Boosting Sortition via Proportional Representation

Soroush Ebadian
University of Toronto
Toronto, Canada
soroush@cs.toronto.edu

Evi Micha
University of Southern California
Los Angeles, USA
evi.micha@usc.edu

ABSTRACT

Sortition is based on the idea of choosing randomly selected representatives for decision making. The main properties that make sortition particularly appealing are *fairness* — where every citizen has an equal chance of being selected— and *proportional representation* — where a randomly selected panel likely reflects the composition of the entire population. When the population lies on a representation metric, we formally define proportional representation by using a notion called the *core*. A panel is in the core if no group of individuals is underrepresented proportional to its size. While uniform selection is fair, it does not always return panels that are in the core. Thus, we ask if we can design a selection algorithm that satisfies fairness and *ex post core* simultaneously. We answer this question affirmatively and present an efficient selection algorithm that is fair and provides a constant-factor approximation to the optimal ex post core. Moreover, we show that uniformly random selection satisfies a constant-factor approximation to the optimal *ex ante core*. We complement our theoretical results by conducting experiments with real data.

CCS CONCEPTS

• **Theory of computation → Algorithmic game theory and mechanism design**; *Approximation algorithms analysis.*

KEYWORDS

sortition, core, proportional representation, fairness, clustering

ACM Reference Format:
Soroush Ebadian and Evi Micha. 2025. Boosting Sortition via Proportional Representation. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

1 INTRODUCTION

The random selection of representatives from a given population has been proposed to promote democracy and equality [30]. Sortition has gained significant popularity in recent years, mainly because of its use for forming *citizens' assemblies*, where a randomly selected panel of individuals deliberates on issues and makes recommendations. Currently, citizens' assemblies are being implemented by more than 40 organizations in over 25 countries [17].

Recently, there has been a growing interest within the computer science research community in designing algorithms that select representative panels fairly and transparently [11, 17–19]. Admittedly, a straightforward method for selecting a representative panel of size k from a given population of size n is to randomly select k individuals uniformly [14]. We refer to this simple procedure as uniform selection. As highlighted by Flanigan et al. [18], two main reasons make this method particularly appealing:

(1) *Fairness*: Each citizen is included in the panel with the same probability, satisfying the requirement of equal participation. Specifically, each citizen is selected with a probability of k/n
(2) *Proportional Representation*: The panel is likely to mirror the structure of the population, since if $x\%$ of the population has specific characteristics, then in expectation, $x\%$ of the panel will consist of individuals with these characteristics.

Indeed, uniform selection seems to achieve proportional representation *ex ante* (before the randomness is realized), since in expectation the selected panel reflects the composition of the population, especially when the size of the panel is very large. However, one of the critiques of this sampling procedure is that with non-zero probability, a panel that completely excludes certain demographic groups can be selected [14]. To address such extreme cases, various strategies have been proposed to ensure proportional representation *ex post* (after the randomness is realized) [27].

A common approach is *stratified sampling* [21], where individuals are divided into disjoint groups, and a proportional number of representatives is sampled uniformly at random from each group. However, this approach becomes impractical when dealing with a large predefined set of features, as the number of possible groups can grow exponentially, and there may not be enough seats in the panel to represent all of them. A more general approach, extensively used in practice, is to set quotas over individual or set of features [18, 31]. Similar to stratified sampling, when aiming for proportional representation across all intersectional features, the number of quotas can become exponential, making it infeasible to satisfy all of them concurrently. Alternatively, one may opt for setting quotas over a subset of intersectional features. For instance, quotas could be set for gender and race simultaneously, along with additional quotas for income. However, this might not ensure the representation of specific subgroups, such as high-income black women.

The presence of the above challenges in existing strategies prompts a need for alternative approaches for ensuring proportional representation. This, in turn, highlights the necessity of rigorously defining proportional representation first. Our work departs from these observations, and we aim to address the following questions:

(1) *What is a formal definition of proportional representation of a population?*
(2) *To what extent does uniform selection satisfy proportional representation?*
(3) *Is it possible to design selection algorithms that enhance representation guarantees while maintaining fairness?*

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1.1 Our approach

Proportional representation via core. We begin by tackling the first question posed above, by borrowing a notion of proportional representation used by recent works on multiwinner elections, fair allocation of public goods and clustering [2, 7–9, 15], called the *core.* The main idea of the core is: *Every subset S of the population is entitled to choose up to $|S|/n \cdot k$ representatives.* Formally, a panel P is called proportionally representative, or is said to be in the core, if there does not exist a subset S of the population that could choose a panel P', with $|P'| \leq |S|/n \cdot k$, under which all of them feel more represented. Note that this notion is not defined over predefined groups using particular features, but it provides proportional representation in the panel to *every* subset of the population.

Representation metric space. A conceptual challenge is to quantify the extent to which a panel represents an individual. To address this, we use the same approach as taken by Ebadian et al. [11] in which it is assumed that the individuals lie in an underlying *representation metric space.* The representation metric space can be constructed as a function of features that are of particular interest for an application at hand, such as gender, age, ethnicity and education. Significantly, our results depend only on the existence of such a metric space without further assumptions.

q-cost. To measure how well an individual is represented by a panel, we adopt the approach of Ebadian et al. [11], building on work of Caragiannis et al. [6] in multiwinner elections. Specifically, an individual's cost for a panel is determined by her distance from the q-th closest member in the panel, for some $q \in [k]$. We argue this cost is more suitable for sortition than other well-known cost functions, such as average cost, for two main reasons.

First, the q-cost does not consider the distance from every representative. In sortition, where the goal is to represent an entire population, it is reasonable to expect that some representatives will be, or should be, distant from certain individuals. For example, a Black woman on a panel may expect the presence of White men with whom she does not identify. However, she would "require" at least a few representatives with a similar background to her. The q-cost function captures this expectation by ensuring that at least q representatives are close to each individual.

Second, the q-cost distinguishes whether an individual has a few very similar representatives rather than many who are moderately distant. Consider a 40-year-old evaluating two panels: the first includes two 40-year-olds, a 20-year-old, and a 60-year-old; the second has two 30-year-olds and two 50-year-olds. If distance is measured by age difference, the average cost is the same for both panels. However, the q-cost (for $q = 2$) highlights a key difference: in the first panel, there are two representatives who are exactly the same age as the individual, while in the second panel, all representatives are somewhat distant in terms of age. Thus, the q-cost effectively differentiates between these two panels (with the appropriate choice of q, though our results hold for all q).

1.2 Our Contribution

In a sortition setting, in addition to proportional representation of all groups, it is important to ensure the *fairness constraint* which is that all individuals have the same chance of being included in the panel. For ensuring that, a selection algorithm should return a distribution over panels of size k. In this work we ask for selection algorithms that are simultaneously in the *ex post core*, meaning that *every* panel that the algorithm might return, is in the core, and simultaneously is fair, meaning that each individual is included in the panel with probability equal to k/n.

In Section 3, we show that uniform selection, despite satisfying fairness by its definition, falls short of achieving any reasonable approximation to the ex post core for almost any q. Moreover, we show that no fair selection algorithm returns always panels in the ex post core. Therefore, we ask for selection algorithms that achieve a multiplicative approximation of the ex post core with respect to the cost improvement of all individuals eligible to choose a different panel. We introduce an efficient selection algorithm, denoted as FairGreedyCapture, that is fair and is in the 6-approximate ex post core for *every* value of $q \in [k]$, concurrently. Our algorithm builds on an algorithm, called Greedy Capture, that was first introduced by Chen et al. [7]. Roughly speaking, the original algorithm partitions the n individuals into at most k parts and returns a deterministic representative from each part. In contrast to Greedy Capture, FairGreedyCapture assigns each individual a probability of being placed into one of k parts, each of size n/k, and then selects a random individual from each part. Then, leveraging Birkhoff's decomposition algorithm, we find a distribution over panels of size k, where each panel contains at least one representative from each part, and each individual is selected with a probability of k/n. We complement this result by showing that no fair selection algorithm provides an approximation better than 2 to the ex post core.

In Section 4, we ask if uniform selection is in the ex ante core, for all values of k and q. In particular, we define a selection algorithm to be in the *ex ante core* if, for any panel P, the expected number of individuals who feel more represented by P than panels chosen from the selection algorithms is less than $|P|/n \cdot k$. As before, we define a multiplicative approximation with respect to the cost improvement and we show that uniform selection provides an approximation of 4 to the ex ante core, while no fair selection algorithm provides an approximation better than 2 to it.

In Section 5, we explore the question of whether, given a panel P, there is any way to determine if it satisfies an approximation of the ex post core for a value of q. This can be useful when a panel has been sampled using a selection algorithm that does not provide any guarantees for the ex post core. We show that given a panel P, we can approximate, in polynomial time, how much it violates the core up to constants.

Finally, in Section 6, we empirically evaluate the approximation of uniform selection and FairGreedyCapture to the ex post core on constructed metrics derived from two demographic datasets. We notice that for large values of q, uniform selection achieves an approximation to the ex post similar to that of FairGreedyCapture. However, for smaller values of q, when the individuals form cohesive parts, uniform selection has often unbounded approximation.

1.3 Related Work

Ebadian et al. [11] recently considered the same question of measuring the representation that a panel or a selection algorithm achieves in a rigorous way. As we mentioned above, they also assume the existence of a representative metric space and use the distance of the q-th closest representative in the panel to measure to what degree a panel represents an individual. However, they use the social cost (i.e. the sum of individual costs) to measure how much a panel represents the whole population. In the full version [12], we show that this measure of representation may fail to achieve the idea of proportional representation.

As we discussed above, a method that is used in practice for enforcing representation is by setting quotas over features. However, a problem that appears is that only a few people volunteer to participate in a decision panel. As a result, the representatives are selected from a pool of volunteers which usually does not reflect the composition of the population, since for example highly educated people are usually more willing to participate in a decision panel than less educated people. Flanigan et al. [17] proposed selection algorithms that, given a biased pool of volunteers, find distributions that maximize the minimum selection probability of any volunteer over panels that satisfy the desired quotas. In this work, similar to Ebadian et al. [11] and Benadè et al. [4], we focus on the pivotal idea of a sortition based democracy that relies on sampling representatives directly from the underlying population [22]. However, later, we discuss how our approach can be modified for being applied in biased pools of volunteers. Benadè et al. [4] focused on the idea of stratified sampling and asked how this strategy may affect the variance of the representation of unknown groups. Flanigan et al. [19] studied how the selection algorithms can become transparent as well. In a more recent work, Flanigan et al. [20] studied the manipulability of different selection algorithms, i.e the incentives of individuals to misreport their features.

The representation of individuals as having an ideal point in a metric space has its roots to the spatial model of voting [1, 13]. The idea of using the core as a notion of proportional representation in a metric space was first introduced by Chen et al. [7], and later revisited by Micha and Shah [28], in a clustering setting. In our model, the clustering setting corresponds to the case where $q = 1$ and there is no fairness constraint. Therefore, only the negative results apply to our setting. Proportional representation in clustering has also been studied by Aziz et al. [3] and Kalayci et al. [23]. The definition by Aziz et al. [3] is quite similar to the core, with the basic difference being that each dense group explicitly requires a sufficient number of representatives. Kalayci et al. [23] consider a version of the core where an agent's cost for the panel is the sum of the distance of each representative, and a group is incentivized to deviate to another solution if the overall group can reduce the sum of costs. A drawback of both the definition of the core we use in this paper and Greedy Capture, which was mentioned by Aziz et al. [3] and Kalayci et al. [23], is that a dense group might end up being represented by just one individual. We stress that while our notion of the core does not explicitly account for this problem, FairGreedyCapture does not suffer from this weakness, since from each dense region returns a proportional number of representatives.

Proportional representation through core has been extensively studied in the context of multiwinner elections as well [2, 15, 16, 26]. The problem of selecting a representative panel can be framed as a committee election problem, where the candidates are drawn from the same pool as the voters. While in these works, the voters and the candidates do not lie in a metric space, but instead the voters hold rankings over candidates, in our model, the rankings could derive from the underlying metric space. Due to impossibility results [8], relaxations of the core have been studied. The ex ante core, as defined here, was introduced by Cheng et al. [8]. They show that, without the fairness constraint, the ex ante core can be guaranteed. In this work, we show that by imposing this fairness constraint, an approximation to the ex ante q-core better than 2 is impossible, for all $q \in [k-1]$.

2 PRELIMINARIES

For $t \in \mathbb{N}$, let $[t] = \{1, \dots, t\}$. We denote the population by $[n]$. A panel P is defined as a subset of the population. The n individuals are embedded in an underlying *representation metric space* with a distance function $d : [n] \times [n] \to \mathbb{R}_{\geq 0}$, where $d(i, j)$ quantifies the dissimilarity between individuals i and j. The distance function d is symmetric, meaning that $d(i, j) = d(j, i)$, and it satisfies the triangle inequality, i.e., $d(i, j) \leq d(i, \ell) + d(\ell, j)$. An instance of our problem is fully specified by the set of individuals, the pairwise distances among them, and an integer $k \in \mathbb{N}$ denoting the desired panel size.

We consider a class of cost functions to measure the cost of an individual i within a panel P. For $q \in [k]$, we define the *q-cost* of i for P as the distance to her q-th closest member in the panel, denoted by $c_q(i, P; d)$. When $q = 1$, the cost of an individual is equal to her distance from her closest representative in the panel, and for $q = k$, the cost is equal to her distance from her furthest representative in the panel. We denote by $\text{top}_q(i, P; d)$ the set of the q closest representatives of i in a panel P (with ties broken arbitrarily). Additionally, $B(i, r; d)$ represents the set of individuals captured from a ball centered at i with a radius of r, i.e., $B(i, r; d) = \{i' \in [n] : d(i, i') \leq r\}$. We may omit d from the notation when clear from the context.

A *selection algorithm*, denoted by $\mathcal{A}_k$, is parameterized by k and takes as input the metric d and outputs a distribution over all panels of size k. We say that a panel is in the support of $\mathcal{A}_k$, if it is implemented with positive probability under the distribution that $\mathcal{A}_k$ outputs. We pay special attention to the *uniform selection* algorithm, denoted by $\mathcal{U}_k$, that always outputs a uniform distribution over all the subsets of the population of size k.

Fairness. As mentioned above, one of the appealing properties of uniform selection is that each individual is included in the panel with the same probability. We call this property *fairness* and we say that a selection algorithm is *fair* if:

$$\forall i \in [n], \quad \Pr_{P \sim \mathcal{A}_k}[i \in P] = k/n.$$

Core. Another appealing property of sortition is proportional representation. Here, we utilize the concept of the core to evaluate the proportional representation of a panel and, by extension, of a selection algorithm. To formalize this, we introduce the *α-q-preference count* of a panel P relative to another panel P', denoted as $V_q(P, P', \alpha)$. For $\alpha \geq 1$, this quantity represents the number of individuals whose q-cost under P exceeds α times their q-cost under P', i.e.,

$$V_q(P, P', \alpha) = |\{i \in [n] : c_q(i, P) > \alpha \cdot c_q(i, P')\}|.$$

A panel P is *in the α-q-core*, if for any panel P', $V_q(P, P', \alpha) < |P'| \cdot n/k$. For $\alpha = 1$, we say that the panel is in the q-core. We define α-q-core for $\alpha > 1$, since even when $q = 1$, a panel in the exact q-core is not guaranteed to exist [7, 28]. Intuitively, $V_q(P, P', \alpha)$ is the number of individuals who feel α times better represented in P' than in P. If this number is sufficiently large, the corresponding set of individuals is eligible to deviate to P'. A panel P is in the α-q-core if no such set of individuals exist.

Ex post q-core. A selection algorithm $\mathcal{A}_k$ is in the *ex post α-q-core* (or ex post q-core, for $\alpha = 1$) if every panel P in the support of $\mathcal{A}_k$ is in the α-q-core, i.e., for all P drawn from $\mathcal{A}_k$ and all P',

$$V_q(P, P', \alpha) < |P'| \cdot n/k.$$

Ex ante q-core. A selection algorithm $\mathcal{A}_k$ is in the *ex ante α-q-core* (or ex ante q-core, for $\alpha = 1$) if for all P':

$$\mathbb{E}_{P\sim\mathcal{A}_k}[V_q(P, P', \alpha)] < |P'| \cdot n/k.$$

The idea of requiring a core-like property over the expected number of preference counts was introduced by Cheng et al. [8] in a multi-winner election setting. Essentially, it states that for any panel P', if, for any realized panel P, we count the number of individuals that reduce their cost by a multiplicative factor of at least α under P', in expectation, this number is less than $|P'| \cdot n/k$. Therefore, in expectation, they are not eligible to choose it.

It is easy to see that ex post α-core implies ex ante α-core, since if for each P in the support of a distribution that $\mathcal{A}_k$ returns and each P', it holds that $V_q(P, P', \alpha) < |P'| \cdot n/k$, then $\mathbb{E}_{P\sim\mathcal{A}_k}[V_q(P, P', \alpha)] < |P'| \cdot n/k$.

3 FAIRNESS AND EX POST CORE

In this section, we investigate if there are selection algorithms that are fair, and in addition, provide a constant approximation to the ex post q-core. Unsurprisingly, uniform selection may fail to provide any bounded approximation to the ex post q-core for $q \in [k-1]$.[1] This happens because each panel has a nonzero probability of selection, and there may exist panels with arbitrarily large violations of the q-core objective.

Theorem 1. *For any $q \in [k-1]$ and $\lfloor n/k \rfloor \geq k$, there exists an instance such that uniform selection is not in the ex post α-q-core for any bounded α.*

All missing proofs can be found in the full version [12].

Therefore, we ask: For every q, is there any selection algorithm that keeps the fairness guarantee of uniform selection and ensures that every panel in its support is in the constant approximation of the q-core? We answer this positively.

We present a selection algorithm, called FairGreedyCapture$_k$, that is fair and in the ex post 6-q-core, for *every* $q \in [k]$. We highlight that the algorithm does not need to know the value of q. Our algorithm leverages the basic idea of the Greedy Capture algorithm introduced by Chen et al. [7], which returns a panel in the $(1 + \sqrt{2}) \cong 2.42$-approximation of the 1-core. Briefly, Greedy Capture begins with an empty panel and gradually expands a ball around each individual at a uniform rate. When a ball captures at least $\lceil n/k \rceil$ individuals for the first time, we "open" the ball by selecting its center for the panel and removing all captured individuals from further consideration. The process continues, expanding balls around all individuals — including those already opened. The algorithm keeps growing balls on all individuals, including the opened balls. As these opened balls grow, any newly captured individuals are immediately disregarded. This algorithm is deterministic and need not satisfy fairness. Additionally, the final panel may contain fewer than k individuals.

[1] For $q = k$, we show in the full version that all panels lie in the 2 approximation of the k-core; hence, any algorithm including uniform selection provides an ex post 2-k-core.

ALGORITHM 1: FairGreedyCapture$_k$

```
Input: [n], d
Output: P_ℓ and λ_ℓ, for ℓ ∈ [L], where each P_ℓ represents a
        panel of size k and λ_ℓ is its probability of selection
/* Create a (k/n)-fractional allocation by
   distributing a k/n fraction of each individual
   among k balls, such that each ball contains a
   total fractional amount equal to 1.            */
X ← [0]^{k×n}; δ ← 0; j ← 1; {y_i ← k/n}_{i∈[n]};
while Σ_{i∈[n]} y_j > 0 do
    Gradually increase δ until the following condition holds;
    while ∃i ∈ [n], such that Σ_{i'∈B(i,δ)} y_{i'} ≥ 1 do
        while X_j = Σ_{i∈[n]} X_{j,i} < 1 do
            Pick i' ∈ B(i, δ) with y_{i'} > 0;
            X_{j,i'} ← min(1 − X_j, y_{i'});
            y_{i'} ← y_{i'} − X_{j,i'};
        end
        j ← j + 1;
    end
end
/* Apply Birkhoff's decomposition                 */
X' ← [1/n]^{(n−k)×n};
Let Y = [ X ; X' ];
Compute a decomposition of Y = Σ_{ℓ=1}^{L} λ_ℓ Y^ℓ using the
  Birkhoff's decomposition (Theorem 2);
for ℓ = 1 to L do
    P_ℓ ← {i ∈ [n] | Y^ℓ_{j,i} = 1 for some j ≤ k}
end
return distribution over L panels {P_ℓ}_{ℓ∈[L]} where P_ℓ is
  selected with probability λ_ℓ
```

At a high level, FairGreedyCapture$_k$ (see Algorithm 1) operates as follows: it greedily opens k balls using the basic idea of the Greedy Capture algorithm, ensuring each ball contains sufficiently many individuals. In contrast to Greedy Capture, which selects ball centers as representatives, our algorithm probabilistically selects precisely one individual from each of the k balls.

Before, we describe the algorithm in more detail, we define a (k/n)-fractional allocation as a non-negative $k \times n$ matrix $X \in [0, 1]^{k\times n}$ where entries in each row sums to 1 and entries in each column sum to k/n, i.e., for each $i \in [n]$, $\sum_{j\in[k]} X_{j,i} = k/n$, and for each $j \in [k]$, $\sum_{i\in[n]} X_{j,i} = 1$. The algorithm, during its execution, generates a (k/n)-*fractional allocation* X of individuals in $[n]$ into

k balls, where $X_{j,i}$ denotes the fraction of individual i assigned to ball j. We say that an individual i is assigned to ball j, if $X_{j,i} > 0$. An individual can be assigned to more than one ball.

The (k/n)-fractional allocation X is generated as follows. Denote the unallocated part of each individual i by y_i. Start with $y_i = k/n$. This corresponds to the fairness criterion that we allocate a k/n probability of selection to each individual. Algorithm 1 grows a ball around every individual in $[n]$ at the same rate. Suppose a ball captures individuals whose combined unallocated parts sum to at least 1. Then, we open this ball and from individuals i' captured by this ball with $y_{i'} > 0$, we arbitrarily remove a total mass of exactly 1 and assign it to the ball. This can be done in various ways, e.g., greedily pick an individual i' with positive $y_{i'}$ and allocate $\min\{1 - \sum_{i\in[n]} X_{j,i}, y_{i'}\}$ fraction of it to the corresponding row (i.e. ball). This procedure terminates when the k/n fraction of each individual is fully allocated. Note that since each time a ball opens, a total mass of 1 is deducted from y_i-s and, for each $i \in [n]$, y_i starts with a fraction of k/n, exactly k balls are opened.

Sampling panels from the (k/n)-fractional allocation. Next, we show a method of decomposing X, the (k/n)-fractional allocation, to a distribution over panels of size k that each contain at least one representative from each ball. We employ the Birkhoff's decomposition [5]. This theorem applies over square matrices that are bistochastic. A matrix is bistochastic if every entry is nonnegative and the elements in each of its rows and columns sum to 1.

Theorem 2 (Birkhoff-von Neumann). *Let Y be a $n\times n$ bistochastic matrix. There exists a polynomial time algorithm that computes a decomposition $Y = \sum_{\ell=1}^{L} \lambda_\ell Y^\ell$, with $L \le n^2 - n + 2$, such that for each $\ell \in [L]$, $\lambda_\ell \in [0,1]$, Y^ℓ is a permutation matrix and $\sum_{\ell=1}^{L} \lambda_\ell = 1$.*

We cannot directly apply the theorem above, since the (k/n)-fractional allocation X is not bistochastic nor a square matrix. However, we can complete X into a square matrix $Y = \begin{bmatrix} X \\ X' \end{bmatrix}$ by adding $n-k$ rows $X' = [1/n]^{(n-k)\times n}$ where all entries are $1/n$. Note that the resulting matrix Y is bistochastic. Indeed, each row of both X and X' sums to 1 by their definition; further, as each column of X sums to k/n and that it is followed by $n-k$ of $1/n$ entries in X', the columns also sum to 1. Note that there are various choices of X' that makes Y a bistochastic matrix, but here we use the uniform matrix for simplicity. Then, the algorithm applies Theorem 2 and computes the decomposition $Y = \sum_{\ell=1}^{L} \lambda_\ell Y^\ell$. For each permutation matrix Y^ℓ, we create a panel P_ℓ consisting of the individuals that have been assigned to the first k rows, i.e. P_ℓ contains all i-s with $Y^\ell_{j,i} = 1$ for some $j \le k$. Finally, the algorithm returns the distribution that selects each panel P_ℓ with probability equal to λ_ℓ.

To prove that FairGreedyCapture$_k$ is fair and ex post $O(1)$-q-core, we need the next two lemmas.

Lemma 1. *Let $S \subseteq [n]$, P' be a panel, and $m = \lfloor |P'|/q \rfloor$.*

1. *There exists a partitioning of S into m disjoint sets $T_1,\dots,T_m$ and an individual $i^*_\ell \in T_\ell$ such that for all $\ell \in [m]$ and $i \in T_\ell$, $c_q(i,P') \le c_q(i^*_\ell,P')$ and $\mathrm{top}_q(i,P') \cap \mathrm{top}_q(i^*_\ell,P') \ne \emptyset$.*
2. *There exists a partitioning of S into m disjoint sets $T_1,\dots,T_m$ and an individual $i^*_\ell \in T_\ell$ such that for all $\ell \in [m]$ and $i \in T_\ell$, $c_q(i,P') \ge c_q(i^*_\ell,P')$ and $\mathrm{top}_q(i,P') \cap \mathrm{top}_q(i^*_\ell,P') \ne \emptyset$.*

Lemma 2. *For any panel P and any $i, i' \in [n]$, it holds that $c_q(i,P) \le d(i,i') + c_q(i',P)$.*

Proof of Lemma 2. Consider a ball centered at i' with radius $c_q(i',P)$. This ball contains at least q representatives of P. Hence, $c_q(i,P)$ is less than or equal to the distance of i to one of the q representatives that are included in $B(i', c_q(i',P))$ which is at most $d(i,i') + c_q(i',P)$. □

Now, we are ready to prove the next theorem.

Theorem 3. *For every $q \in [k]$, FairGreedyCapture$_k$ is fair and in the ex post 6-q-core.*

Proof. Seeing that the algorithm is fair is straightforward. For a matrix A, let $A[1:k,:]$ be the submatrix induced by keeping its first k rows. First, note that for each panel P^ℓ we choose the individuals that have been assigned to $Y^\ell[1:k,:]$ and second, recall that $Y[1:k,:] = X$. The fairness of the algorithm follows by the facts that $Y[1:k,:] = X = \sum_{\ell=1}^{L} \lambda_\ell Y^\ell[1:k,:]$ and for each $i \in [n]$, $\sum_{j=1}^{k} X_{j,i} = k/n$.

We proceed by showing that FairGreedyCapture$_k$ is in the ex post 6-q-core, for all $q \in [k]$. First, note that if an individual i is assigned to a ball j in some Y^ℓ, then we must have $X_{j,i} > 0$. Now, since each individual $i \in [n]$ is assigned to a ball $j \in [k]$ in the permutation, we get that at least one individual is selected from each ball.

Let P be any panel that the algorithm may return. Suppose for contradiction that there exists a panel P' such that $V_q(P,P',6) \ge |P'|\cdot n/k$. Thus, there exists $S \subseteq [n]$ with $|S| \ge |P'|\cdot n/k$ such that

$$\forall i \in S, \qquad c_q(i,P) > 6\cdot c_q(i,P'). \tag{1}$$

Let $T_1,\dots,T_m$ be a partition of S with respect to P', as given in the first part of Lemma 1. Since $m \le \lfloor |P'|/q \rfloor$ and $|S| \ge |P'|\cdot n/k$, we conclude that there exists a part, say T_ℓ, that has size at least $q\cdot n/k$. From Lemma 1, we know that there exists $i^*_\ell \in T_\ell$ such that for each $i \in T_\ell$ it holds that $c_q(i,P') \le c_q(i^*_\ell,P')$ and $\mathrm{top}_q(i,P') \cap \mathrm{top}_q(i^*_\ell,P') \ne \emptyset$. Therefore, we can conclude that for each $i \in T_\ell$, $d(i^*_\ell, i) \le 2\cdot c_q(i^*_\ell,P')$, as follows: Pick an arbitrary representative in $\mathrm{top}_q(i,P') \cap \mathrm{top}_q(i^*_\ell,P')$ and denote it as r_i. Then,

$$d(i,i^*_\ell) \le d(i,r_i) + d(r_i,i^*_\ell) \le c_q(i,P') + c_q(i^*_\ell,P') \le 2\cdot c_q(i^*_\ell,P').$$

This implies that the ball centered at i^*_ℓ with a radius of $2\cdot c_q(i^*_\ell,P')$ captures all individuals in T_ℓ.

Now, consider all the balls that FairGreedyCapture$_k$ opens and contain individuals from T_ℓ. Since $|T_\ell| \ge q\cdot n/k$ and each ball is assigned a total fraction of 1, there are at least q such balls. Next, we claim that at least q of them have radius at most $2\cdot c_q(i^*_\ell,P')$. Suppose for contradiction that at most $q-1$ of them have radius at most $2\cdot c_q(i^*_\ell,P')$. This means that a total fraction of at least 1 from individual in T_ℓ is assigned to balls with radius strictly larger than $2\cdot c_q(i^*_\ell,P')$. However, the ball centered at i^*_ℓ with radius $2\cdot c_q(i^*_\ell,P')$ would have captured this fraction, and therefore we reach a contradiction.

Next, denote with $B_1,\dots,B_q$, q balls that are opened, and each contain individuals from T_ℓ and have radius at most $2\cdot c_q(i^*_\ell,P')$. Due to the definition of FairGreedyCapture$_k$, each panel that is returned, contains at least one representative from each ball.

Therefore, each ball B_j contains at least one representative, denoted by r_j. Now, note that since each B_j contains at least one individual from T_ℓ, denoted by i_j, we have that

$$\forall j \in [q], \quad d(i^*_\ell, r_j) \le d(i^*_\ell, i_j) + d(i_j, r_j) \le 6 \cdot c_q(i^*_\ell, P'),$$

where the first inequality follows from the triangle inequality and the last inequality follows from the facts that for each $i \in T_\ell$, $d(i^*_\ell, i) \le 2 \cdot c_q(i^*_\ell, P')$, and each B_j has radius at most $2 \cdot c_q(i^*_\ell, P')$ and both i_j and r_j belong to this ball. Therefore, there are at least q representatives in P that have distance at most $6 \cdot c_q(i^*_\ell, P')$ from i^*_ℓ. But then, $c_q(i^*_\ell, P) \le 6 \cdot c_q(i^*_\ell, P')$ which contradicts Eq. (1). ◻

As discussed, the ex post α-q-core implies the ex ante α-q-core which means that FairGreedyCapture$_k$ is also in the ex ante 6-q-core for all $q \in [k]$. In the next section, we show that no fair algorithm provides an approximation better than 2 to the ex ante q-cost, for any q. Therefore, we get that no fair selection algorithm provides an approximation better than 2 to the ex post q-core either. Thus, FairGreedyCapture$_k$ is optimal within a factor of 3.

3.1 Ex Post Core and Quotas over Features

In the introduction, we discussed a common approach used to ensure proportional representation, which involves setting quotas based on individual or groups of features. For instance, a quota might mandate that at least 45% of representatives are female. While the concept of the core aims to achieve proportional representation across intersecting features, it may not guarantee the same across individual features. This raises the question of whether it's possible to achieve both types of representation to the degree that is possible. We argue that this is feasible and show how the core requirement can be translated into a set of quotas.

FairGreedyCapture$_k$ generates k balls, with each individual assigned to one or more balls. A sufficient condition to achieve an ex post $O(1)$-q-core is to have at least one representative from each ball. This condition can be transformed into quotas by introducing an additional feature, b_i, for each individual i, indicating the balls they belong to. We then impose quotas ensuring that for each ball $j \in [k]$, the panel includes at least one individual belonging to that ball. In other words, we can think of each ball as a subpopulation from which we want to draw a representative. We can then utilize the methods proposed by Flanigan et al. [17] to identify panels that meet these quotas, along with others as much as possible, while maximizing fairness. Note that framing proportional representation as a set of quotas also allows us to use the algorithms of the aforementioned paper for sampling from a biased pool of volunteers.

4 UNIFORM SELECTION AND EX ANTE CORE

We have already discussed that uniform selection fails to provide any reasonable approximation to the ex post q-core, for almost all values of q. However, as we mentioned in the introduction, it seems to satisfy the ex ante q-core, at least when k is very large. In this section, we ask whether indeed uniform selection satisfies a constant approximation of the ex ante q-core, in a rigorous way, for all values of q and k. We show that uniform selection is in the ex ante 4-q-core, for every q.[2] The main reason is that, for $q = k$, it suffices to show that the grand coalition does not deviate ex ante.

[2] In fact, for $q = k$, uniform selection is in the ex ante k-core (see the full version).

ALGORITHM 2: Auditing Algorithm

Input: P, $[n]$, d, k, q,
Output: $\hat{\alpha}$
for $j \in [n]$ **do**
  $\hat{P}_j \leftarrow \{j\} \cup \{q-1 \text{ closest neighbors of } j\}$
  $\hat{\alpha}_j \leftarrow \lceil qn/k \rceil$ largest value among $\left\{ c_q(i,P)/c_q(i,\hat{P}_j) \right\}_{i \in [n]}$
end
return $\hat{\alpha} \leftarrow \arg\max_{j \in [n]} \hat{\alpha}_j$

Since each panel is selected with non-zero probability, the marginal probabilities of deviation is strictly less than one, and the ex ante k-core is satisfied.

Theorem 4. *For any $q \in [k]$, uniform selection is in the ex ante 4-q-core. That is, for any panel P', $\mathbb{E}_{P \sim \mathcal{U}_k} \left[V_q(P, P', 4) \right] < |P'| \cdot n/k$.*

In the next theorem, we show that for any $q < k$, no selection algorithm that is fair, is guaranteed to achieve ex ante α-q-core with $\alpha < 2$. Hence uniform selection is optimal up to a factor of 2.

Theorem 5. *For any $q \in [k-1]$, when $n \ge 2k^2/(k-q)$, there exists an instance such that no selection algorithm that is fair is in the ex ante α-q-core with $\alpha < 2$.*

5 AUDITING EX POST CORE

In this section, we turn our attention to the following question: Given a panel P, how much does it violate the q-core, i.e. what is the maximum value of α such that there exists a panel P' with $V_q(P, P', \alpha) \ge |P'| \cdot n/k$? This auditing question can be very useful in practice for measuring the proportional representation of a panel formed using a method that does not guarantee any panel to be in the approximate core, such as uniform selection.

Chen et al. [7] ask the same question for the case where the cost of an individual for a panel is equal to her distance from her closest representative in the panel, i.e. when $q = 1$. In this case, it suffices to restrict our attention to panels of size 1, which are subsets of the population that individuals may prefer to be represented by. In other words, given a panel P, we can simply consider every individual as a potential representative and check if a sufficiently large subset of the population prefers this individual to be their representative over P. Thus, we can find the maximum α such that there exists P', with $V_q(P, P', \alpha) \ge n/k$ as follows: For each $j \in [n]$, calculate α_j which is equal to the $\lceil n/k \rceil$ largest value among the set $\{c_q(i,P)/c_q(i,\{j\})\}_{i \in [n]}$ containing the q-cost ratios of P to P'. Then, α is equal to the maximum value among all α_j's.

For $q > 1$, this question is more challenging. We approximate the maximum α by generalizing the above procedure as follows: For each $j \in [n]$, define $\hat{P}_j$ as the panel containing j and its $q-1$ closest neighbors. Compute $\hat{\alpha}_j$ as the $\lceil q \cdot n/k \rceil$ largest value in $\{c_q(i,P)/c_q(i,\hat{P}_j)\}_{i \in [n]}$. Finally, return the maximum value among all $\hat{\alpha}_j$'s as $\hat{\alpha}$. Algorithm 2 executes this procedure. We show that the maximum α such that there exists a panel P' with $V_q(P, P', \alpha) \ge |P'| \cdot n/k$ is at most $3 \cdot \hat{\alpha} + 2$.

Theorem 6. *There exists an efficient algorithm that for every panel P and $q \in [k]$ returns $\hat{\alpha}$-q-core violation that satisfies $\hat{\alpha} \le \alpha \le 3\hat{\alpha} + 2$, where α is the maximum amount of q-core violation of P.*

6 EXPERIMENTS

In previous sections, we examined uniform selection from a worst-case perspective and found that it cannot guarantee panels in the core for any bounded approximation ratio. But, what about the average case? In this section, we aim to address this question through empirical evaluations of both algorithms using real databases.

6.1 Datasets

In accordance with the methodology proposed by Ebadian et al. [11], we utilize the same two datasets used by the authors as a proxy for constructing the underlying metric space. These datasets capture various characteristics of populations across multiple observable features. It is reasonable to assume that individuals feel closer to others who share similar characteristics. Therefore, we construct a random metric space using these datasets.

Adult. The first is the Adult dataset, extracted from the 1994 Current Population Survey by the US Census Bureau and available on the UCI Machine Learning Repository under a CC BY 4.0 license [10, 24]. Our analysis focuses on five demographic features: `sex`, `race`, `workclass`, `marital.status`, and `education.num`. The dataset is comprised of 32,561 data points, each with a sample weight attribute (`fnlwgt`). We identify 1513 unique data points by these features and treat the sum of the weights associated with each unique point as a distribution across them.

ESS. The second dataset we analyze is the European Social Survey (ESS), available under a CC BY 4.0 license [29]. Conducted biennially in Europe since 2001, the survey covers attitudes towards politics and society, social values, and well-being. We used the ESS Round 9 (2018) dataset, which has 46,276 data points and 1451 features across 28 countries. On average, each country has around 250 features (after removing non-demographic and country-unrelated data), with country-specific data points ranging from 781 to 2745. Each ESS data point has a post-stratification weight (`pspwght`), which we use to represent the distribution of the data points. Our analysis focuses on the ESS data for the United Kingdom (ESS-UK), which includes 2204 data points.

6.2 Representation Metric Construction

In line with the work of Ebadian et al. [11], we apply the same approach to generate synthetic metric preferences, which are used to measure the dissimilarity between individuals based on their feature values. Our datasets consist of two types of features: *categorical* features (e.g. sex, race, and martial status) and *continuous* features (e.g. income). We define the distance between individuals i and j with respect to feature f as follows:

$$d(i,j;f) := \begin{cases} \mathbb{1}[f(i) \neq f(j)], & \text{if } f \text{ is a \textit{categorical} feature;} \\ \frac{|f(i)-f(j)|}{\max_{i',j'} |f(i')-f(j')|}, & \text{if } f \text{ is a \textit{continuous} feature,} \end{cases}$$

where the normalization factor for continuous features ensures that $d(i,j;f) \in [0,1]$ for all i, j, and f, and that the distances in different features are comparable. Next, we define the distance between two individuals as the weighted sum of the distances over different features, i.e. $d(i,j) = \sum_{f \in F} w_f \cdot d(i,j;f)$, where the weights w_f's are randomly generated. Each unique set of randomly generated feature weights results in a new representation metric.

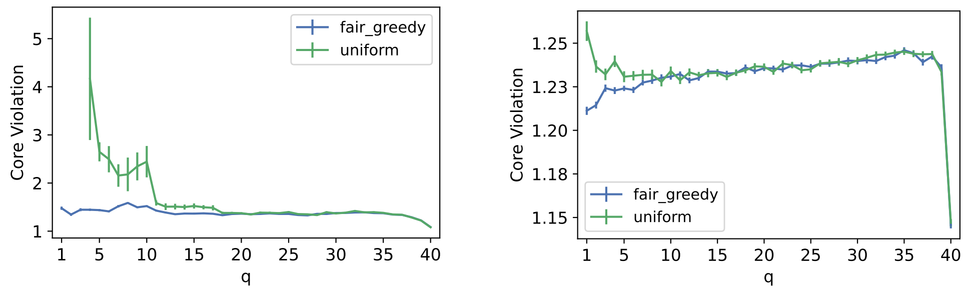
We generate 100 sets of randomly-assigned feature weights per dataset, calculate a representation metric for each set, and report the performance metrics averaged over 100 instances. Given that our datasets are samples of a large population (i.e. millions) and represented through a relatively small number of unique data points (i.e. few thousands), we assume that each data point represents a group of at least k people, which takes a maximum value of 40 in our study. To empirically measure ex post core violation, for each of the 100 instances, we sample one panel from an algorithm and compute the core violation using Algorithm 2. We note that this is not exactly equal to the worst-case core violation, but a very good approximation of it.

6.3 Results

6.3.1 Results for ex post core violation. In Adult dataset, we observe an unbounded ex post core violation for Uniform when $q \leq 4$. Specifically, for $q \in \{1, 2, 3\}$, we observed unbounded core violation in 84%, 9%, and 36% of the instances respectively. This happens since ~8.3% of the population is mapped to a single data point and that Uniform fails to select q individuals from this group. When $q \leq 3$, we have $q/k \leq 8.4\%$, and this cohesive group is entitled to select at least q members of the panel from themselves, which results in q-cost of 0 for them and an unbounded violation of the core. However, FairGreedyCapture captures this cohesive group and selects at least q representatives from them. Furthermore, we see significantly higher ex post core violation for Uniform compared to FairGreedyCapture for smaller values of q (up to 12) and comparable performance for larger values of q. This is expected as FairGreedyCapture tends to behave more similarly to Uniform as q increases because it selects from fewer yet larger groups ($\lfloor k/q \rfloor + 1$ groups of size qn/k).

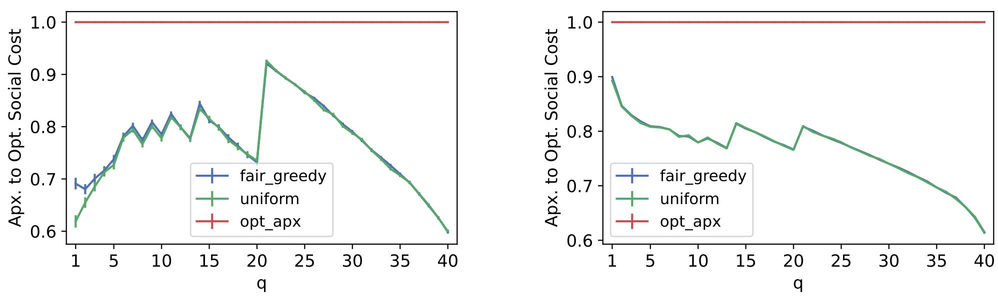
We observe a similar pattern in ESS-UK that Uniform obtains worse ex post core violations when q is smaller and similar performance as FairGreedyCapture for larger values of q. However, in contrast to Adult, we do not observe similar unbounded violations for Uniform in ESS-UK. The reason is that ESS-UK consists of 250 features (compared to the 5 we used from Adult) and any data points represent at most 0.2% of the population. Thus, no group is entitled to choose enough representatives from their own to significantly improve their cost or make it 0. The decline in core violation for $q = k$ happens as it measures the minimum improvement in cost over the whole population, which is more demanding than lower values of q. Lastly, FairGreedyCapture performs consistently for all values of q and achieves an ex post core violation less than 1.6 and 1.25 in Adult and ESS-UK respectively.

6.3.2 Evaluating approximation to optimal social cost. As we mentioned in the introduction, Ebadian et al. [11] use a different approach to measure the representativeness of a panel by considering the social cost (sum of q-costs) over a panel. In particular, they define the representativeness of an algorithm as the worst-case ratio between the optimal social cost and the (expected) social cost obtained by the algorithm. Ebadian et al. [11], in their empirical analysis, measure the average approximation to the optimal social cost of an algorithm $\mathcal{A}$ over a set of instances $\mathcal{I}$, defined as



(a) Adult **(b) ESS-UK**

Figure 1: Ex post core violation of FairGreedyCapture and Uniform with $k = 40$.



(a) Adult **(b) ESS-UK**

Figure 2: Approximation to the optimal social cost of FairGreedyCapture and Uniform with with $k = 40$.

$\frac{1}{|\mathcal{I}|}\sum_{I\in\mathcal{I}} \frac{\min_P \sum_{i\in[n]} c_q(i,P)}{\sum_{i\in[n]} c_q(i,\mathcal{A}(I))}$. Since finding the optimal panel is a hard problem and the dataset and panel sizes are large, Ebadian et al. [11] use a proxy for the minimum social cost, specifically, an implementation of the algorithm of Kumar and Raichel [25] for the fault-tolerant k-median problem that achieves a constant factor approximation of the optimal objective — which is equivalent to minimizing the q-social cost. We use the same approach and report the average approximation to the optimal social cost of FairGreedyCapture and Uniform.

In Figure 2, the reader can see the performance of the two different algorithms over this objective. For ESS-UK, we observe a similar behaviour from the two algorithms, while for Adult, FairGreedyCapture outperforms Uniform for $q \in [3]$, which is again due to FairGreedyCapture capturing the cohesive group. All considered, we observe that FairGreedyCapture can maintain at least the same level or even better optimal social cost approximation as Uniform would, while achieving significantly better empirical core guarantees in the two datasets.

7 DISCUSSION

This work introduces a notion of proportional representation, called the core, within the context of sortition. The core serves as a measure to ensure proportional representation across intersectional features. While uniform selection achieves an ex ante $O(1)$-q-core, it fails to provide a reasonable approximation to the ex post q-core. To overcome this limitation, we propose a selection algorithm, FairGreedyCapture, that retains the benefits of uniform selection — namely, fairness and an ex ante $O(1)$-q-core — while also satisfying the ex post $O(1)$-q-core criterion.

It is worth emphasizing that the limitations of uniform selection in satisfying ex post guarantees stem from the possibility of selecting non-proportionally representative panels. In the full version, we explore a natural variation where the core property is required to hold over the expected q-costs of panels chosen from a selection algorithm. We demonstrate that this variation is incomparable with the ex post q-core. More importantly, uniform selection fails to offer any meaningful multiplicative approximation for this variation.

Several directions remain for future work. One immediate challenge is closing the gaps between our lower and upper bounds for both the ex ante and ex post core. Furthermore, Micha and Shah [28] show that for $q = 1$, Greedy Capture [7] offers better guarantees in Euclidean space. This raises an interesting question: can FairGreedyCapture achieve better guarantees when the metric d is defined by common distance functions such as L^2, L^1, or L^∞ norms? Lastly, in this work, we assume that all individuals share a common metric space. However, in practice, individuals may prioritize different features. Ensuring representation across subjective metric spaces is an intriguing direction for future research.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] K. Arrow. 1990. *Advances in the spatial theory of voting*. Cambridge University Press.

[2] H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh. 2017. Justified representation in approval-based committee voting. *Social Choice and Welfare* 48, 2 (2017), 461–485.

[3] H. Aziz, B. E Lee, and S. M. Chu. 2023. Proportionally Representative Clustering. *arXiv preprint arXiv:2304.13917* (2023).

[4] G. Benadè, P. Gölz, and A. D. Procaccia. 2019. No Stratification Without Representation. In *Proceedings of the 20th ACM Conference on Economics and Computation (EC)*. 281–314.

[5] G. Birkhoff. 1946. Three observations on linear algebra. *Univ. Nac. Tacuman, Rev. Ser. A* 5 (1946), 147–151.

[6] I. Caragiannis, N. Shah, and A. A. Voudouris. 2022. The Metric Distortion of Multiwinner Voting. *Artificial Intelligence* 313 (2022), 103802.

[7] X. Chen, B. Fain, L. Lyu, and K. Munagala. 2019. Proportionally fair clustering. In *International Conference on Machine Learning*. 1032–1041.

[8] Y. Cheng, Z. Jiang, K. Munagala, and K. Wang. 2020. Group fairness in committee selection. *ACM Transactions on Economics and Computation (TEAC)* 8, 4 (2020), 1–18.

[9] V Conitzer, R Freeman, N Shah, and J. W. Vaughan. 2019. Group Fairness for the Allocation of Indivisible Goods. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. 1853–1860.

[10] D. Dua and C. Graff. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

[11] S. Ebadian, G. Kehne, E. Micha, A. D. Procaccia, and N. Shah. 2022. Is Sortition Both Representative and Fair?. In *Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS)*. 25720–25731.

[12] S. Ebadian and E. Micha. 2024. Boosting Sortition via Proportional Representation. arXiv:2406.00913 [cs.GT] https://arxiv.org/abs/2406.00913

[13] J. M. Enelow and M. J. Hinich. 1984. *The spatial theory of voting: An introduction*. CUP Archive.

[14] F. Engelstad. 1989. The assignment of political office by lot. *Social Science Information* 28, 1 (1989), 23–50.

[15] B. Fain, K. Munagala, and N. Shah. 2018. Fair allocation of indivisible public goods. In *Proceedings of the 19th ACM Conference on Economics and Computation (EC)*. 575–592.

[16] P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon. 2017. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice* 74, 2017 (2017), 27–47.

[17] B. Flanigan, P. Gölz, A. Gupta, B. Hennig, and A. D. Procaccia. 2021. Fair Algorithms for Selecting Citizens' Assemblies. *Nature* 596 (2021), 548–552.

[18] B. Flanigan, P. Gölz, A. Gupta, and A. D. Procaccia. 2020. Neutralizing Self-Selection Bias in Sampling for Sortition. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*. 6528–6539.

[19] B. Flanigan, G. Kehne, and A. D. Procaccia. 2021. Fair Sortition Made Transparent. In *Proceedings of the 35th Annual Conference on Neural Information Processing Systems (NeurIPS)*. 25720–25731.

[20] B. Flanigan, J. Liang, A. D. Procaccia, and S. Wang. 2024. Manipulation-Robust Selection of Citizens' Assemblies. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*.

[21] A. Gąsiorowska. 2023. Sortition and its Principles: Evaluation of the Selection Processes of Citizens' Assemblies. *Journal of Deliberative Democracy* 19, 1 (2023).

[22] J. Gastil and E. O. Wright (Eds.). 2019. *Legislature by Lot: Transformative Designs for Deliberative Governance*. Verso, London.

[23] Y. H. Kalayci, D. Kempe, and V. Kher. 2024. Proportional Representation in Metric Spaces and Low-Distortion Committee Selection. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*. Article 1093, 9 pages.

[24] R. Kohavi and B. Becker. 1996. Adult data set. *UCI machine learning repository* 5 (1996), 2093.

[25] N. Kumar and B. Raichel. 2013. Fault Tolerant Clustering Revisited. In *Proceedings of the 25th Canadian Conference on Computational Geometry (CCCG) Waterloo, Ontario, Canada*. Carleton University, Ottawa, Canada, Article 36, 6 pages.

[26] M. Lackner and P. Skowron. 2023. *Multi-winner voting with approval preferences*. Springer, Cham.

[27] B. Martin and L. Carson. 1999. *Random selection in politics*. Bloomsbury Publishing.

[28] E. Micha and N. Shah. 2020. Proportionally fair clustering revisited. In *47th International Colloquium on Automata, Languages, and Programming (ICALP)*. Schloss Dagstuhl, Saarbrücken, Germany, 85:1–85:16.

[29] ESS Round 9: European Social Survey (2021): ESS-9 2018 Documentation Report. 2021. Edition 3.1. Bergen, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC. (2021). https://doi.org/10.21338/NSD-ESS9-2018

[30] D. Van Reybrouck. 2018. *Against Elections: The Case for Democracy*. Seven Stories Press, New York.

[31] A. Vergne. 2018. Citizens' Participation Using Sortition: A practical guide to using random selection to guarantee diverse democratic participation. http://aei.pitt.edu/102678/