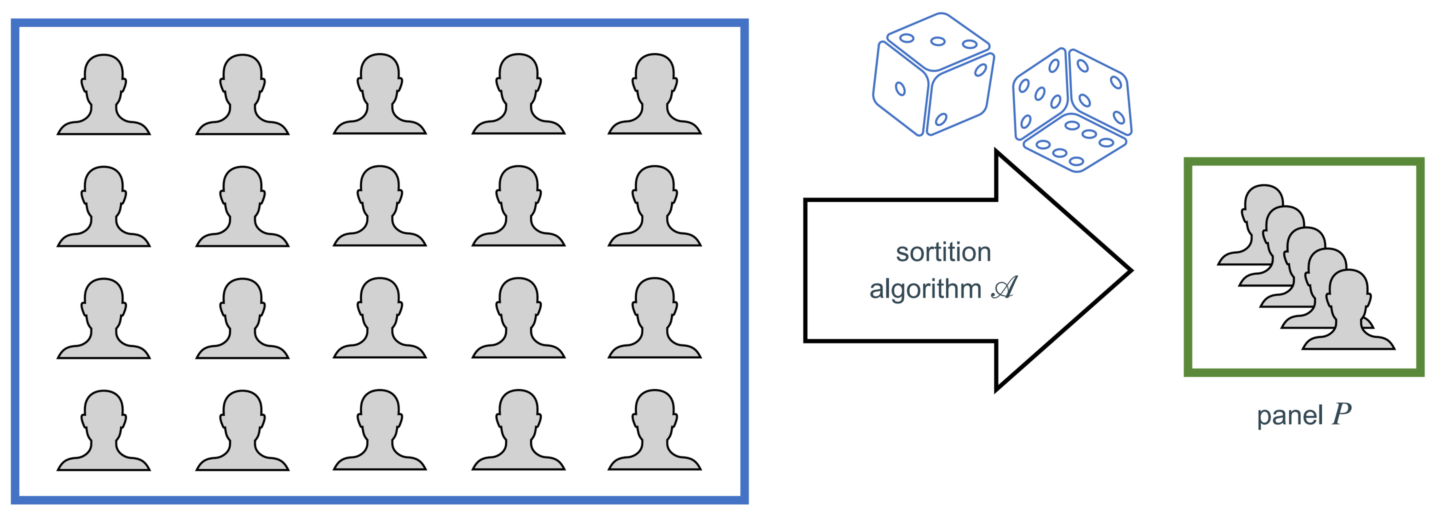
Is Sortition Both Representative and Fair?

Soroush Ebadian, Gregory Kehne, Evi Micha, Ariel D. Procaccia, and Nisarg Shah

University of Toronto · Harvard John A. Paulson School of Engineering and Applied Sciences · Neural Information Processing Systems

What is Sortition?



- Democratic paradigm which randomly selects a panel (jury) from a population
- Origins in ancient Athens, used today e.g. in constructing *citizens' assemblies*
- Randomness guarantees that all individuals have some chance of being selected for P, while satisfying a collection of constraints which ensure that the chosen panel represents the underlying population (see Flanigan et al.)

Main Question:

How do fairness and representation trade off ...

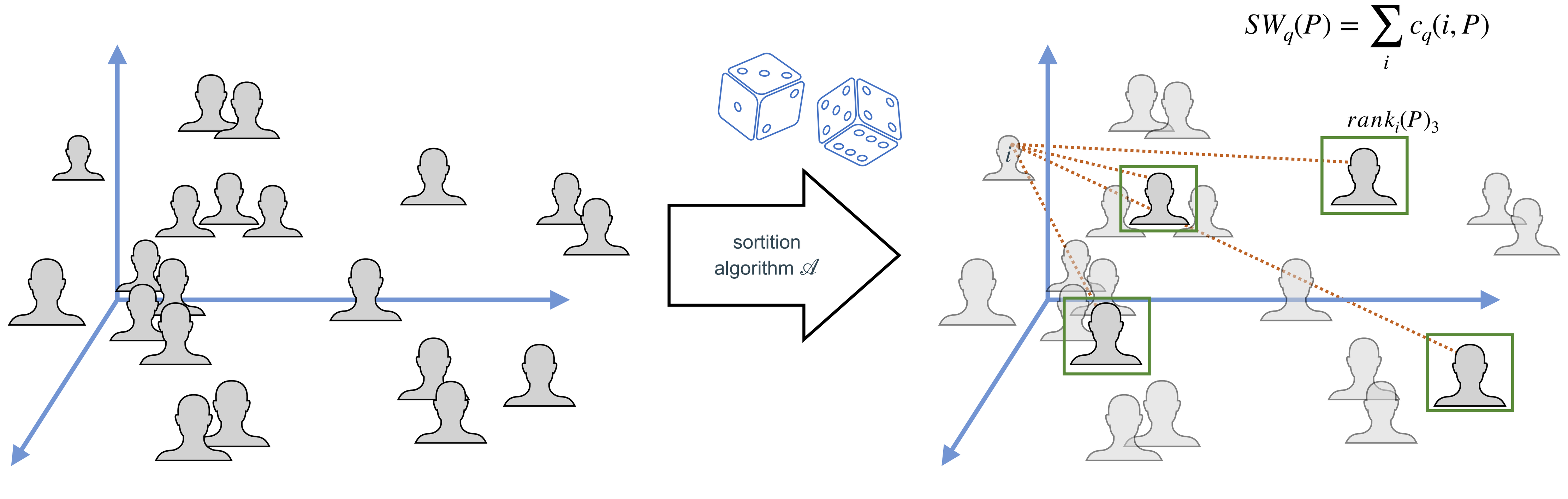
when representativeness within the population has some structure, and perfect representation is unattainable?

Sortition in a Metric Setting

- Measure the fairness of a sortition algorithm $\mathscr{A}$ by the minimum probability of any individual's selection
- Encode representation in a distance metric between individuals
- Representation distance from i to a panel P is the distance from i to their q^{th} closest representative in P, given by $d(i, rank_i(P)_q)$ (see Caragiannis et al.)
- Representation of a sortition algorithm $\mathscr{A}$ is the fraction of the best-possible representation it attains (in expectation), *in the worst case over all metrics*

$$c_q(i,P) = d(i, rank_i(P)_q)$$

$$SW_q(P) = \sum_i c_q(i,P)$$



Our Results

Theoretical upper and lower bounds on the best achievable tradeoff between representation and fairness.

less fair / more representative ↑ — ↓ more fair / less representative

> OPT and OPT *Approximation* have $repr_q(OPT) = \Omega(1)$ but $fair_q(OPT) = 0$.

> RandomReplace satisfies $fair_q(RR) \geq q/k$ and $repr_q(RR) = \Omega(1/q)$.

> For $q \leq k/2$, if $fair_q(\mathscr{A}) \geq q/k + \epsilon$ then $repr_q(\mathscr{A}) = 0$.

> For $q \leq k/2$, if $fair_q(\mathscr{A}) > q/k$ then $repr_q(\mathscr{A}) \leq k/q^2$.

> RandomSelection satisfies $repr_q(RS) = \Omega\left(\frac{k-q+1}{k}\right)$, which is tight if $q > k/2$ and $fair_q(\mathscr{A}) = 1$.

ALGORITHM 1: OPT

Input: metric d on $[n]$, $k \leq n$
Output: optimal panel $P \subseteq [n]$
1: $P^* \leftarrow \arg\min_{P \in \binom{[n]}{k}} \sum_{i \in [n]} c_q(i,P)$
2: **return** P^*

ALGORITHM 2: OPT Approximation

Input: metric d on $[n]$, $k \leq n$, $q \leq k$, algo KMEDIANSPROXY
Output: panel $P \subseteq [n]$ an approx to OPT
1: $k' \leftarrow \lfloor k/q \rfloor$, $P \leftarrow \emptyset$, $S \leftarrow [n]$
2: $Q \leftarrow$ KMEDIANSPROXY(d, k')
3: **for** $c \in Q$ **do**
4: $P_c \leftarrow$ the q closest $i \in S$ to c
5: $P \leftarrow P \cup P_c$, $S \leftarrow S \setminus P_c$
6: **end for**
7: augment P arbitrarily until $|P| = k$
8: **return** P

algorithm due to Kumar and Raichel

ALGORITHM 3: RANDOMREPLACE$_q$

Input: metric d on $[n]$, $k \leq n$, panel P with $\text{repr}_q(P) = \alpha$
Output: P with $\leq q$ random replacements
1: Pick $S \in \mathcal{S}_q$ uniformly at random
2: Set $P_S \leftarrow P$ and $\overline{S} \leftarrow S \setminus P$
3: **for** $i \in \overline{S}$ **do**
4: Pick an arbitrary $j_i \in \text{top}_q(i,P) \setminus S$
5: $P_S \leftarrow P_S \cup \{i\} \setminus \{j_i\}$
6: **end for**
7: **return** P_S

ALGORITHM 4: RANDOMSELECTION

Input: $[n]$, $k \leq n$
Output: u.a.r. panel $P \subseteq [n]$
1: Sample $P \sim \binom{[n]}{k}$ uniformly at random
2: **return** P

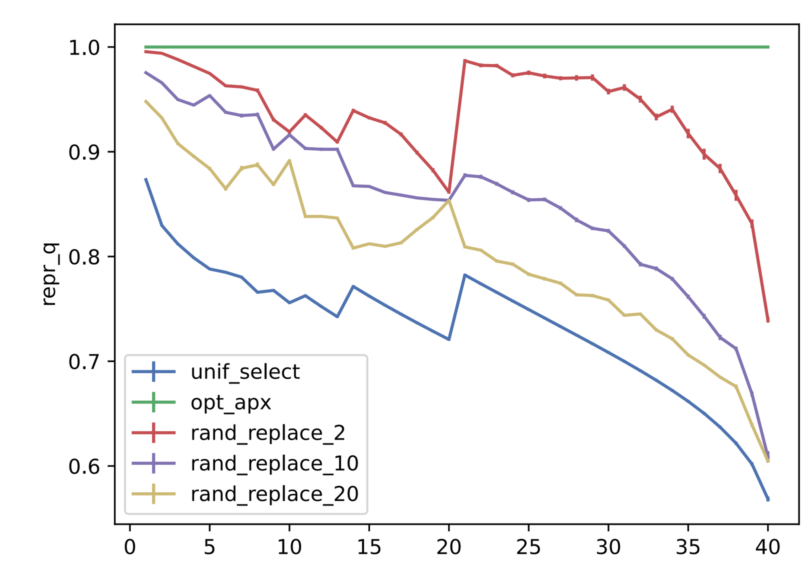
fairness: $fairness_q(\mathscr{A}) = \inf_d \dfrac{\min_{i \in [n]} \Pr[i \in \mathscr{A}(d)]}{k/n}$

representation: $repr_q(\mathscr{A}) = \inf_d \dfrac{\min_{P'} SC_q(P', d)}{\mathbb{E}[SC_q(\mathscr{A}(d))]}$

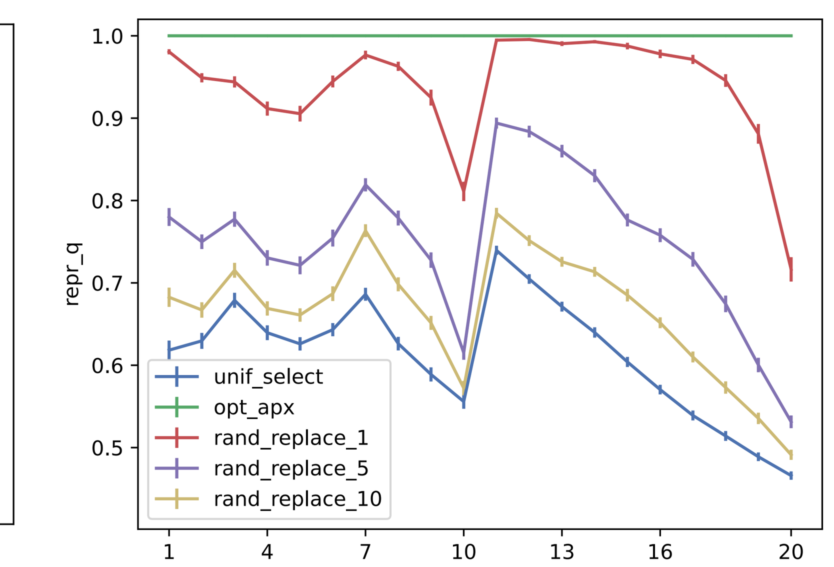
Experimental Findings

Testing the performance of OPT *Approximation* and $RandomReplace_q$ on synthetic representation metrics

- Constructed collections of randomized representation metrics using the UCI Adult dataset (based on 1994 US Census data) and the European Social Survey from 2018.
- Experiments show a discontinuity at $q = k/2$, reflected in the theory

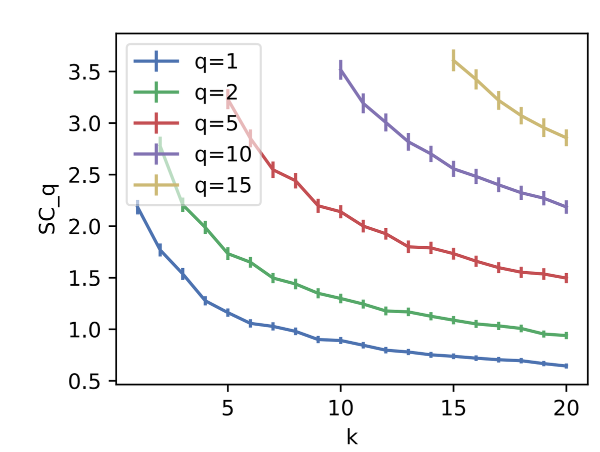


ESS dataset with $k = 40$

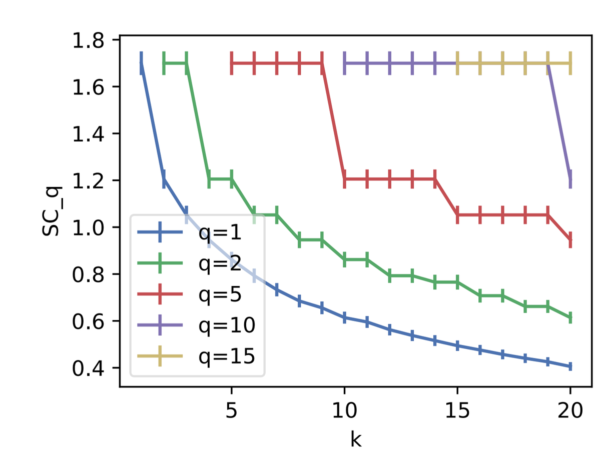


Adult dataset with $k = 20$

Representativeness of $RandomReplace$ for increasing values of q, as compared to OPT *Approximation*.



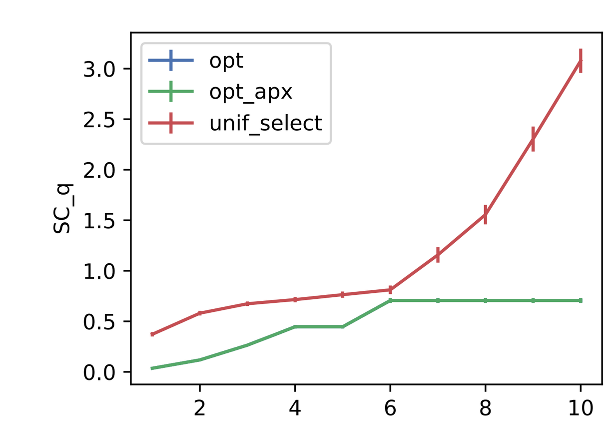
RandomSelection



OPT Approximation

An analysis of the Adult dataset for fixed values of q and increasing panel sizes k (above).

A benchmarking of OPT *Approximation* against OPT and $RandomSelection$ on a smaller instance (right).



Adult dataset with $k = 10$

References

Ioannis Caragiannis et al. (2022) "The metric distortion of multiwinner voting." Artificial Intelligence, 2022.

Bailey Flanigan et al. (2021) "Fair algorithms for selecting citizens' assemblies." Nature, 596:548–552, 2021.

Nirman Kumar and Benjamin Raichel (2013). "Fault tolerant clustering revisited." In: Proc. Proceedings of the 25th Canadian Conference on Computational Geometry (CCCG) 2013.