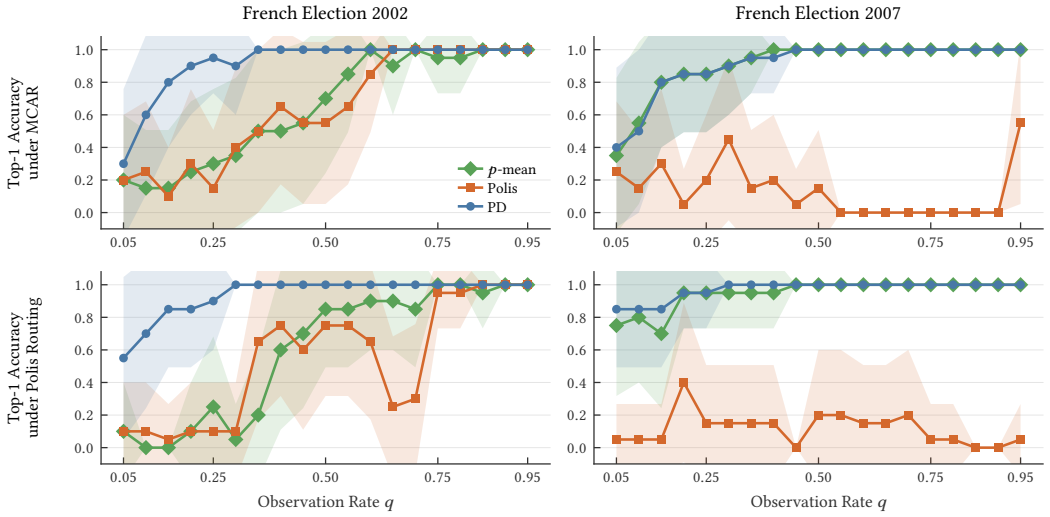


Fig. 7. Top-1 accuracy of the estimated ranking with respect to the ground-truth rankings as a function of observation rate. Top row: MCAR. Bottom row: simulated Polis routing. Left: 2002 French election. Right: 2007. Lines show means over 20 trials. Shaded bands show ±1 standard deviation.

C.3 Other Values of p in p-Mean

In our experiments in Section 8, we used $p = -\infty$ for p-mean. In this experiment, we look at p-mean with other values of p. We test $p = \{1, 0, -1, -2, -10\}$ and compare them to $p = -\infty$ qualitatively using the approval-heterogeneity plots and in terms of their robustness under MCAR missingness.

Results. The robustness results are in Figure 8 and Figure 11 shows the approval-heterogeneity plots. In regard to robustness, all values of p are comparable in terms of the Kendall's τ correlation between the estimated ranking and true ranking across observation rates. In the 2002 French election data, however, the top-1 accuracy tends to be higher for more positive values of p.

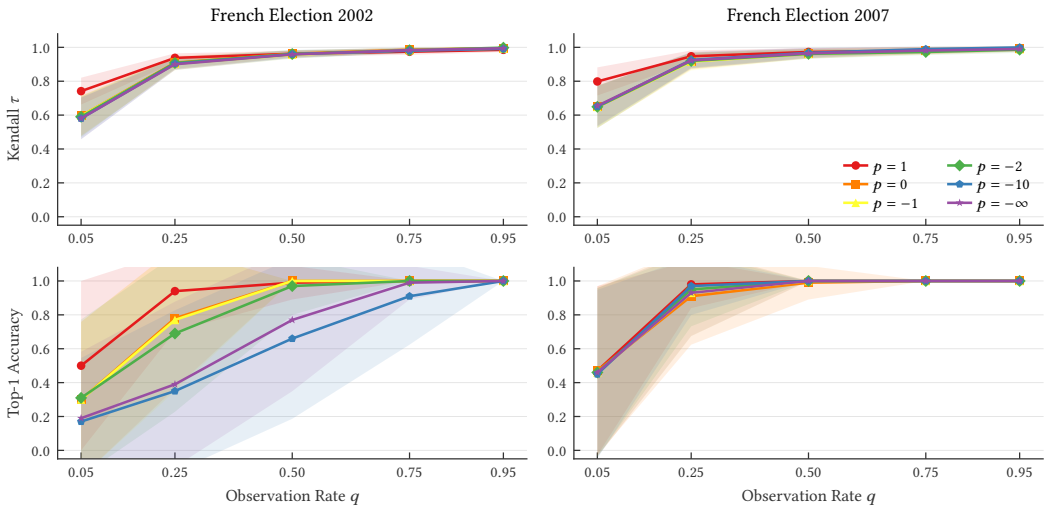


Fig. 8. Kendall's τ (top) and Top-1 accuracy (bottom) of the estimated ranking with respect to the ground-truth rankings as a function of observation rate under MCAR missingness. Left: 2002 French election. Right: 2007. Lines show means over 20 trials. Shaded bands show ±1 standard deviation.

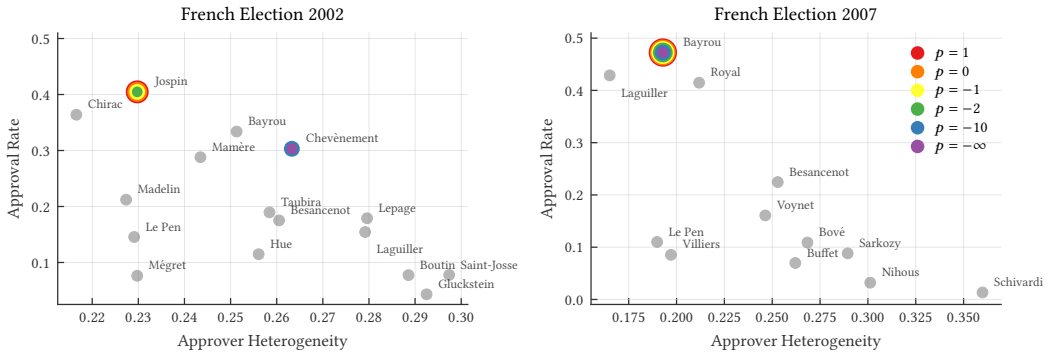


Fig. 9. Candidates plotted in approval-diversity space for the 2002 (left) and 2007 (right) French presidential elections. Diversity is measured as the mean Hamming distance among a candidate's approvers. The candidate with the highest score is highlighted for each value of p.

C.4 Additional Results on Complete Approval Matrices

We conducted further experiments on three additional complete approval datasets obtained from PrefLib [Mattei and Walsh, 2013]. These datasets, the *San Sebastian Poster Competition* (groups 1 and 2) and the *Czech Technical University (CTU) Tutorial Selection*, contain full approval ballots from smaller electorates compared to those in our main experiments. Our goal in this section is to complement the main results presented in Section 8.

C.4.1 Qualitative Comparisons. Figure 10 shows the top 5 candidates by p-mean bridging score along with ridgeline plots of approvers projected onto the first principal component (PC1). In all datasets, the overall rankings produced by PD, Polis GAC, and p-mean are in close agreement.

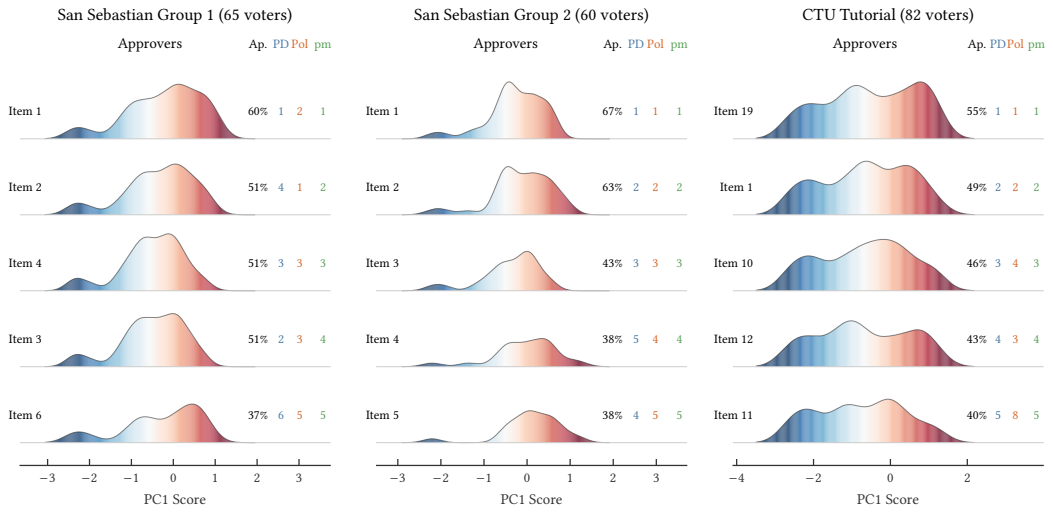


Fig. 10. Top 5 candidates by p-mean bridging score in the San Sebastian Poster Competition (left and middle) and the Czech Technical University tutorial selection dataset (right). Ridgeline plots show the distribution of approving voters along the first principal component (PC1). Columns report overall approval rate and each candidate's rank under PD, Polis's GAC (Pol), and p-mean (pm).

In the combined approval–diversity space shown in Figure 11, we observe that p-mean and PD select identical candidates across all datasets. In both the San Sebastian Group 2 and the CTU Tutorial dataset, all three metrics agree on the top candidate. In the San Sebastian Group 1 data, however, the Polis GAC selects a candidate with lower approver heterogeneity and a lower approval rate compared to the top candidates selected by PD and p-mean; this candidate is Pareto dominated in approval–diversity space.

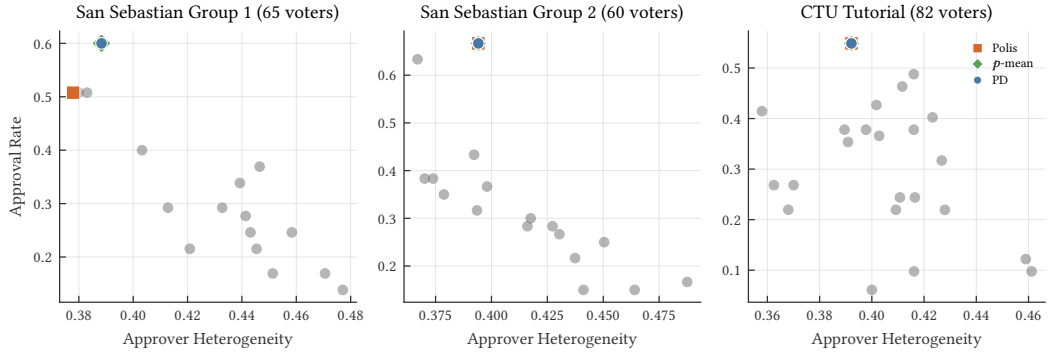


Fig. 11. Candidates plotted in approval–diversity space for the San Sebastian Poster Competition (left and middle) and the Czech Technical University tutorial selection dataset (right). Diversity is measured as the mean Hamming distance among a candidate's approvers. The candidate with the highest score is highlighted for each function.

C.4.2 Robustness Under Missingness. We further evaluated how the ranking functions perform under simulated missing data. Figures 12 and 13 report Kendall's τ correlation with ground-truth

rankings and Top-1 accuracy as functions of the observation rate under two missingness regimes: (i) Missing Completely at Random (MCAR) and (ii) missingness induced by Polis-style routing.

On these smaller datasets, the Polis GAC is more comparable to PD and p-mean in terms of robustness than observed in the larger datasets studied in Section 8.4. Nonetheless, PD and p-mean continue to outperform Polis GAC overall. In particular, in the San Sebastian Group 1 data, Polis GAC exhibits poor Top-1 accuracy under both missingness regimes.

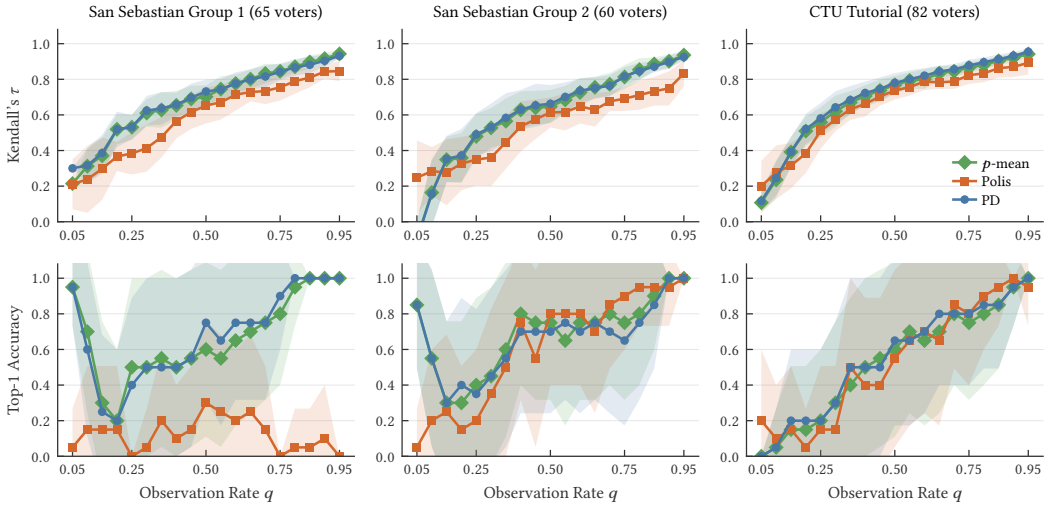


Fig. 12. Kendall's τ (top) and Top-1 accuracy (bottom) of the estimated ranking with respect to the ground-truth rankings as a function of observation rate under MCAR missingness. Left and middle: San Sebastian Poster Competition. Right: Czech Technical University tutorial selection dataset. Lines show means over 20 trials. Shaded bands show ±1 standard deviation.

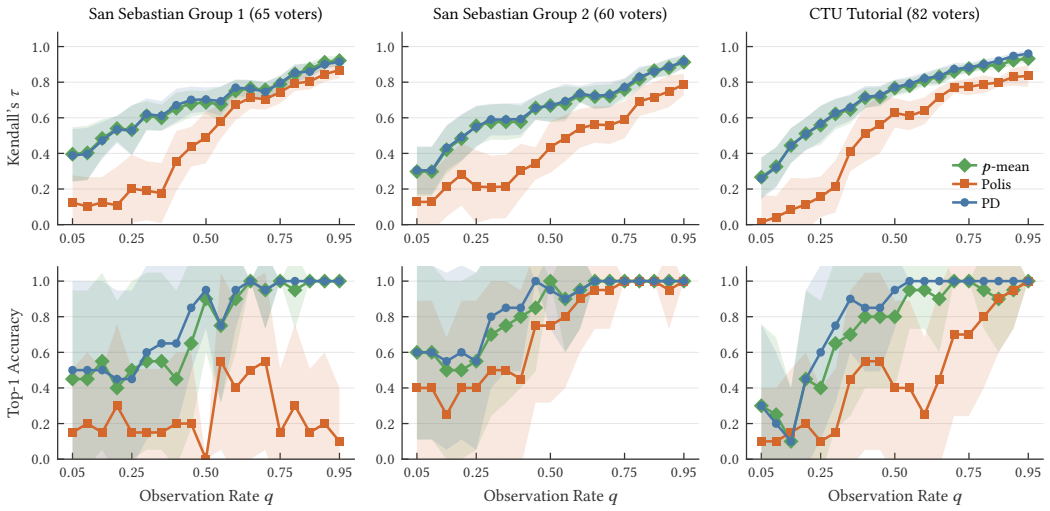


Fig. 13. Kendall's τ (top) and Top-1 accuracy (bottom) of the estimated ranking with respect to the ground-truth rankings as a function of observation rate under missingness induced by simulated Polis routing. Left and middle: San Sebastian Poster Competition. Right: Czech Technical University tutorial selection dataset. Lines show means over 20 trials. Shaded bands show ±1 standard deviation.