Social Choice for Social Good

Paul Gölz

In this thesis, we study ways in which computational social choice can contribute to social good, by providing better democratic representation and by allocating scarce goods.

On the side of political representation, we study multiple emerging innovations to the democratic process. If *legislative elections with proportional representation* allow for approval-based ballots rather than the choice of a single party, we give voting rules that satisfy attractive axioms of proportionality. For *liquid democracy*, a kind of transitive proxy voting, we show how an extension to multiple delegation options can decrease the concentration of power in few hands. Finally, for *sortition*, a system in which representatives are randomly selected citizens, we develop sampling algorithms, both for the case where all citizens participate if sampled and for the case in which participants self select.

Concerning the allocation of scarce goods, we investigate the applications of refugee resettlement and kidney exchange. For *refugee resettlement*, we show how submodular optimization can lead to more diverse matchings that might increase employment by reducing competition for jobs between similar refugees. In *kidney exchange*, we give approximately incentive-compatible mechanisms for transplant chains in a semi-random model of compatibilities.

Finally, we present three directions for future research, revisiting the topics of sortition, refugee resettlement, and semi-random models.
## Contents

1 Introduction .......................................................... 3  
   1.1 Improving Democratic Representation .......................... 3  
   1.2 Allocating Scarce Goods ......................................... 5  

2 Approval-Based Legislative Apportionment ......................... 6  
   2.1 Related Work ..................................................... 8  
   2.2 Voting Rules with Strong Proportionality Guarantees ......... 8  

3 Liquid Democracy ..................................................... 10  
   3.1 Related Work ..................................................... 11  
   3.2 Minimizing Supervoter Weight in Liquid Democracy ......... 12  

4 Sortition .............................................................. 14  
   4.1 Related Work ..................................................... 15  
   4.2 The Case of Full Participation .................................. 16  
   4.3 Limited Participation ............................................. 19  

5 Refugee Resettlement .................................................. 21  
   5.1 Related Work ..................................................... 22  
   5.2 Migration as Submodular Optimization ......................... 23  

6 Kidney Exchange ........................................................ 24  
   6.1 Related Work ..................................................... 26  
   6.2 Mechanism Design for Kidney Exchange in a Semi-Random Model ..................................................... 27  

7 Future Directions ..................................................... 30  
   7.1 Sortition with Fairness to the Pool ............................. 30  
   7.2 Online Refugee Matching ........................................ 31  
   7.3 Semi-Random Models in Fair Division ......................... 32
1 Introduction

In advising young students of science, Thomas Jefferson hoped to “keep their attention fixed on the main objects of all science, the freedom & happiness of man” [65]. How can computer science — which was certainly not on Jefferson’s mind — further the same goals?

There is reason to believe that computer science can have a positive impact on society. Quite obviously, what computers can do has increased tremendously over the recent past, and computation has already deeply impacted many areas of society. However, technology need not be the only way in which computer science can contribute to the social good. Indeed, in the field of computational social choice, computational techniques have proved to bring new perspectives and solutions to theoretical social questions that once were the sole domain of political scientists, economists, and some mathematicians [26].

The two parts of this thesis loosely correspond to Jefferson’s ideals of freedom and happiness. In the first part (“freedom”), we study three democratic innovations: liquid democracy, approval-based legislative voting, and sortition. These innovations have in common that they identify new ways of selecting democratic representatives, hopefully in a way that gives citizens more freedom to have their voices impact the decisions of society. By informing the development of these nascent paradigms of democratic participation, we hope to strengthen the effectiveness and legitimacy of liberal democracy.

The second part (“happiness”) studies two allocation problems that deeply impact people’s lives: the allocation of resettled refugees to cities and the allocation of kidney transplants in a paired-donation system. By placing refugees where they can thrive best, and by encouraging hospitals to pool their patients and allow for more kidney transplants, the welfare of affected people can increase.

General Background  The field of computational social choice is concerned with the algorithmic aspects of group decision making. In the first part of the thesis, the decision being studied is not an individual policy decision, but a decision on a set of representatives. This line of work has its roots in the area of multi-winner elections, in which a committee of candidates has to be selected [75]. However, our problems deviate from the classical setup in terms of the ballot format considered and in terms of the committees produced (whether a single entity can be chosen multiple times). The problems in the second part (in a very general sense) produce matchings in a graph, and borrow tools from matching and mechanism design.

1.1 Improving Democratic Representation

The first part of this thesis studies three extensions to the current practice of democracy.

On a high level, the history of democracy is a story of great success. Over the span of just 70 years, the share of the world population living in democratic systems increased from 11% in 1945 to 55% in 2015.\(^1\) While democracy proper only refers to the form of governance, political freedom and other civil liberties are often closely linked, with Mulligan, Gil, and Sala-i-Martin [87] finding that democracies are less “likely to torture, execute, regulate religion, censor the press and spend a lot of money on the military,” as compared to authoritarian countries.

\(^1\)https://ourworldindata.org/grapher/world-pop-by-political-regime?stackMode=relative
Despite encouraging long-term trends, recent signs show that democracy’s success cannot be taken as granted in the future. Not only have multiple political systems become more authoritarian in the past decade, but support for democracy has dwindled even in supposedly consolidated democracies such as France and the United States [52]. It remains to hope that this dissatisfaction is not directed against the values of democracy itself but against the current implementation of this idea. For instance, in the median over 34 countries across the globe,\textsuperscript{2} 64\% disagree with the statement that “[m]ost elected officials care what people like me think,” and this view is highly correlated with low satisfaction with democracy.

If, as Churchill said, “democracy is the worst form of Government except for all those other forms that have been tried,” it is worth investigating alternative ways of democratic decision making, in which citizens have a more ways of making their voices heard. We study three such systems, ordered from least disruptive to most disruptive:

First, we study approval-based legislative apportionment, which extend legislative elections in systems of proportional representation. In the status quo, voters can vote for a single party and a fixed number of seats in the legislature is apportioned to the parties in a way that aims to represent parties proportionally to their number of votes. The restriction to indicating a single party can lead to coordination problem, where different voter groups could profit from voting for a consensus candidate, but fail to do so without coordination. Approval-based apportionment lifts the restriction of indicating a single party and instead allows voters to support any subset of the parties. We propose apportionment rules for this setting that guarantee attractive axioms of representation [30].

Liquid democracy goes even further by no longer requiring that political influence be mediated by parties and professional politicians. Instead, direct democracy is extended by allowing citizens to (transitively) delegate their voting power to any other citizen of their choice. In the largest liquid-democracy experiment so far, observers were concerned that too much power ended up being concentrated in the hands of few voters. We suggest an extension to liquid democracy in which agents can indicate multiple possible delegations that they are indifferent between, and in which a centralized algorithm chooses one of these delegations per agent such that the maximum power of any voter is minimized. In a stochastic model of delegation network formation, we observe a “power of two choices” effect, which suggests that this extension could drastically reduce the concentration of power in liquid democracy [55].

Finally, the system of sortition uses randomization to overcome the requirement that every member of the population must make a decision. Once a representative panel is selected from the population, it is possible to spend resources on informing the panel on the topic (which allows the panel to make an informed choice) and to have the panel deliberate (which allows the panel to take into account perspectives from all across the population). The democratic legitimacy of such a panel rests on the properties of sampling process, which ensure equal access to individuals and groups. If everybody in the population participates when sampled,\textsuperscript{3} we design a variant of stratified sampling that selects every member of the population with equal probability, that reduces the variance in representation for groups of interest, and that cannot substantially increase the variance in representation for any other subset of the population [17]. If—as is current practice—participation is voluntary with different groups in the population participating at different rates, we design an algorithm that attempts to recover guarantees on

\textsuperscript{2}https://www.pewresearch.org/global/2020/02/27/democratic-rights-popular-globally-but-commitment-to-them-not-always-strong/

\textsuperscript{3}Some authors advocate for mandatory participation [77, 22], somewhat like jury duty or a military draft.
1.2 Allocating Scarce Goods

While the first part of the thesis deals with ways of improving general decision making, the second part of the thesis focuses on two individual challenges: refugee resettlement and kidney exchange. Common to both is the fact that they allocate limited resources (government-approved slots in a host locality or transplants) which to receive can greatly improve the recipient’s life. In other ways, the two studied problems differ substantially: In our refugee resettlement, we treat information as reliable, which makes it a problem of optimization. By contrast, in kidney exchange, hospitals are modeled as strategic actors that must be incentivized to cooperate via mechanism design without money. Another difference between the two problems is the nature of the scarce resource: Whereas the yearly slots in refugee resettlement are fixed, every kidney recipient makes a corresponding donor available for donation, which means that a single initial organ can lead to a long chain of transplantations.

As of 2018, 26 million refugees are forced to find shelter from violence and persecution in countries other than their own.\(^4\) A lot of attention has been raised towards the limited willingness of nations in the global North to host refugees. But, when a refugee reaches her host country, this is only the beginning of a difficult path of integration. A major factor in successful integration is early access to employment [79], and different refugees have different prospects of employment depending on where in the country they arrive. In the case of refugee resettlement, where initial placement can be externally decided, there have been recent pushes [12, 100] to match according to predicted employment prospects. Given enough political will, the same systems could be used for placing refugees in an international setting, say in the European Union. We argue that the existing approaches based on maximum-weight matching can lead to placing very homogenous sets of refugees to each locality. If similar refugees compete for similar employment, the model might overstate the employment prospects in such an allocation, which might reduce the efficiency of the solution. We suggest modeling competition effects via submodular functions, and optimizing them to obtain more diverse matchings with less internal competition [54].

The second problem that we study is kidney exchange, the problem of identifying feasible exchanges of donor kidneys across multiple pairs consisting of a patient and a donor each. There is a rich literature on the case where these exchanges take on the shape of 2-cycles and 3-cycles. However, the case where the matching is an arbitrary-length chain going out from an altruistic donor has been less conducive to positive theoretical results. A major challenge in kidney exchange is the behavior of hospitals, who have been found not to register patient-donor pairs in the exchange in the hope that this could lead to more of their patients being matched. As a result, a matching algorithm that simply returns the longest path from the altruist, might encourage hospitals to hide many nodes, and might therefore drive down the number of patients receiving a kidney. Instead, our objective is to find mechanisms that limit incentives for such manipulation while producing long paths. Due to impossibility results on worst-case instances, we adopt a semi-random model where a small number of random edges are added to worst-case instances. While it remains impossible for truthful mechanisms to compete with the overall longest path, we give a truthful mechanism that competes with an only slightly

\(^4\)https://www.unhcr.org/globaltrends2018/
weaker benchmark: the length of any path whose subpaths within each player have a minimum average length [19].

2 Approval-Based Legislative Apportionment

This work is set in the context of legislative elections in systems of proportional representation. Unlike in “first-past-the-post” systems, in which each seat in the legislature is awarded to the plurality winner of a certain district, proportional representation prescribes that the total number of representatives championing a particular opinion in a legislature be proportional to the number of voters who favor that opinion.

In most democratic institutions, proportional representation is implemented via party-list elections: Candidates are members of political parties and voters are asked to indicate their favorite party; each party is then allocated a number of seats that is (approximately) proportional to the number of votes it received. The problem of transforming a voting outcome into an allocation of seats is known as apportionment. Analyzing the advantages and disadvantages of different apportionment methods has a long and illustrious political history and has given rise to a deep and elegant mathematical theory [10, 89].

Unfortunately, forcing voters to choose a single party prevents them from communicating any preferences beyond their most preferred alternative. For example, if a voter feels equally well represented by several political parties, there is no way to express this preference within the voting system.

In the context of single-winner elections, approval voting has been put forward as a solution to this problem as it strikes an attractive compromise between simplicity and expressivity [23, 76]. Under approval voting, each voter is asked to specify a set of candidates she “approves of,” i.e., voters can arbitrarily partition the set of candidates into approved candidates and disapproved ones. Proponents of approval voting argue that its introduction could increase voter turnout, “help elect the strongest candidate,” and “add legitimacy to the outcome” of an election [23, p. 4–8].

Due to the practical and theoretical appeal of approval voting in single-winner elections, a number of scholars have suggested to also use approval voting for multi-winner elections, in which a fixed number of candidates needs to be elected [71]. In contrast to the single-winner setting, where the straightforward voting rule “choose the candidate approved by the highest number of voters” enjoys a strong axiomatic foundation [48], several ways of aggregating approval ballots have been proposed in the multi-winner setting [6, 63].

Most studies of approval-based multi-winner elections assume that voters directly express their preference over individual candidates; we refer to this setting as candidate-approval elections. This assumption runs counter to widespread democratic practice, in which candidates belong to political parties and voters indicate preferences over these parties (which induce implicit preferences over candidates). In this paper, we therefore study party-approval elections, in which voters express approval votes over parties and a given number of seats must be distributed among the parties. We refer to the process of allocating these seats as approval-based apportionment.

We believe that party-approval elections are a promising framework for legislative elections in the real world. Allowing voters to express approval votes over parties enables the aggregation mechanism to coordinate like-minded voters. For example, two blocks of voters might currently
vote for parties that they mutually disapprove of. Using approval ballots could reveal that the blocks jointly approve a party of more general appeal; allocating more seats to this party leads to mutual gain. This cooperation is particularly necessary for small minority opinions that are not centrally coordinated. In such cases, finding a commonly approved party can make the difference between being represented or votes being wasted because the individual parties receive insufficient support.

In contrast to approval voting over individual candidates, party-approval voting does not require a break with the current role of political parties—it can be combined with both “open list” and “closed list” approaches to filling the seats allocated to a party.

As illustrated in Figure 1, party-approval elections can be positioned between two well-studied voting settings:

First, approval-based apportionment generalizes standard apportionment, which corresponds to party-approval elections in which all approval sets are singletons. This relation (depicted as arrow (i) in Figure 1) provides a generic two-step approach to define aggregation rules for approval-based apportionment problems: transform a party-approval instance to an apportionment instance, and then apply an apportionment method.

Second, our setting can be viewed as a special case of approval-based multi-winner voting, in which voters cast candidate-approval votes. A party-approval election can be embedded in this setting by replacing each party by multiple candidates belonging to this party, and by interpreting a voter’s approval of a party as approval of all of its candidates. This embedding establishes party-approval elections as a subdomain of candidate-approval elections (see arrow (ii) in Figure 1), and allows us to transform arbitrary candidate-approval voting rules to our setting.

By using either of the two approaches, we obtain algorithms for approval-based apportionment that satisfy attractive guarantees for which no algorithms are known in the candidate-approval setting.
2.1 Related Work

To the best of our knowledge, we are the first to formally develop and systematically study approval-based apportionment. That is not to say that the idea of expressing and aggregating approval votes over parties has not been considered before. Indeed, several scholars have explored possible generalizations of existing aggregation procedures.

For instance, Brams et al. [25] study multi-winner approval rules that are inspired by classical apportionment methods. Besides the setting of candidate approval, they explicitly consider the case where voters cast party-approval votes. They conclude that these rules could “encourage coalitions across party or factional lines, thereby diminishing gridlock and promoting consensus.”

Such desire for compromise is only one motivation for considering party-approval elections, as exemplified by recent work by Speroni di Fenizio and Gewurz [95]. To allow for more efficient governing, they aim to concentrate the power of a legislature in the hands of few big parties, while nonetheless preserving the principle of proportional representation. To this end, they let voters cast party-approval votes and transform these votes into a party-list election by assigning each voter to one of her approved parties. One method for doing this (referred to as majoritarian portioning later in this paper) assigns voters to parties in such a way that the strongest party has as many votes as possible.

Several other papers consider extensions of approval-based voting rules to accommodate party-approval elections [24, 83, 63, 64]. All of these papers have in common that they study specific rules or classes of rules, rather than exploring the party-approval setting in its own right.

2.2 Voting Rules with Strong Proportionality Guarantees

Model A party-approval election is a tuple \((N, P, A, k)\) consisting of a set of voters \(N = \{1, \ldots, n\}\), a finite set of parties \(P\), a ballot profile \(A = (A_1, \ldots, A_n)\) where each ballot \(A_i \subseteq P\) is the set of parties approved by voter \(i\), and the committee size \(k \in \mathbb{N}\). We assume that \(A_i \neq \emptyset\) for all \(i \in N\). For questions of computational complexity, we assume \(k\) to be encoded in unary.

A committee is a multiset \(W : P \rightarrow \mathbb{N}\) over parties, which determines the number of seats \(W(p)\) assigned to each party \(p \in P\). The size of a committee \(W\) is given by \(|W| = \sum_{p \in P} W(p)\), and we denote multiset addition and subtraction by \(+\) and \(-\), respectively. A party-approval rule is a function that takes a party-approval election \((N, P, A, k)\) as input and returns a committee \(W\) of valid size \(|W| = k\).

Axioms We focus on two axioms capturing proportional representation: extended justified representation and core stability [6]. Both axioms are derived from their analogs in multi-winner elections and can be defined in terms of quota requirements: For a party-approval election \((N, P, A, k)\) and a subset \(S \subseteq N\) of voters, define the quota of \(S\) as \(q(S) = \lfloor k \cdot |S|/n \rfloor\). Intuitively, \(q(S)\) corresponds to the number of seats that the group \(S\) “deserves” to be represented by (rounded down).

\(5\)This definition implies that rules are resolute, that is, only a single committee is returned. In the case of a tie between multiple committees, a tiebreaking mechanism is necessary. Our results hold independently of the choice of a specific tiebreaking mechanism.
Definition 1. A committee $W : P \rightarrow \mathbb{N}$ provides \textit{extended justified representation (EJR)} for a party-approval election $(N, P, A, k)$ if there is no subset $S \subseteq N$ of voters such that $\bigcap_{i \in S} A_i \neq \emptyset$ and $\sum_{p \in A_i} W(p) < q(S)$ for all $i \in S$.

In words, EJR requires that for every voter group $S$ with a commonly approved party, at least one voter of the group should be represented by $q(S)$ many candidates. A party-approval rule is said to \textit{satisfy EJR} if it only produces committees providing EJR.

We can obtain a stronger representation axiom by removing the requirement of a commonly approved party.

Definition 2. A committee $W : P \rightarrow \mathbb{N}$ is \textit{core stable} for a party-approval election $(N, P, A, k)$ if there is no nonempty subset $S \subseteq N$ and committee $T : P \rightarrow \mathbb{N}$ of size $|T| \leq q(S)$ such that $\sum_{p \in A_i} T(p) > \sum_{p \in A_i} W(p)$ for all $i \in S$. The core of a party-approval election is defined as the set of all core-stable committees.

Core stability requires adequate representation even for voter groups that cannot agree on a common party, by ruling out the possibility that the group can deviate to a smaller committee that represents all voters in the group strictly better. It follows from the definitions that core stability is a stronger requirement than EJR: If a committee violates EJR, there is a group $S$ that would prefer any committee of size $q(S)$ that assigns all seats to the commonly approved party.

A final, non-representational axiom that we will discuss is \textit{committee monotonicity}. A party-approval rule $f$ satisfies this axiom if, for all party-approval elections $(N, P, A, k)$, it holds that $f(N, P, A, k) \subseteq f(N, P, A, k + 1)$. Committee monotonic rules avoid the so-called \textit{Alabama paradox}, in which a party loses a seat when the committee size increases. Besides, committee monotonic rules can be used to construct proportional rankings [93].

Result 1: EJR & Committee Monotonicity   Exploiting the relations described in Fig. 1, we resolve problems that remain open in the more general setting of approval-based multi-winner voting. First, we prove that committee monotonicity is compatible with extended justified representation by providing a rule that satisfies both properties. As mentioned earlier, this rule proceeds in two steps: It first reduces the approval-based apportionment instance into a regular apportionment instance, by choosing a single party out of every agent’s approval set in a process which we call “majoritarian portioning.” In a second stage, we apportion the reduced instance using Jefferson’s method, a classical apportionment method.

\textit{Majoritarian portioning} was proposed by Speroni di Fenizio and Gewurz [95], and proceeds in rounds $j = 1, 2, \ldots$. Initially, all parties and voters are active. In iteration $j$, select the active party $p_j$ that is approved by the highest number of active voters. For each active voter approving of $p_j$, drop all other parties from her approval set, and mark $p_j$ and all voters in $N_j$ as inactive. If active voters remain, the next iteration is started. Else, every agent votes now for a single party and apportionment rules can be applied.

\textit{Jefferson’s method} (also known as the D’Hondt method) assigns the $k$ seats iteratively, each time giving the next seat to the party $p$ with the largest quotient $r(p)/(s(p) + 1)$, where $s(p)$ denotes the number of seats already assigned to $p$.

Theorem 3. The party-approval rule composing majoritarian portioning and Jefferson’s method satisfies EJR and committee monotonicity.
In fact, Jefferson’s method can be replaced by any other apportionment method satisfying a property known as *lower quota*. What makes this result surprising is that a lot of candidates that on the outset would seem more promising than majoritarian portioning fail to guarantee EJR when combined with promising apportionment methods. Majoritarian portioning was originally proposed without axiomatic justification, which our result gives for the first time.

**Result 2: Core Stability** Second, we show that the core of an approval-based apportionment problem is always nonempty, a “tantalizing open question” [35] in the general setting of candidate-approval elections. Indeed, this is obtained by the embedding of the well-known proportional approval voting rule from candidate-approval elections [98].

In our setting, given a party-approval election \((N, P, A, k)\), the *proportional approval voting (PAV)* rule chooses the committee \(W\) maximizing the PAV score

\[
PAV(W) = \sum_{i \in N} H \left( \sum_{p \in A_i} W(p) \right),
\]

where \(H(i)\) denotes the \(i\)th harmonic number, \(\sum_{j=1}^i 1/j\).

**Theorem 4.** For every party-approval election, PAV chooses a core-stable committee.

This result is surprising given that PAV is know not to satisfy core stability in the candidate-approval setting and given that the question of whether the core of candidate-approval election is always non-empty has proved to be a hard open problem. Since PAV is NP-hard to evaluate, we also give a local-search variant of PAV that produces core-stable committees in polynomial time.

### 3 Liquid Democracy

*Liquid democracy* is a potentially disruptive approach to democratic decision making. As in direct democracy, agents can vote on every issue by themselves. Alternatively, however, agents may delegate their vote, i.e., entrust it to any other agent who then votes on their behalf. Delegations are transitive; for example, if agents 2 and 3 delegate their votes to 1, and agent 4 delegates her vote to 3, then agent 1 would vote with the weight of all four agents, including herself. Just like representative democracy, this system allows for separation of labor, but provides for stronger accountability: Each delegator is connected to her transitive delegate by a path of personal trust relationships, and each delegator on this path can withdraw her delegation at any time if she disagrees with her delegate’s choices.

Although the roots of liquid democracy can be traced back to the work of Miller [82], it is only in recent years that it has gained recognition among practitioners. Most prominently, the German Pirate Party adopted the platform *LiquidFeedback* for internal decision-making in 2010. At the highest point, their installation counted more than 10,000 active users [72]. More recently, two parties — the Net Party in Argentina, and Flux in Australia — have run in national elections on the promise that their elected representatives would vote according to decisions made via their respective liquid-democracy-based systems. Although neither party was able to win any seats in parliament, their bids enhanced the promise and appeal of liquid democracy.
However, these real-world implementations also exposed a weakness in the liquid-democracy approach: Certain individuals, the so-called super-voters, seem to amass enormous weight, whereas most agents do not receive any delegations. In the case of the Pirate Party, this phenomenon is illustrated by an article in Der Spiegel, according to which one particular super-voter’s “vote was like a decree,” even though he held no office in the party. As Kling et al. [72] describe, super-voters were so controversial that “the democratic nature of the system was questioned, and many users became inactive.” Besides the negative impact of super-voters on perceived legitimacy, super-voters might also be more exposed to bribing. Although delegators can retract their delegations as soon as they become aware of suspicious voting behavior, serious damage might be done in the meantime. Furthermore, if super-voters jointly have sufficient power, they might find it more efficient to organize majorities through deals between super-voters behind closed doors, rather than to try to win a broad majority through public discourse. Finally, recent work by Kahng et al. [67] indicates that, even if delegations go only to more competent agents, a high concentration of power might still be harmful for social welfare, by neutralizing benefits corresponding to the Condorcet Jury Theorem.

While all these concerns suggest that the weight of super-voters should be limited, the exact metric to optimize for varies between them and is often not even clearly defined. In this work, we choose to minimize the weight of the heaviest voter. As is evident in the Spiegel article, the weight of individual voters plays a direct role in the perception of super-voters. But even beyond that, we are confident that minimizing this measure will lead to substantial improvements across all presented concerns.

Just how can the maximum weight be reduced? One approach might be to restrict the power of delegation by imposing caps on the weight. However, as argued by Behrens et al. [16], delegation is always possible by coordinating outside of the system and copying the desired delegate’s ballot. Pushing delegations outside of the system would not alleviate the problem of super-voters, just reduce transparency. Therefore, we instead adopt a voluntary approach: If agents are considering multiple potential delegates, all of whom they trust, they are encouraged to leave the decision for one of them to a centralized mechanism. With the goal of avoiding high-weight agents in mind, our research challenge is twofold:

First, investigate the algorithmic problem of selecting delegations to minimize the maximum weight of any agent, and, second, show that allowing multiple delegation options does indeed provide a significant reduction in the maximum weight compared to the status quo.

3.1 Related Work

Kling et al. [72] conduct an empirical investigation of the existence and influence of super-voters. The analysis is based on daily data dumps, from 2010 until 2013, of the German Pirate Party installation of LiquidFeedback. As noted above, Kling et al. find that super-voters exist, and have considerable power. The results do suggest that super-voters behave responsibly, as they “do not fully act on their power to change the outcome of votes, and they vote in favour of proposals with the majority of voters in many cases.” Of course, this does not contradict the idea that a balanced distribution of power would be desirable.

There are only a few papers that provide theoretical analyses of liquid democracy [56, 37, 67]. We would like to stress the differences between our approach and the one adopted by Kahng
et al. [67]. They consider binary issues in a setting with an objective ground truth, i.e., there is one “correct” outcome and one “incorrect” outcome. In this setting, voters are modeled as biased coins that each choose the correct outcome with an individually assigned probability, or competence level. The authors examine whether liquid democracy can increase the probability of making the right decision over direct democracy by having less competent agents delegate to more competent ones. By contrast, our work is completely independent of the (strong) assumptions underlying the results of Kahng et al. In particular, our approach is agnostic to the final outcome of the voting process, does not assume access to information that would be inaccessible in practice, and is compatible with any number of alternatives and choice of voting rule used to aggregate votes. In other words, the goal is not to use liquid democracy to promote a particular outcome, but rather to adapt the process of liquid democracy such that more voices will be heard.

3.2 Minimizing Supervoter Weight in Liquid Democracy

**Problem Statement** Let $G = (N, E)$ be finite and directed graph, whose nodes represent the agents, and where the outgoing edges from a node represent the agent’s indicated delegation options. In addition, a set $V \subseteq \text{sinks}(G)$ represents the voters, i.e., the agents that vote for themselves. We call a graph $(N, E)$ with voters $V$ a delegation graph if it is acyclic, its sinks are exactly the set $V$, and every other vertex has outdegree one. In such a graph, define the weight $w(n)$ of a node $n \in N$ as

$$w(n) := 1 + \sum_{(m,n) \in E} w(m).$$

This is well-defined because $E$ is a well-founded relation on $N$.

Resolving the delegations of a graph $G$ with voters $V$ can now be described as the Min-MaxWeight problem: Among all delegation subgraphs $(N', E')$ of $G$ with voting vertices $V$ of maximum $|N'|$, find one that minimizes the maximum weight of the voting vertices.

**Computational Complexity** MinMaxWeight is closely connected to the problem MinMaxCongestion of minimizing the maximum congestion for confluent flow. Indeed, we can see voting weight as a flow with each agent having demand 1, and only voters being able to be flow sinks. The fact that every agent delegates to at most one agent corresponds to the confluence requirement on the flow, i.e., that every node has at most one outgoing edge with positive flow. The problems mainly differ in what to do with agents from whom no voter is reachable. Resolving this complication allows us to reduce in both directions between the problems and to transfer both results on the hardness of approximation of MinMaxCongestion and approximation algorithms for MinMaxCongestion [33] to the MinMaxWeight problem:

**Theorem 5.** There is a polynomial-time $(1 + \ln |V|)$-approximation algorithm for MinMaxWeight.

**Theorem 6.** It is NP-hard to approximate the MinMaxWeight problem to a factor of $\frac{1}{2} \log_2 |V|$ (even when each node has outdegree at most 2).
Stochastic Model of Delegation To evaluate the benefits of allowing multiple delegations, we propose a probabilistic model for delegation behavior — inspired by the well-known preferential attachment model [14] — in which we add agents successively. The model is parametrized by three parameters: \( d \in (0, 1) \), \( k \in \mathbb{N} \), and \( \gamma \geq 0 \).

The first agent will always be a voter. Any subsequent voter delegates with the delegation probability \( d \); otherwise, she votes herself. If she delegates, she chooses \( k \) many delegation options among the previously inserted agents. These choices are independent and choose each preexisting agent \( j \) with probability proportional to \((\text{indegree} + 1)^\gamma\).\(^6\) As shown in Fig. 2, the parameter \( \gamma \) controls the bias of this selection towards agents who already receive many delegations. This probabilistic model allows us to theoretically and empirically analyze whether having multiple delegation options per delegator allows to reduce the maximum weight in the graph.

Lower Bound for Single Uniform Delegation In the case of \( \gamma = 0 \), which we term uniform delegation, a delegator is equally likely to attach to any previously inserted node. Already in this case, a “rich-get-richer” phenomenon can be observed, i.e., voters at the end of large networks of potential delegations will likely see their network grow even more. Indeed, a larger network of delegations is more likely to attract new delegators. If every delegator names only a single delegation option, the maximum weight scales at least as fast as a power-law function with high probability:

**Theorem 7.** In the preferential delegation model with \( k = 1 \), \( \gamma = 0 \), and \( d \in (0, 1) \), with high probability, the maximum weight of any voter at time \( t \) is in \( \Omega(t^{\beta}) \), where \( \beta > 0 \) is a constant that depends only on \( d \).

In traditional liquid democracy, where \( k = 1 \) and all potential delegations will be realized, this explains the emergence of super-voters with excessive weight observed by Kling et al. [72].

Upper Bound for \( k = 2 \) and Uniform Delegation We aim to show that for \( k \geq 2 \), the resolution of potential delegations can strongly outweigh these effects. Our analysis draws on a phenomenon called the power of choice that can be observed in many different load balancing models. In fact, even a greedy mechanism that selects a delegation option to locally minimize the maximum weight as agents arrive exhibits this asymptotic behavior, which upper-bounds the maximum weight for optimal resolution. We find that, at least for \( \gamma = 0 \), the “power-of-choice” effect outweighs the “rich-get-richer” dynamic described earlier:

\(^6\)A single delegation option might be sampled multiple times, which reduces the effective number of delegation options below \( k \).
Theorem 8. In the preferential delegation model with \( k = 2, \gamma = 0, \) and \( d \in (0, 1) \), the maximum weight of any voter at time \( t \) is \( \log_2 \log t + \Theta(1) \) with high probability.

This establishes a doubly-exponential separation between the cases of single delegation and double delegation; more delegation options can only decrease the maximum weight further.

**Empirical Evaluations** In empirical simulations, we confirm that the gap between single and double delegation continues to exist for values of \( \gamma \) up to 1, which is usually seen as a proxy for real-world social networks. We also find that the maximum weight can already be brought down if only a fraction of the delegating agents give two delegations, with all other delegators giving a single option.

### 4 Sortition

The preceding two sections built up on the well-known paradigms of representative democracy and direct democracy. Whether an agent’s vote was a choice of party, of a set of potential delegates, or of a direct vote on an issue, every agent in the population always had to make a decision.

On the one hand, that seems very natural—after all, the principle of “one man, one vote” guarantees all individuals in the population their fair say. However, this concept does bring problems with it: A frequent criticism of direct democracy is that voters have too little information to make an informed choice on complicated policy issues. Both (electoral) representative democracy and liquid democracy build on the assumption that, nonetheless, voters are able to identify representatives that serve their values and interests.

**Sortition** chooses a different path by sampling a representative panel of citizens, who then decide [86, 31, 50, 22, 101, 97, 42]. The fact that this panel is small—typical panel sizes are around around 100 citizens — allows to spend resources on informing the panel members of the facts pertaining to the policy issue, and allows the members to deliberate. In contrast to opinion polls and plebiscites, sortition reveals an *informed* opinion, which can act as a proxy for the will of a (utopian) perpetually-informed public. Nonetheless, this proposal is radical in that it is predicated on the ability of average citizens to govern themselves and their peers.

It may, therefore, come as a surprise that there is extensive historical precedent for governance by sortition — and that present-day experiments give grounds for optimism. In 4th-century BC Athens, the Council of 500 and the People’s Court, described by contemporary sources as equal to the better-known Assembly, were comprised of randomly selected citizens [60]. Sortition also played a role in the Italian city-states of the Renaissance, most notably in Venice, where the doge was selected through a complicated sequence of rounds alternating election and sortition [85]. After sortition was abandoned in favor of elections in the aftermath of the American and French revolutions, it was largely forgotten and persisted chiefly as a method for selecting jurors.

Over the last two decades, sortition saw a revival in the formation of citizens’ panels around the world, with encouraging outcomes. In his excellent survey on sortition, Van Reybrouck [101] gives a detailed account of several of these experiments, in which such randomly selected panels deliberated on topics as varied as energy policy, electoral reform, and gay marriage.

The legitimacy of the sortition process rests on the stochastic properties of the sampling process.
With respect to the population, every agent should have equal probability of ending up in the panel, guaranteeing her that the system values her input as much as that of everyone else. When every individual is selected with equal probability, it also follows that any group making up \( x\% \) of the population will, in expectation, make up \( x\% \) of the panel.

With respect to the panel, the sampled panels should not be unreasonably unrepresentative, such that the panel weighs the perspectives of all groups in the population.\(^7\) The fair representation of groups can be deterministically enforced via quotas and stratification, or it can be made likely by concentration effects.

It is instructive to think of the sampling step as selecting a panel of fixed size \( k \) members among the population of size \( n \) by uniform sampling without replacement. In this case, fair selection probabilities in the population are clearly given, and the panel is likely to represent a group close to its proportional share by concentration of measure.

4.1 Related Work

While randomization is a common tool in computational social choice [27], few papers explicitly consider sortition. One exception is an article by Walsh and Xia [102], in which they prefix standard voting rules with random subsampling of the voters. They find that sortition can add hardness of manipulation to otherwise manipulable voting rules. Thus, they consider sortition as an additional step in choosing a societal outcome, whereas we are interested in the composition of the panel itself.

Saran and Tumennasan [91] study which social choice rules can be implemented in Nash equilibrium under sortition. Agents can vote strategically, but must commit to a vote before knowing who is chosen into the panel. Then, under mild assumptions, every rule that can be implemented by consulting the entire population can also be implemented under (uniform) sortition, as long as panels have at least size four.

We share our goal of fairly representing subgroups with multiple areas in social choice, namely apportionment and multi-winner elections (both of which we discussed in Section 2). While all standard apportionment mechanisms are deterministic, Grimmett [57] proposes a randomized apportionment mechanism. Whereas, in our setting, agents are part of many groups, apportionment generally considers every agent as belonging to a single state or party. In our work on approval-based legislative apportionment, this restriction is lifted in the sense that agents can belong to multiple known groups, and each group receives a nonnegative number of positions in the committee. All representation axioms that can be ensured in this setting provide an agent with guarantees on the number of positions summed up over all of the agent’s groups. By contrast, a sortition panel is composed of agents who themselves belong to multiple groups, and our representation guarantees hold for every group simultaneously.

Party-approval elections, in turn, are a special case of multi-winner elections [7]. Since they allow agents to directly approve arbitrary subsets of candidates, these preferences cannot directly be interpreted as group memberships. By constraining the set of possible preferences, one recovers apportionment [29] or party-approval elections.

On a technical level, our work is connected to questions of (stratified) sampling in statistics [40, 62] and of sample reweighting. Whereas we want a proportional representation of a

\(^7\)This distinguishes sortition from a randomized dictatorship.
certain feature in our sample, pollsters use the sample to obtain an accurate estimate of the prevalence of the feature. In this framework, our requirement of fair expected representation translates into requiring an unbiased estimator. In both cases, mechanisms leading to lower variance are preferred, and it is known that sampling from a continuous pool cannot increase variance. Despite these similarities, our setting is more restrictive, since we cannot weight agents differently, which is an important technique in polling. While, in final votes, weighting representatives might be defensible, a representative’s influence on a debate cannot be weighted.

4.2 The Case of Full Participation

Let us return to the ideal of sampling $k$ panel members uniformly from the population (without replacement). If sortition is administered with such a sampling technique, a unique advantage compared to other forms of democracy is that sortition affords fair representation to all possible groups in the population, in expectation. As long as every agent has the same probability of being selected, any subpopulation $M$ is expected to send $\frac{|M|}{n} k$ representatives to a council of size $k$, where $n$ is the total number of agents. This is a drastic improvement over the status quo, which widely fails to provide similar representation even to high-profile groups such as women or racial minorities.

Still, these guarantees only hold ex ante, i.e., in expectation over the random sampling. Panels are typically selected to be large enough that any given group is unlikely to be grossly over- or underrepresented. Still, practitioners seem unhappy with the plausible deviations in demographic representation. To mitigate these deviations, they often stratify their samples—for example, by choosing half of the representatives among women and half among men. Assuming that this reflects the composition of the population, stratification does not change an individual’s probability of being selected, nor the expected representation of groups. However, it guarantees ex-post fair representation to the strata, i.e., to women and men.

But what if the exact group is unknown to the stratifier? We would argue that the largest need for equal representation concerns political convictions. If the panels are constituted repeatedly, fluctuations in demographic representation might impact optics; the guarantee on expected representation ensures that all demographic groups wield their fair share of power over time. This argument does not apply to political opinions: Certainly, fair expected representation is a big improvement over the status quo, and we know of no other system that could give similar guarantees. But a negative fluctuation in representation can prevent a suggestion from gaining traction or could afford an undeserved majority to the other side. If the opinion was relevant to the topic of this specific panel, higher representation in future panels might not compensate for a bad decision.

We cannot directly stratify for these opinions either since they are hidden features—informed opinions will only be revealed through information and deliberation. Furthermore, we might not even know which opinions will be relevant, and the opinions could be too numerous to stratify for all of them.

Nonetheless, we show that stratification is an effective tool for promoting fair representation of such groups, by reducing the variance of their representation. Taking the perspective of the organizer of a citizens’ panel, our aim is to

... characterize the effect of stratification on the variance of the number of representatives from unknown groups, and to demonstrate that this knowledge can help more accurately represent opinion holders.
The key insight underlying our work is that the benefits of stratification extend, beyond the strata themselves, to all groups which correlate with the strata. Informally, this observation has been made as early as 1972 [86]. Consider our earlier example of stratification by gender. Let $M$ capture a certain political opinion held by, say, half of the population, and let a random variable $A_M$ denote the number of agents from $M$ in the panel. If the opinion is highly prevalent among women and rare among men, the distribution of $A_M$ is much more concentrated with stratification than without, as can be seen in Fig. 3a. Quite amazingly, there is virtually nothing to be lost in stratifying: If, as in Fig. 3b, $M$ is equally split between the strata—which is the worst case—stratification only increases the variance by a minuscule amount.

Beyond our toy example, these observations suggest a strategy of more elaborate stratification than is common today: When planning a citizens’ panel, an organizer can use public data to partition the population into many small strata, each corresponding to a small number of seats in the panel. The goal is to group together citizens who are as similar as possible. Since demographic data is highly predictive of political views [88, p. 92], most $M$ of interest should “polarize” the strata, i.e., most strata should either have a very high or very low prevalence of $M$, in which case we profit from the reduction in variance observed above. To do so, a panel organizer needs to reason about the variance of $A_M$, which can be difficult.

Our main contribution is a tight bound showing that—up to a factor very close to one—stratification cannot increase variance. While this is known in the statistics literature for sampling from a continuous population, our setting implies additional steps of rounding—if women make up for 51% of the electorate, they should randomly receive 10 or 11 out of 20 seats—and these rounding steps might increase the variance of $A_M$. We propose block rounding, a way of randomly rounding seat allocations for the strata, which maintains equal probability of participation for all agents. We show that the variance generated by block rounding nicely dovetails with the variance generated by the stratification itself, such that a uniform and tight upper bound can be proved:

**Theorem 9.** For some $n$ and $k$, let $A(N, k)$ be a stratifying algorithm based on block rounding with an arbitrary stratification satisfying that that every stratum $i$ has size $n_i \geq \frac{N}{k}$ (i.e., that the expected number $k_i$ of selected representatives from $i$ is at least 1). Then, for any $M \subseteq N$, we have...
let $A_M$ denote the random variable denoting the number of panel members from $M$ using the algorithm $A(N,k)$, and let $U_M$ denote the same variable but using uniform sampling as the algorithm. Then,

$$\text{Var}(A_M) \leq \frac{n-1}{n-k} \text{Var}(U_M).$$

Note that, since $n$ is typically much larger than $k$, the factor of $(n-1)/(n-k)$ is very close to one. To gauge the benefits of stratification, we give a second upper bound which characterizes the reduction in variance due to stratification in terms of the concentration of $M$ in every stratum. Given a set $M$ and a specific stratum $S_i$, define $\epsilon_i$ as $|M \cap S_i| - |S_i|/n |M|$, i.e., as the difference between the actual number of agents from $M$ in the stratum and the number that we would get if the $|M|$ agents were proportionally distributed over the strata. Note that the sum of $\epsilon_i$ is always 0. If all $\epsilon_i$ equal 0, the stratification is essentially useless. The more the fraction of $M$ agents differs between strata, i.e., the more $\epsilon_i$ are far from 0, the larger the reductions in variance due to stratification:

**Theorem 10.** In the setting of Theorem 9, assume furthermore that every stratum has size $n_i \geq \alpha \frac{n}{k}$ for some constant $\alpha \geq 1$. Then, for any $M \subseteq N$,

$$\text{Var}(A_M) \leq \frac{n-1}{n-k} \text{Var}(U_M) - \left(1 - \frac{1}{\alpha}\right) \frac{k}{n} \sum_i \frac{\epsilon_i^2}{n_i}.$$

Next, we explore the space of sampling algorithms that uphold expected representation for all groups, but are not necessarily based on stratification. We show that no such algorithm dominates any other algorithm:

**Theorem 11.** Let there be two algorithms $A(N,k)$ and $A'(N,k)$, and let their corresponding random variables be called $A_M$ and $A'_M$, respectively. If $\text{Var}(A_M) < \text{Var}(A'_M)$ for some $M$, there is an $M'$ such that $\text{Var}(A_{M'}) > \text{Var}(A'_{M'})$.

Furthermore, we show that uniform sampling is optimal from a worst-case perspective:

**Theorem 12.** Uniform sampling minimizes $\max_{M \subseteq N} \text{Var}(A_M)$ among all algorithms.

Since no sampling algorithm is best, assumptions like ours—that relevant $M$ will correlate with visible features—are needed. Since stratified sampling with block sampling can never increase variance by more than a minuscule amount over uniform sampling, our stratified sampling algorithm provides close-to-optimal worst-case guarantees.

Finally, we investigate the effect of stratification on variance using a large dataset containing information about demographic features and political attitudes. We find that random stratifications are helpful overall, but that only a few of them lead to significant reductions. In a case study simulating the situation of a panel organizer, we find that, using insights from our theoretical analysis, a human stratifier can simultaneously reduce the variance with respect to multiple unknown, attitude-based groups. Compared to uniform sampling, these decreases in variances correspond to an increase in panel size by multiple positions. Manual stratification also clearly outperforms a simple stratification by gender and race. Stratifications automatically generated via $k$-means clustering fall short of the manual stratification, but show promise.
4.3 Limited Participation

In the previous section, we assumed that we could sample from the whole population and that every sampled agent would participate in the panel. Over the past year, we have spoken with several nonprofit organizations whose role it is to sample and facilitate sortition panels [32]. As we have learned, sampling from the whole population is not an option for them due to limited participation—typically, only between 2 and 5\% of citizens are willing to participate in the panel when contacted. Moreover, those who do participate exhibit self-selection bias, i.e., they are not representative of the population, but rather skew toward certain groups with certain features.

![Sampling process in current practice](image)

Figure 4: Sampling process in current practice

To address these issues, sortition practitioners introduce additional steps into the sampling process (see Fig. 4). Initially, they send a large number of invitation letters to a random subset of the population, the recipients. If the recipients are willing to participate in a panel, they can opt into a pool of volunteers. Ultimately, the panel of size $k$ is sampled from the pool. Naturally, the pool is unlikely to be representative of the population, which means that uniformly sampling from the pool would yield panels whose demographic composition is unrepresentative of that of the population. To prevent grossly unrepresentative panels, many practitioners impose quotas on groups based on orthogonal demographic features such as gender, age, or residence inside the country. These quotas ensure that the ex-post number of panel members belonging to such a group lies within a narrow interval around the proportional share. Since it is hard to construct panels satisfying a set of quotas, practitioners typically sample using greedy heuristics. While these heuristics tend to be successful at finding valid panels, the probability with which an individual is selected is not controlled in a principled way.

Since individual selection probabilities are not deliberately chosen, the current panel selection procedure gives up most of the fairness guarantees associated with sortition via sampling from the whole population. Where uniform sampling selects each person with equal probability $k/n$, currently-used greedy algorithms do not even guarantee a minimum selection probability for members of the pool, let alone fair “end-to-end” probabilities with which members of the population will end up on the panel. As a further downside, the greedy algorithms we have seen being applied may need many attempts to produce a valid panel and might take exponential time to produce a valid panel even if one exists.

Our main contribution is a more principled sampling algorithm that, even in the setting of limited participation, retains the individual fairness of sampling without replacement while allowing the deterministic satisfaction of quotas. In particular, our algorithm satisfies the following desiderata:

- **End-to-End Fairness:** The algorithm selects the panel via a process such that all members of the population appear on the panel with probability asymptotically close to $k/n$. This also implies that all groups in the population have near-proportional expected representation.
- **Deterministic Quota Satisfaction**: The selected panel satisfies certain upper and lower quotas enforcing approximate representation for a set of specified features.

- **Computational Efficiency**: The algorithm returns a valid panel (or fails) in polynomial time.

Deterministic quota satisfaction is a guarantee of group fairness, while end-to-end fairness, which recovers most of the ex ante guarantees of sampling without replacement, can be seen primarily as a guarantee of individual fairness. The phrase *end-to-end* refers to the fact that we are fair to individuals with respect to their probabilities of going from *population* to *panel*, across the intermediate steps of being invited, opting into the pool, and being selected for the panel.

The key challenge in satisfying these desiderata is self-selection bias, which can result in the pool being totally unrepresentative of the population. In the worst case, the pool can be so skewed that it contains no representative panel—in fact, the pool might not even contain $k$ members. As a result, no algorithm can produce a valid panel from every possible pool. However, we are able to give an algorithm that succeeds with high probability, under weak assumptions mainly relating the number of invitation letters sent out to $k$ and the minimum participation probability over all agents.

Crucially, any sampling algorithm that gives (near-)equal selection probability to all members of the population must reverse the self-selection bias occurring in the formation of the pool. We formalize this self-selection bias by assuming that each agent $i$ in the population agrees to join the pool with some positive participation probability $q_i$ when invited. If these $q_i$ values are known for all members of the pool, our sampling algorithm can use them to neutralize self-selection bias. To do so, our algorithm selects agent $i$ for the panel with a probability (close to) proportional to $1/q_i$, conditioned on $i$ being in the pool. This compensates for agents’ differing likelihoods of entering the pool, thereby giving all agents an equal end-to-end probability. On a given pool, the algorithm assigns marginal selection probabilities to every agent in the pool. Then, to find a distribution over valid panels that implements these marginals, the algorithm randomly rounds a linear program using techniques based on discrepancy theory. Since our approach aims for a fair *distribution* of valid panels rather than just a single panel, we can give probabilistic fairness guarantees.

We give an algorithm which ensures, under natural assumptions, that every agent ends up on the panel with probability at least $(1 - o(1)) k/n$ as $n$ goes to infinity. Furthermore, the panels produced by this algorithm satisfy non-trivial quotas, which ensure that the ex-post representation of each feature-value pair cannot be too far from being proportional.

Our algorithm proceeds in two phases: I. *assignment of marginals*, during which the algorithm assigns a marginal selection probability proportional to $1/q_i$ to every agent in the pool, and II. *rounding of marginals*, in which the marginals are dependently rounded to 0/1 values, the agents’ indicators of being chosen for the panel. As we discussed previously, our algorithm succeeds only with high probability, rather than deterministically; it may fail in phase I if the desired marginals do not satisfy certain conditions. Fortunately, under reasonable conditions, we show that the pool will satisfy all required conditions with high probability. Then, in Phase II, we randomly round the marginals using ideas based on the Beck-Fiala theorem [15].

Let $n$ be the size of the population $N$, $r$ the number of recipients, and $k$ the size of the panel. Fix a set of features (e.g., gender and education), each with its corresponding values
corresponding values (e.g., male and female for gender; college and no college for education), and for each feature-value pair \( f, v \), let the number of agents in the population with that value be \( n_{f,v} \). Finally, set \( \alpha := (\min_{i \in N} q_i) r/k \). Then, our algorithm satisfies the following theorem, guaranteeing close-to-equal end-to-end selection probabilities for all members of the population as well as the satisfaction of quotas.

**Theorem 13.** Suppose that \( \alpha \to \infty \) as \( n \to \infty \) and that \( n_{f,v} \geq n/k \) for all feature-value pairs \( f, v \). Then, for the random panel \( \text{Panel} \) produced by our sampling algorithm,\(^8\) it holds for all \( i \) in the population that

\[
P[i \in \text{Panel}] \geq (1 - o(1)) \frac{k}{n}.
\]

All non-empty panels produced by this process have size \( k \) and contain at least \( \ell_{f,v} := (1 - \alpha^{-49}) \frac{k n_{f,v}}{n} - |F| \) agents of each feature value pair \( f, v \) and at most \( u_{f,v} := (1 + \alpha^{-49}) \frac{k n_{f,v}}{n} + |F| \) of them.

The guarantees of the theorem grow stronger as the parameter \( \alpha = q^* r/k \) tends toward infinity, i.e., as the number \( r \) of invitations grows.

As we mentioned, our theoretical and algorithmic results take the probabilities \( q_i \) of all pool members \( i \) as given in the input. While these values are not observed in practice, we show that they can be estimated from available data using maximum-likelihood estimation. We cannot directly train a classifier predicting participation, however, because practitioners collect data only on those who do join the pool, yielding only positively labeled data. In place of a negatively labeled control group, we use publicly available survey data, which is unlabeled (i.e., includes no information on whether its members would have joined the pool). To learn in this more challenging setting, we use techniques from contaminated controls, which combine the pool data with the unlabeled sample of the population to learn a predictive model for agents’ participation probabilities. We validate our approach on data from a real sortition panel.

### 5 Refugee Resettlement

The three preceding sections aimed to enable society to identify the right people for making decisions, without specifying what decisions were being made. By contrast, this section and the next zero in on specific questions and on how to make decisions that promote welfare.

Migration is one of the greatest societal challenges facing humanity in the 21st Century. In 2015 there were 244 million international migrants in the world, suggesting a larger increase in the rate of international migration than was previously anticipated [80]. A key reason behind this increase is the widely understood fact that migration often has significant benefits to the migrants and their families, as well as the host country and the country of origin. For example, migrants working in the United States earn wages that are higher by a median factor of 4.11 than those they would have earned in their home countries [38]; and the money they send back to their countries of origin is a reliable source of foreign currency. But the increase in the rate of migration is also driven by globalization, social disparities, and, in no small part, by a vast number of refugees— including, most recently, millions who have fled war in Syria, persecution in Myanmar, and economic calamity in Venezuela.

---

\(^8\)Panel might be empty when the algorithm fails.
These events have fueled a surge of attempts to put migration, and, especially, refugee resettlement, on scientific footing. Alvin Roth, a Nobel laureate in economics, nicely summarizes the problem in a 2015 opinion piece [90]:

“Refugee relocation is what economists call a matching problem, in the sense that different refugees will thrive differently in different countries. Determining who should go where, and not just how many go to each country, should be a major goal of relocation policy.”

This observation underlies work in market design, which draws on the extensive literature on matching problems such as school choice. It focuses on refugee resettlement mechanisms that elicit refugees’ preferences over localities, and output a matching that is desirable with respect to these preferences [84, 66, 41].

By contrast, a recent paper by Bansak et al. [13], published in Science, takes a markedly different, data-driven approach to refugee resettlement, which employs machine learning and optimization. Their goal is to maximize the expected number of refugees who find employment. To be precise, they consider families and maximize the expected number of families such that at least one member is employed, but this does not make a significant technical difference, so we focus on individuals for ease of exposition.

Using historical data from the United States and Switzerland, they train predictors that estimate the probability $p_{i \ell}$ that a given refugee $i$ would find employment in a given locality $\ell$. The optimal solution is then an assignment of refugees to localities that maximizes the sum of probabilities, subject to capacity constraints. Assuming their predictions of employment probabilities are accurate, Bansak et al. demonstrate that their approach leads to a $40\%-70\%$ increase in the number of employed refugees, compared to the actual outcomes in the United States and Switzerland.

On a high level, we subscribe to the data-driven approach, and believe that the assumptions made by Bansak et al. [13] are quite reasonable in the context of refugee resettlement — most notably, the implicit assumption that the probability $p_{i \ell}$ can be estimated based only on information about the refugee $i$ and the locality $\ell$, and does not depend on where other refugees are assigned. But as this approach gains traction, we envision it being deployed on a larger scale, and ultimately informing international migration policy more broadly, especially in the context of labor migration.

The key observation behind our work is that, at that scale, competition effects would invalidate the foregoing assumption. Indeed, the larger the number of, say, engineers, who settle in a specific area, the less likely it is that any particular engineer would find a job. The immigration of approximately 331,000 Jews from the former Soviet Union to Israel in 1990 and 1991 serves as a case in point. A disproportionate number of these highly-educated immigrants were engineers and doctors, leading to a saturation of the local job markets: “a state with an oversupply of doctors could not possibly double the number of its doctors in several years” [94]. Our goal in this paper, therefore, is to enhance the data-driven approach to migration by explicitly modeling competition effects, and directly optimizing for them.

5.1 Related Work

Our work is most closely related to that of Ahmed et al. [1]. Motivated by diversity in problems such as movie recommendation and assignment of papers to reviewers, they formulate
a weighted bipartite $b$-matching problem with a specific submodular objective. They show that this problem can be formulated as a quadratic program (which may be intractable at a large scale). They also use the greedy algorithm, but only note in passing that it would give worst-case guarantees in a degenerate special case that essentially coincides with having no constraints. By contrast, our worst-case guarantees hold for any \emph{approximately} submodular objective function \emph{under capacity constraints}, or even under an arbitrary number of matroid constraints. Another key difference is that Ahmed et al. focus on one specific objective function, whereas we explore several different models, which are tailored to the migration domain. Perhaps the most striking difference is more fundamental: For Ahmed et al., diversity is an objective that is orthogonal to (additive) efficiency. By contrast, we consider diversity as an inherent part of efficiency since matchings lacking in diversity will suffer from competition effects.

A bit further afield, there is a large body of work on submodular optimization, and its applications in AI. The papers of [49] and [61] are especially relevant. Applications of submodular optimization include influence in social networks [68], sensor placement [74], and human computation [92], just to name a few.

5.2 Migration as Submodular Optimization

On a technical level, our objective function $f$ receives a set of migrant-locality pairs as input, and returns the predicted number of employed migrants under the corresponding assignment. Crucially, we assume that $f$ is (monotone) submodular: individual elements provide diminishing marginal returns. Formally, if $S$ and $T$ are two subsets of migrant-locality pairs such that $S \subseteq T$, and $(i, \ell)$ is a migrant-locality pair such that $(i, \ell) \notin T$, then

$$f(S \cup \{(i, \ell)\}) - f(S) \geq f(T \cup \{(i, \ell)\}) - f(T).$$

This captures the idea that the larger the number of migrants that compete with $i$ for a job at locality $\ell$, the less likely it is that $i$ herself would find employment — especially if it is a skilled job — and the smaller the marginal contribution of $(i, \ell)$ to the objective function (overall number of migrants who find employment).

We can therefore cast our optimization problem as the maximization of a monotone submodular function subject to matching constraints, which represent caps on the number of migrants each locality can absorb.\footnote{Capacity constraints can be transformed into matching constraints on the complete bipartite graph with migrants on one side, and localities on the other, where the number of copies of each locality is equal to its capacity.} In fact, we allow the objective function to be \emph{approximately} submodular since most ways of estimating employment for a matching will introduce noise that invalidates submodularity.

The matching constraints in question can be naturally described as the intersection of two matroids. Adapting a classical result by Fisher et al. [49], we show that a simple greedy algorithm — which is known to perform well when given access to an exactly submodular function — gives an approximation guarantee of $(P + 1 + \frac{4\epsilon}{1-\epsilon} k)^{-1}$ when maximizing a submodular function $f$ subject to $P$ many matroids if we only have access to an $\epsilon$-approximately submodular function approximating $f$ and if $k$ is the size of the largest set in the intersection of the matroids:
**Theorem 14.** Let \( z : 2^N \to \mathbb{R} \) be \( \epsilon \)-approximately submodular and let \( \hat{z} \) be the underlying (monotone, normalized) submodular function, i.e., let

\[
(1 - \epsilon) \hat{z}(S) \leq z(S) \leq (1 + \epsilon) \hat{z}(S)
\]

for all \( S \subseteq N \). Let \( \mathcal{F} \) be the intersection of \( P \) matroids, and let \( k \) denote the size of the largest \( S \in \mathcal{F} \). Then the greedy algorithm selects a set \( S \in \mathcal{F} \) such that

\[
\left( P + 1 + \frac{4\epsilon}{1 - \epsilon} k \right) \hat{z}(S) \geq \max_{S' \in \mathcal{F}} \hat{z}(S').
\]

In our setting, this result gives us a guarantee—for \( P = 2 \)—on the performance of our greedy matching with respect to the optimal matching.

The submodular objective function can potentially be learned, or optimized directly, from historical data [9, 11] without any further structural assumptions. As an alternative, we propose three models of how migrants find employment, all of which induce submodular objective functions. The purpose of these models is twofold: First, they represent different ways in which competition effects might arise, thereby helping to understand them. Second, in practice they may prove to be more accurate as they are defined by relatively few parameters, and are easier to train.

We compare the employment generated by the greedy heuristic on the submodular model to the baseline approach of assuming additivity. We find that the benefits of accounting for submodularity almost always outweigh the loss associated with using an inexact optimization technique. Across our three models and a variety of settings, the greedy approach frequently increases employment by 10% and more, suggesting that substantial gains can be made in practice.

### 6 Kidney Exchange

We now transition to a second problem of resource allocation, kidney exchange. For many years, the patients in need of a kidney transplant have far outnumbered the organs available from deceased donors [81]. While many patients can find a friend or relative willing to donate a live organ to them, medical incompatibilities frequently prevent such a direct donation. Kidney-exchange platforms address this problem by matching patients to kidneys across multiple patient–donor pairs. The overarching goal in designing these platforms is to maximize the number of patients who receive transplants, i.e., the welfare of the matching. The input to the mechanism can be represented as a *compatibility graph*, a directed graph whose nodes represent patient–donor pairs, and where an edge from one pair to another indicates that the first pair’s donor can donate to the second pair’s patient.

Traditionally, research on kidney exchange has focused on matchings\(^{11}\) that have the form of a union of disjoint 2-cycles and 3-cycles. Cycles are a natural choice because the paired donors can only be expected to donate a kidney if their patient receives one; the short length of cycles is due to the fact that all surgeries have to proceed simultaneously to prevent donors from

\(^{11}\)As is customary in the kidney-exchange literature, we use the word “matching” in a more general sense than in most of computer science. A matching is a set of edges in a directed graph such that each node has indegree and outdegree at most one, and for each node (except for the altruist, which we define later), the outdegree is at most the indegree.
reneging. In this body of work, compatibility graphs are often modeled by classes of densely connected random graphs. If these random graphs are sufficiently large, 2-cycles and 3-cycles can match essentially all pairs that one could hope to match [2]. However, these properties of the random model do not reflect the observation that, in real kidney-exchange platforms, more complex matchings can match substantially more patient–donor pairs [4].

Indeed, subsequent work [4, 43] has highlighted that welfare can be increased substantially by one such form of matchings in particular: long chains starting at non-directed altruistic donors. These donors, which we simply refer to as altruists in the rest of the paper, are donors willing to donate a kidney without requiring a transplant in exchange. When an altruist donates a kidney, a compatible patient can receive a kidney before her donor donates to another patient. Thus, even if one of the donors should renege, no pair will have donated without having received a kidney in return, which means that simultaneous transplantations are no longer required. Since donors still renege very infrequently [39], an altruistic donor can in principle be at the beginning of an arbitrary-length chain of donations.12

To take full advantage of the path started by an altruist, it is in general necessary to extend this path across patient–donor pairs in multiple hospitals. However, finding such a high-welfare path across hospitals is complicated by the fact that the centralized path-finding mechanism cannot directly observe the patient–donor pairs at each hospital and must rely on hospitals to accurately report them. Since hospitals feel primarily bound to their own patients’ well-being, a hospital might choose not to register certain pairs if this will lead the platform to match more of its other pairs. Thus, the problem of finding long paths in kidney exchange is one of mechanism design, i.e., we have to design mechanisms that incentivize hospitals to truthfully report all their clients, which is known as incentive compatibility.13

On worst-case graphs, simple counterexamples show that welfare maximization and incentive compatibility are incompatible (??). This impossibility holds even for generous approximate notions of welfare maximization and incentive compatibility. On the other extreme, if the compatibility graphs were purely random Erdős-Rényi graphs, the problem would become easy for most parameter settings: either there are no long paths or there is a Hamiltonian cycle. Unfortunately, in practice, kidney-exchange compatibility graphs do not resemble Erdős-Rényi graphs due to issues such as sensitization and the dynamic nature of how patients enter and exit the system.

Adopting a semi-random approach, we show that one can obtain interesting positive results by adding a small number of random edges to worst-case graphs. Notably, edges may be added with probability $1/n^c$ for $1 < c < 2$, meaning that only a vanishing fraction of vertices will be the endpoint of a random edge. To our knowledge, these are the first positive results for semi-random graphs with edge probabilities in $o(1/n)$.12

In contrast to purely random models, our semi-random model captures a wide range of compatibility graphs, including very sparse and heterogeneous ones. While the random edges do not have a direct correspondence in biology, we interpret them as a mild regularity condition on the graph, expressing that any two large subsets of the graph are unlikely to have no edges between them. Furthermore, these random edges are only assumed to exist between pairs of different hospitals, which means that the hospital subgraphs can be purely worst case, somewhat like in the semi-random model of Makarychev, Makarychev, and Vijayaraghavan [78].

As of June 2020, one particular chain has led to 114 successful transplants and is still ongoing.12

For ease of exposition, we assume that hospitals can hide nodes, not individual edges. However, as discussed in ??, our results generalize to the case where edges can also be hidden.13

[25]
6.1 Related Work

Whenever a problem is intractable for worst-case inputs or specific input distributions are hard to justify, semi-random models can allow for positive results while retaining substantial generality.

In most works that apply semi-random models, the motivating intractability is computational complexity. For instance, Blum and Spencer [20] pioneered semi-random models in the context of 3-coloring, which is NP hard in the worst case. One of their models starts from an adversarially chosen 3-colored graph and then, independently for each pair of differently colored nodes, toggles the existence of an edge between the two nodes with some small probability $p$. If $p \geq n^{-0.6+\epsilon}$ (and all colors have $\Omega(n)$ nodes), an efficient algorithm 3-colors the resulting graph with high probability. Since then, many authors have investigated other computational problems in related models [47, 73, 53]; we refer the reader to Feige [46] for a survey of the literature. Semi-random models are also used in smoothed analysis [96, 45], again driven by complexity considerations. As in our work, one can understand positive results as saying that worst-case instances have to be constructed in a very brittle way, which hopefully makes them unlikely to appear in practice.

Only in few cases have semi-random models been applied for desiderata other than complexity. The most prominent example is the use of random-order models for online algorithms [58, 70, 69], where a random arrival order makes the problem more tractable in an information-theoretic sense. In certain versions of the secretary problem, for instance, an adversary chooses a set of quality scores, but the order in which the algorithm sees these scores is chosen uniformly at random. Here, assuming a random arrival order is not a tool for easier computation; instead, it increases the best approximation ratio of any algorithm from $1/n$ to $1/e$. Babaioff, Immorlica, and Kleinberg [8] combine these ideas with mechanism design in order to design auctions that are incentive compatible when a set of buyers with worst-case valuations arrive in random order. The fact that monetary transfers are available in this setting gives their mechanisms a decidedly different flavor than ours.

We are not aware of existing work on semi-random models that gives positive results with on edge probabilities as low as ours. As two examples of work giving relatively low edge probabilities, Makarychev et al. [78] find planted balanced cuts for edge probabilities in $\text{polylog}(n)/n$ for balanced cuts; Chen, Sanghavi, and Xu [34] recover clusters for a similar range of probabilities in a different semi-random model. By contrast, our approach can handle edge probabilities as low as $1/n^{2-\epsilon}$.

There is a rich body of research on algorithm and mechanism design for kidney exchange [18, 2, 99, 5, 59, 21], some of which we have already mentioned. Ashlagi and Roth [3] identify hospital incentives and chains started by altruistic donors as the two main challenges in increasing the welfare of kidney exchange in practice. If matchings are restricted to unions of 2-cycles, Ashlagi, Fischer, Kash, and Procaccia [5] give an incentive-compatible mechanism that approximates optimal welfare by a factor of 2. In a repeated setting, Hajaj, Dickerson, Hassidim, Sandholm, and Sarne [59] give a mechanism that is incentive compatible, maximizes welfare, and includes long chains from altruistic donors. Because matchings happen in infinitely many rounds and the stochastic arrival rate of pairs at each hospital is publicly known, the mechanism can punish hospitals for hiding nodes by withholding matches in subsequent rounds. This is not possible in our setting.

Our work is most closely related to that of Blum et al. [21], who design mechanisms for kidney
exchange in a different semi-random model. Their model is semi-random in that, in a worst-case compatibility graph, the nodes are allocated randomly to the hospitals. For matchings composed of disjoint constant-length cycles, their mechanism guarantees optimal welfare and individual rationality up to lower-order terms, a much weaker notion than matching-time incentive compatibility. Blum et al. acknowledge that, unfortunately, their approach “does not seem to extend beyond individual rationality,” and they provide an impossibility result showing that individual rationality cannot be obtained if long paths from altruistic donors are allowed.

6.2 Mechanism Design for Kidney Exchange in a Semi-Random Model

Let there be a finite, directed base graph $G_{base} = (V, E_{base})$, where $n := |V|$. Let the vertices $V$ be partitioned into subsets $V_i$ for each of finitely many hospitals $i$. Let $E_p$ denote the random variable ranging over subsets of $\bigcup_{i \neq j} (V_i \times V_j)$, where each potential edge is included in $E_p$ independently with some given edge probability $p$. While our results hold for a wider range of $p$, we are primarily interested in very low values of $p$, which might, for example, scale as $1/n^c$ for $1 < c < 2$. The compatibility graph is defined by adding the random edges to the base graph, i.e., as $G := (V, E_{base} \cup E_p)$. Let one node in $G$ be labeled as the altruist $\alpha$. The hospital owning the altruist node is called the altruist owner. We consider an instance to be the base graph together with node ownership, the value of $p$, and the identity of the altruist.

Each hospital $i$ reports a subset $V'_i \subseteq V_i$ to the mechanism. If $V'_i = V_i$, we say that $i$ reports truthfully. The mechanism then learns about all edges in the subgraph of $G$ induced by $\bigcup V'_i$, which means that the mechanism knows this subgraph, node ownership and the identity of the altruist, but does not have access to the remaining nodes and edges, to whether edges are random or deterministic, and to the value of $p$. Based on this knowledge, the mechanism selects a (simple) path starting at the altruist.

Let the length $|\pi|$ of a path $\pi$ be its number of nodes, and let the total length $||P||$ of a set of vertex-disjoint paths $P$ be the sum of their lengths. We say that paths are disjoint when they are vertex disjoint. We refer to the length of the path produced by the mechanism as the welfare, and to the number of vertices in this path that belong to a certain hospital as the utility of this hospital.

We say that edges between nodes of the same hospital are internal edges; all other edges are cross edges. A path entirely consisting of internal edges of hospital $i$ is called an internal or $i$-internal path, and we refer to the maximal internal subpaths of a path as the segments of the path.

The welfare of a mechanism can be measured relative to multiple benchmarks, all of whom we define on the full compatibility graph (rather than just the base graph). We define our strongest benchmark $Opt$ as the longest path starting at the altruist. We also define two weaker benchmarks. For some $s$, let the high-averages benchmark, $AvgOpt^s$, be the longest path from the altruist such that each hospital has either no segments in the path, or her mean segment length is at least $s$. This weaker benchmark might not be defined; in this case,

\footnote{For well-definedness, we assume that the altruist is always reported. This is without loss of generality: If the altruist is hidden, the mechanism cannot output anything and, thus, no hospital can receive more utility from the mechanism. The more promising strategy of using the altruist for internal matching rather than bringing it to the mechanism will be captured by the definition of matching-time incentive compatibility even without hiding nodes.}
consider the benchmark as the empty path.

We search for a mechanism that, with high probability over the random edges of a given instance, satisfies the following desiderata:

**Efficiency:** If all hospitals report truthfully, the welfare of the mechanism is at least a \((1 - o(1))\) fraction of the welfare of the high-averages benchmark.

**Incentive Compatibility:** If all hospitals report truthfully except for hospital \(i\), \(i\)’s utility from the path produced by the mechanism is at most a \((1 + o(1))\) fraction of its utility when all hospitals report truthfully.

**Individual Rationality:** Under truthful reports, the utility of the altruist owner is always at least a \((1 - o(1))\) fraction of the length of its longest internal path \(\pi_{IR}\) starting at the altruist.

**Matching-Time Incentive Compatibility:** As for regular incentive compatibility, let \(\pi\) be the path produced by the mechanism when all hospitals except for possibly \(i\) report truthfully. Not only must \(i\)’s utility from \(\pi\) be at most a \((1 + o(1))\) fraction of its truthful utility, but the same must be true for any path \(\pi’\) created by diverting \(\pi\) at one of \(i\)’s nodes \(v\) into an internal continuation.\(^{15}\)

Note that matching-time incentive compatibility implies incentive compatibility. Matching-time incentive compatibility also generalizes individual rationality since the altruist owner may obtain her longest internal path \(\pi_{IR}\) starting at the altruist by reporting all her nodes and diverting the path at the altruist into \(\pi_{IR}\).

Even in our semi-random model, maximum welfare is at odds with incentive compatibility:

**Proposition 15.** For any \(p \in o(1/n)\), there exists a family of instances of size going to infinity such that every deterministic mechanism will, with \(\Omega(1)\) probability, on any instance either obtain \(\Theta(n)\) less welfare than the maximum-welfare path or give some hospital \(\Theta(n)\) incentive for hiding nodes.

The instances constructed in the proof are problematic in the sense that any high-welfare path must visit one hospital’s subgraph many times but spend few steps there each time, a scenario which incentivizes hospitals to hide outgoing edges to other hospitals. It is natural, then, that we can overcome this impossibility result by benchmarking our mechanism against the longest path such that, for each hospital, the subpaths the hospital’s subgraph have some minimum average length.

The random edges in our model are so sparse that they, by themselves, do not form paths of significant length. The main insight, however, is that they allow us to “stitch together” sufficiently long paths of worst-case edges inside different hospitals. Indeed, suppose that \(s\) is a parameter such that \(s^2 p \gg 1\). As illustrated in Fig. 5, let there be two paths \(\pi_1, \pi_2\) of length greater than \(s\), each internal to a different hospital. Then, the expected number of edges from the the \(s\) last nodes of \(\pi_1\) to the first \(s\) nodes of \(\pi_2\) is high. If any of these edges do exist, we can use them to “stitch” \(\pi_1\) and \(\pi_2\) together. Then, nearly all nodes of \(\pi_1\) and \(\pi_2\) will lie on the stitched path except at most \(2s\) nodes in the suffix of \(\pi_1\) and the prefix of \(\pi_2\).

\(^{15}\)Formally: For some node \(v \in V_i\) on \(\pi\), write \(\pi = \pi_1 v \pi_2\) for the prefix and suffix, and let \(\pi_{ext}\) be an internal path of hospital \(i\), starting at \(v\), disjoint from \(\pi_1\), and not necessarily contained in the nodes reported by \(i\). Then, \(\pi' = \pi_1 \pi_{ext}\).
Roughly speaking, our mechanism has the following structure: From the graph reported by the hospitals, the mechanism chooses a set of disjoint paths within each hospital so that total length is large, and so that truthful reporting maximizes the total length of a hospital’s chosen paths. Then, the mechanism stitches these internal paths together, which will succeed with high probability. Successfully stitched paths preserve the welfare and hospital utility except for small stitching losses, which gives approximate guarantees on welfare and incentive compatibility.

Since our mechanism can stitch long enough paths with high probability, it can select the internal paths of hospital $i$ (mostly) based on the subgraph reported by hospital $i$. As a result, each hospital wants to reveal all its nodes to increase the total length of its chosen paths. If, by contrast, the mechanism chose its path by how well nodes are connected via cross edges, hospitals might hide connections to other hospitals to sway the mechanism to match more of their own nodes instead.

Challenges in translating this high-level structure into concrete mechanisms include:

- **The choice of original paths.** The paths within each hospital initially selected by the mechanism must compete with the benchmark in terms of welfare, and must lend themselves to stitching.
- **The number of paths per hospital.** Stitching only works if we alternate the hospitals on a chain, so we cannot stitch too many paths by a single hospital.
- **Internal paths too short for stitching.** The altruist might only be able to reach long paths via an initial sequence of short internal paths connected by worst-case cross edges. In this case, we do have to consider worst-case cross edges in deciding where to stitch while also ensuring that incentives are limited.

Our main result is the following:

**Theorem 16.** Let there be at least two hospitals. Let $s, p$ be parameters varying in $n$ such that $ps^2 \in \omega(\log(n)^5)$. Let $\text{AvgOpt}^s$ be the longest path from the altruist such that, for each hospital, the average length of the subpaths in this hospital’s subgraph is at least $s$ (or the hospital does not own any nodes on the path).

Then, there is a mechanism that, with high probability, produces a path of length at least $(1 - o(1)) |\text{AvgOpt}^s|$ assuming truthful responses, and that ensures that each hospital can only increase its utility by lower-order terms by manipulation, including matching-time manipulation.
7 Future Directions

We conclude by presenting three directions of current and future work, each building up on work presented in the previous sections of the proposal.

7.1 Sortition with Fairness to the Pool

In Section 4.3, we gave an algorithm for sampling sortition panels in the face of limited participation. Based on knowledge of individual participation probabilities $q_i$, we were able to give guarantees on the end-to-end probability of an agent making it from the population to the panel. However, we see two obstacles to the practical implementation of this algorithm:

- First, the fairness and transparency of the process crucially relies on the accuracy of the $q_i$ estimates. A member of the pool (who only observes this one pool) might not believe that it is fair that our algorithm assigns her much lower marginal probability than to a second agent in the pool.

- Second, practitioners precommit to much stricter quotas than what the Beck-Fiala rounding can guarantee.\textsuperscript{16} Even deciding whether a certain set of marginal probabilities is implementable for a given pool and given quotas is NP hard, which—even ignoring computational cost—is a hurdle in proving that a random pool will allow an algorithm to succeed. Thus, end-to-end guarantees are hard to give in this setting.

Due to these obstacles, we have started developing algorithms that only provide individual fairness guarantees to the members of the pool, but which might be more directly applicable. That is, rather than equalizing end-to-end probabilities, we aim to equalize the marginal selection probabilities $\pi_{i,P}$ for a fixed pool $P$ within the constraints of our externally agreed quotas. Because we are essentially allocating the scarce resource “marginal selection probability” over the pool members, we inherit fairness objectives and axioms from the fair division literature.

We currently have implemented two algorithms, MaxiMin and Nash, which respectively maximize the minimum marginal selection probability and the product of marginal selection probabilities. While neither runs in polynomial time,\textsuperscript{17} these algorithms use practically efficient optimization techniques such as integer linear programming, column generation, and convex optimization, and run in reasonable time on realistic inputs. Both algorithms are currently implemented in the open source sampling software\textsuperscript{18} used by the Sortition Foundation and others, and have been successfully used in sampling multiple sortition panels.

Besides the Sortition Foundation, we started a collaboration with Of By For, an NGO hoping to organize sortition panels on COVID-19 in the United States.\textsuperscript{19} Of By For puts a particular focus on the transparency of the final selection stage, in which a panel is selected from the distribution over panels returned by our algorithm. Because of this, we are thinking about how physical randomness (dice, lottery balls) can be used to make this step more transparent to a general audience.

\textsuperscript{16}Sometimes, there is no gap at all between upper and lower bounds, or maybe a gap of up to 2. By contrast, our theorem will always have a gap of $2|F|$ plus losses due to imperfect concentration.

\textsuperscript{17}Indeed, they solve NP hard problems.

\textsuperscript{18}https://github.com/sortitionfoundation/stratification-app

\textsuperscript{19}https://citizenscouncil.us/
Another big question is how to deal with agents that, after opting into the pool, still refuse to participate when sampled. While this does not seem to be a big problem for the sortition foundation, it seems to be a substantial problem for some of the other NGOs we have talked with.

7.2 Online Refugee Matching

In Section 5, we pointed out one potential issue in AI-based refugee resettlement, i.e., the assumption that employment is modeled as a linear function. We believe that this assumption would be problematic if large numbers of refugees are being matched, so that they have a significant impact on the employment market. However, current practice is marked by a heavily reduced influx of refugees.

With our collaborators at Refugees.ai (who implemented an employment based matching system for the US resettlement agency HIAS), we currently tackle the more pressing challenge of online arrival. Since HIAS only learns about arriving refugees shortly before their arrival and immediately has to place them in a city within the country, offline optimization is not applicable. Based on our experiments, the currently employed greedy matching leads to employment scores that lie between 10 and 20% below the offline optimum. The matching problem has the following characteristics:

- refugees arrive in cases, i.e., families that cannot be separated,
- each locality has a capacity of refugees that it can host (i.e., a knapsack constraint over the sizes of cases),
- we optimize the sum of edge weights in the matching, and
- arrivals are stochastic rather than worst-case, but the distribution is non-static in unpredictable ways.\(^\text{20}\)

These properties put our problems outside of the matching domains that are well understood. While our research is still ongoing, we have seen good performance by algorithms based on pricing the slots in the localities. That is, these algorithms still place cases into localities in a greedy way, but the employment score of a case at a locality is decreased by an estimate of how valuable the slots in that locality would be if they remained free for placing future cases. For example, if the current case would obtain a small employment boost by being placed in Pittsburgh rather than Boston, the algorithm might still place the case in Boston if it expects future cases to profit more from being matched to Pittsburgh and the capacities in Pittsburgh are binding.

We built a synthetic distribution of refugee arrivals based on Chow-Liu trees [36], which allows us to choose between different prices for the slots. So far, techniques based on shadow prices in matching LPs have outperformed more involved approaches based on machine-learning aided optimization.

\(^{20}\)In particular, we have found that policy decisions abruptly change the arrival patterns; international events also change the distribution of characteristics of the arriving refugees.
7.3 Semi-Random Models in Fair Division

In Section 6, choosing a semi-random model enabled us to obtain positive results where purely worst-case models faced hard impossibilities. While guarantees on worst case instances are easier to apply to the real world, semi-random models with low amounts of randomness can still be understood as guarantees holding under mild regularity conditions. As we highlighted in the related work, the general tool of semi-random models has mainly been leveraged against few impossibilities besides computational hardness. Since the field of computational social choice is riddled with impossibility results, we would like to investigate whether some of these impossibilities are brittle under low amounts of randomness.

The concrete problem that we are investigating is the existence of envy-free allocations: Let there be \( n \) agents and \( m \) goods, and let each agent \( i \) have a utility \( u_i(\alpha) \in \mathbb{R}_{\geq 0} \) for each good \( \alpha \). The set of goods has to be partitioned into allocations \( A_i \) for each agent \( i \). We assume that agents have additive utilities, i.e., that agent \( i \) values a bundle \( B \) of goods as \( u_i(B) = \sum_{\alpha \in B} u_i(\alpha) \). An allocation is said to be envy-free if each agent \( i \) weakly prefers her own allocation to every other agent’s allocation, i.e., if \( u_i(A_i) \geq u_i(A_j) \) for all agents \( i \) and \( j \). While envy-free allocations do not exist for all instances—for instance, a single item cannot be fairly split between two agents who want it—allocations satisfying a slightly weaker axiom, envy freeness up to one good, always exist. Furthermore, for random preferences, envy-free allocations exist with high probability \([44]\).

These results give reason to hope that envy-free allocations might also exist if worst-case preferences are randomly perturbed, even if only a vanishing fraction of item utilities is perturbed per agent. If, say, we can show that the round robin mechanism (which guarantees envy freeness up to one good) will return a truly envy-free allocation with high probability, this would give additional justification to using it in the real world.

References


