manuscript No.
(will be inserted by the editor)

A Rule for Committee Selection with Soft Diversity Constraints

Haris Aziz

Abstract Committee selection with diversity or distributional constraints is a ubiquitous problem. However, many of the formal approaches proposed so far have certain drawbacks including (1) computational intractability in general, and (2) inability to suggest a solution for instances where the hard constraints cannot be met. We propose a cubic-time algorithm for diverse committee selection that satisfies natural axioms and draws on the idea of using soft bounds.

Keywords Social choice theory · committee voting · multi-winner voting · diversity constraints · computational complexity
JEL Classification: C70 · D61 · D71

1 Introduction

Selecting a target number of candidates is a ubiquitous problem that occurs in faculty hiring, scholarship selection, corporate board election, and formation of representative bodies (Aziz et al., 2017a,b; Ratliff, 2006; Faliszewski et al., 2017). In many of these settings, there may be natural distributional constraints motivated for example by diversity. For example, in certain European countries, there is a requirement of having a minimum percentage of females in corporate boards. In some school admission guidelines, there are quotas for less-advantaged groups.

Finding the best set of candidates subject to diversity constraints has also been formally studied in social choice. In several works, the problem of diverse committee selection is viewed as the problem with candidates having different (possibly multiple) types and the committee having distributions constraints on each of the types (see e.g., (Brams and Potthoff, 1990; Bredereck et al., 2018; Celis et al., 2018; Potthoff, 1990; Straszak et al., 1993)). There are a few drawbacks of approaches that use hard

H. Aziz
UNSW Sydney and Data61 CSIRO, Sydney 2052, Australia
Tel.: +61-2-8306 0490
Fax: +61-2-8306 0405
E-mail: haris.aziz@unsw.edu.au

distributional constraints. The drawbacks include the following: (1) there may be instances of the diverse committee selection problem that do not admit any feasible solution that satisfies the constraints (for example, there simply may not be enough female applicants) and (2) the hard constraints make the problem of committee selection computationally hard (for example, if we require that each type should have at least one representative, the problem of checking whether there exists a committee satisfying the requirement is NP-complete (see e.g., (Aziz et al., 2016))). Drawback (1) can especially arise if there are many types of applicants and there are diversity targets for all of them. For example, in large-scale admissions programs in India, there are quotas for several types of backgrounds. For approaches that are NP-hard, the lack of a simple polynomial-time algorithm may render them impractical for large enough instances. Even for smaller instances, these approaches cannot be used without resorting to a computer. Finally, not imposing hard constraints may allow for outcomes that are more preferred by more agents.

Some other approaches consider distances between candidates or committees based on their type attributes and then view diversity not as a constraint but as an optimisation objective based on the distances (see e.g., (Kuo et al., 1993; Lang and Skowron, 2016)). The approaches do not generally take into account the excellence of the candidates and the underlying problems are NP-hard. Apart from imposing hard distributional constraints, another approach that is often used in real-life committee selection to achieve diversity is to give bonus points or ranking boosts to candidates who are from under-represented groups.[1] When these rules are imposed centrally, they may come across as arbitrary fixes to solving diversity issues. If the decision makers or voters internally take diversity issues into account while formulating an objective linear ranking, it puts a cognitive burden on the voters to mix diversity prioritisation with objective excellence estimation.

In this paper, we consider the committee selection problem with distributional constraints and focus on the most common constraints whereby at least certain fraction of the candidates should satisfy a given type.[2] Our approach is to view the distributional constraints as *soft* constraints which should be satisfied as much as possible. Often real-life diversity guidelines need not be hard constraints but general rules of thumb to achieve procedural fairness. We present a simple cubic-time algorithm that simultaneously satisfies two axioms called *type optimality* and *justified envy-freeness*. The axiom justified envy-freeness is inspired by the matching market literature. The combination of the two axioms can be viewed as finding a committee that is as close to satisfying the hard distributional constraints and also selecting the best candidates.

2 Setup

The setting involves a set of candidates C, a weak order $\succsim$ over C, a set of types T, a matrix τ that specifies whether a candidate is of a certain type, and a vector $\underline{q}$ that

[1] The model where synergies or presence of diverse agents provide additional points to the committee has been considered in a general model by Izsak et al. (2018).

[2] More general models also allow for expressing upper quotas. The goal of the upper quotas can easily be met by setting lower quotas on the complement of the set of types.

specifies the lower quota bound for each type. A diverse committee selection instance can be summarized as $(C, \succsim, T, \tau, \underline{q})$ where

- $C = \{c_1, \ldots, c_m\}$ is the set of candidates.
- The weak order $\succsim$ over C is the priority order over the candidates.
- $T = \{t_1, t_2, ..., t_\ell\}$ is the set of types. We will use t to refer to some generic type in T.
- τ is a matrix consisting of each candidate c's type vector τ_c where
  - τ_c is a type row vector of candidate c consisting of 1's and 0's
  - $\tau_c^t = 1$ if c belongs to type t and $\tau_c^t = 0$ otherwise.
- $\underline{q}$ is a vector consisting of all type-specific lower bounds. The value $\underline{q^t}$ denotes the lower bound for type t.[3]

We will denote the set of all types that a candidate c belongs to by $\eta(c)$. For $c, d \in C$, if $c \succsim d$ but $d \not\succsim c$, we will write the strict part of the relation as $c \succ d$. Note that the model is powerful enough to capture the following kind of lower bounds: "there should be at least x members who are of one of the types from set $S \subset T$." In that case, one can create an 'artificial' type t_S such that for any $c \in C$, $\tau_c^{t_S} = 1$ if $\tau_c^t = 1$ for some $t \in S$.

For a committee $W \subset C$, we will denote $\sum_{c \in W} \tau_c$ by τ_W. We will denote the number of candidates of type t in W by $\tau_W(t)$. For some committee $W \subset C$, if $\tau_W(t) < \underline{q^t}$, we will say that type t is *under-represented* in W.

The linear ranking over C could be based on some objective measure that reflects the global quality of the candidate such as entrance examination scores. It could also be based on the aggregate scores based on some positional scoring voting done by voters who vote on the candidates (Brams and Potthoff, 1990; Bredereck et al., 2018).

The goal in the committee selection problem is to select a target number of candidates. We will denote the target size by k.

Example 1 Consider the following instance.

- $C = \{c_1, c_2, c_3, c_4\}$
- $c_1 \succ c_2 \succ c_3 \succ c_4$
- $T = \{t_1, t_2, t_3, t_4, t_5\}$
- $\tau = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$
- $\underline{q} = \begin{pmatrix} 0 & 1 & 2 & 1 & 0 \end{pmatrix}$

Suppose the target committee size is two. There is no feasible committee that can satisfy the hard constraints. The committee $\{c_3, c_4\}$ satisfies all the constraints except the one corresponding to t_4.

[3] Note that $\underline{q^t}$ is not a vector. The underscore denotes that it is a lower bound.

3 Axioms for Diverse Committee Selection

We formalize some axiomatic properties that are desirable in our context. A committee W satisfies type distribution $(x_1, \dots, x_\ell)$ if for each i, it has at least x_i members of type t_i.

Definition 1 (domination between type distributions) A type distribution $x = (x_1, \dots, x_\ell)$ weakly dominates another type distribution $y = (y_1, \dots, y_\ell)$ if

(i) for each t_i such that $y_i \geq \underline{q}^i$, we also have $x_i \geq \underline{q}^i$, and

(ii) for each t_i such that $y_i < \underline{q}^i$, either $x_i \geq \underline{q}^i$ or $|x_i - \underline{q}^i| \leq |y_i - \underline{q}^i|$

When x weakly dominates y, we denote it by $x \geq y$. We say x *dominates* y if $x \geq y$ but $y \ngeq x$.

Next, we observe the following lemma regarding the transitivity of the weak domination relation between type distributions.

Lemma 1 *If $\tau_X \geq \tau_Y$ and $\tau_Y \geq \tau_Z$, then $\tau_X \geq \tau_Z$.*

Proof. By the definition of weak domination, if some type t_i is not under-represented in Z, then it also not under-represented in Y. If t_i is under-represented in Z, then its representation is at least as much in Y. By the same reasoning, if some type is not under-represented in Y, then it also not under-represented in X. If t_i is under-represented in Y, then its representation is at least as much in X. Hence we obtain the following implications. (i) if some type t_i is not under-represented in Z, then it also not under-represented in X, and (ii) if t_i is under-represented in Z, then its representation is at least as much in X. By the definition of domination between types, we note that $\tau_X \geq \tau_Z$. □

Based on the notion of dominance between type distributions, we are now in a position to define type optimality of a committee. Roughly speaking, a committee W is type optimal if there exists no other committee obtained by swapping two candidates whose type distribution dominates that of W.

Definition 2 (Type optimal) A committee W is type optimal if there exists no candidate $c' \in W$ and $c \notin W$ such that $\tau_{((W \setminus \{c'\}) \cup \{c\})}$ dominates τ_W.

We note that type optimality is desirable in terms of distributional constraints but does not take into account the excellence of the candidates. Type optimality has been defined in a local sense based on swaps of candidates. If we define it in a global sense by allowing swaps of subsets of candidates with subsets of candidates, then checking whether a given type distribution is optimal is NP-complete.

We now present an axiom that avoids scenarios where a candidate may feel that she deserves the place of a lesser ranked candidate. The axiom is adapted from the literature on stable matching with distributional constraints (Kurata et al., 2017; Goto et al., 2017; Kojima et al., 2014; Ehlers et al., 2014; Kamada and Kojima, 2015). The intuition behind the axiom is that a candidate $c \notin W$ has justified envy towards $c' \in W$ if c has higher priority than c', and replacing c' with c will not make some type t_i reaching the desired quota to being under-represented, or from being under-represented to being even more under-represented.

Definition 3 (Justified envy-freeness) A committee W satisfies justified envy-freeness if there are no candidates $c \notin W$ and $c' \in W$ such that $c \succ c'$ and there exists no type $t_i \in \eta(c') \setminus \eta(c)$ such that the number of candidates in W of type t_i is less than or equal to $\underline{q}^i$.

We note that justified envy-freeness by itself can be trivially satisfied by a committee that selects the top k ranked candidates. Such a committee is *score-optimal*, i.e., maximizes the total score, if there were points associated with ranks of ordinal ranks in the priority list. However, such a committee may not respect any of the distributional constraints.

We have identified justified envy-freeness and type optimality as two desirable axioms for our setting. The two axioms are necessarily satisfied by any committee that is score-optimal and meets the hard distributional constraints. If a committee is not type optimal, then it does not satisfy the distributional constraints. If it does not satisfy justified envy-freeness, then a swap of two candidates can increase the total points of the committee without violating its hard constraints which means that the committee was not score-optimal subject to the constraints. In the next section, we present an algorithm that returns a committee satisfying both axioms.

4 A Rule for Diverse Committee Selection

We are in a position to present our algorithm (Algorithm 1) to find a diverse committee. In the first stage (Steps 1 to 7), the algorithm checks if there is a type that is under-represented and then selects the highest priority candidate who satisfies such a type. The first stage is along similar lines as the Greedy Algorithm I of Bredereck et al. (2018). If no type is under-represented, then the algorithm adds the required number of highest ranking candidates. The second stage (Steps 8 to 10) is geared towards obtaining a good type distribution. If the committee does not satisfy type optimality, candidates are exchanged with the goal to satisfy type optimality until it is satisfied. In the final while loop (Steps 11 to 13), the algorithm allows swaps of candidates if there is justified envy. The algorithm stops when the committee satisfies justified envy-freeness. In certain steps of the algorithm, there may be multiple options for selecting or choosing candidates to be included or excluded from the committee. In this sense the algorithm represents a class of rules rather than a single rule. The exact choice can be established by choosing any tie-breaking rule. We have presented the algorithm without specifying the tie-breaking rule because the axiomatic properties satisfied by the algorithm are not dependent on tie-breaking.

Let us illustrate the algorithm.

Example 2 Consider the following instance.

- $C = \{c_1, c_2, c_3, c_4\}$
- $c_1 \succ c_2 \succ c_3 \succ c_4$
- $T = \{t_1, t_2, t_3, t_4, t_5\}$

Algorithm 1 Rule for finding a desirable committee satisfying soft distributional constraints on types

Input: $(C, \succsim, T, \tau, \underline{q})$.
Output: $W \subseteq C$ such that $|W| = k$

1: $W \longleftarrow \emptyset$
2: **while** $|W| < k$ and there exists some candidate in $C \setminus W$ of some type t that is underrepresented in W **do**
3: Add a highest priority candidate of that type t to W.
4: **end while**
5: **if** $|W| < k$ **then**
6: Select highest ranked $k - |W|$ candidates from $C \setminus W$ into W.
7: **end if**
8: **while** there is a candidate $c \notin W$ and candidate $c' \in W$ such that $\tau_{(W\setminus\{c'\})\cup\{c\}}$ dominates τ_W **do**
9: $W \longleftarrow (W \setminus \{c'\}) \cup \{c\}$
10: **end while**
11: **while** there is a candidate $c \notin W$ who has justified envy for some candidate $c' \in W$ **do**
12: $W \longleftarrow (W \setminus \{c'\}) \cup \{c\}$
13: **end while**
14: **return** W

– $\tau = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$
– $\underline{q} = \begin{pmatrix} 0 & 1 & 2 & 1 & 0 \end{pmatrix}$

Suppose the target committee size is 2. The algorithm first considers a type t_2 that is under-represented and selects c_2 because $c_2 > c_4$. It then considers type t_3 that is under-represented and selects c_3. At this point the type distribution is not optimal and can be improved by the exchange of c_2 with c_4 so that $W = \{c_3, c_4\}$. At this point the committee is type optimal and also satisfies justified envy-freeness.

Proposition 1 *Algorithm 1 returns in time $O(|C|^3)$ a committee that satisfies justified envy-freeness.*

Proof. Note that by Step 7, we already have a committee of size k. The first while loop takes time at most $O(k|C|)$. In the second while loop, the type distribution keeps improving since the type domination relation is transitive (Lemma 1), but there can be at most $|C|^2$ such improvements. The while loop requires the scan of at most $|C|$ candidates so the time taken for the second while loop is at $O(|C|^3)$. Finally, in the last while loop, with each swap of candidates, a candidate is replaced by a candidate with a higher ranking. This can happen at most $O(|C|^2)$ times. The while loop requires the scan of at most $|C|$ candidates so the time taken for the last while loop is at $O(|C|^3)$.

The while loop terminates when no candidate has justified envy for some candidate in W. □

Although Algorithm 1 finds a type optimal committee by Step 10, the committee undergoes further changes in the final while loop to achieve justified envy-freeness. We now argue that the returned committee still satisfies type optimality.

Proposition 2 *Algorithm 1 returns a committee satisfying type optimality.*

Proof. In the second while loop, we start from a committee of size k. Due to the argument in the proof of Proposition 1, we know that by the end of Step 10, W is a committee that is type optimal.

We show that in the final while loop, W remains type optimal. In the final while loop, we implement swaps if there is a candidate $c \notin W$ who has justified envy for some candidate $c' \in W$. Suppose we swap c' with c. We claim that $(W \setminus \{c'\}) \cup \{c\}$ is also type optimal. Since c had justified envy against c', by definition of justified envy, there exists no type $t_i \in \eta(c') \setminus \eta(c)$ such that the number of candidates in W of type t_i is less than or equal to $\underline{q}^i$. If there is a type t_i such that $\tau_W(t_i) \geq \underline{q}^i$, then $\tau_{(W\setminus\{c'\})\cup\{c\})}(t_i) \geq \underline{q}^i$. In words, if a type is not under-represented in W, it is not under-represented in $(W \setminus \{c'\}) \cup \{c\}$. If there is a type t_i such that $\tau_W(t_i) < \underline{q}^i$, then $\tau_W(t_i) \leq \tau_{(W\setminus\{c'\})\cup\{c\})}(t_i) \leq \underline{q}^i$. In words, if a type is under-represented in W, it is at most as under-represented in $(W \setminus \{c'\}) \cup \{c\}$. Thus we have established that $\tau_{(W\setminus\{c'\})\cup\{c\})} \geq \tau_W$. Since W was type optimal, $(W \setminus \{c'\}) \cup \{c\}$ is type optimal as well. □

We have shown that our algorithm simultaneously satisfies both type optimality and justified envy-freeness. Since these are two basic properties satisfied by optimal committees satisfying hard constraints, our algorithm provides a computationally easy and principled rule to find desirable committees that 'almost' satisfy the distributional constraints. It will be interesting to see if other desirable axioms can be simultaneously satisfied in conjunction with the ones which our algorithm satisfies. Our cubic-time algorithm may not return a committee satisfying the distributional constraints even if such a committee exists. However this may not be viewed as criticism of the algorithm since the problem of checking whether such a committee exists is NP-complete, and a polynomial-time algorithm is unlikely to exist unless P-NP (Papadimitriou, 1994).

Our general algorithm can have more precise specifications that prioritise certain types in a lexicographical manner or implement swaps according to some predetermined pattern. The $\succsim$ ranking order can be derived by using some social welfare function for a set of voters voting on the quality of the candidates. Finally, it will be interesting to undertake a more experimental comparison of our rule with other methods proposed in the literature.

References

H. Aziz, J. Lang, and J. Monnot. Computing Pareto Optimal Committees. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 60–66, 2016.

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

H. Aziz, E. Elkind, P. Faliszewski, M. Lackner, and P. Skowron:. The Condorcet principle for multiwinner elections: From shortlisting to proportionality. In *proc 26th ijcai*, pages 84–90, 2017b.

S. J. Brams and R. F. Potthoff. Constrained approval voting: A voting system to elect a governing board. *Interfaces*, 20(2):67–80, 1990.

R. Bredereck, P. Faliszewski, A. Igarashi, M. Lackner, and P. Skowron. Multiwinner elections with diversity constraints. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2018.

L. E. Celis, L. Huang, and N. K. Vishnoi. Group fairness in multiwinner voting. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 144–151, 2018.

L. Ehlers, I. E. Hafalir, M. B. Yenmezb, and M. A. Yildirimc. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153:648—683, 2014.

P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon. Multiwinner voting: A new challenge for social choice theory. In U. Endriss, editor, *Trends in Computational Social Choice*, chapter 2. 2017.

M. Goto, F. Kojima, R. Kurata, A. Tamura, and M. Yokoo. Designing matching mechanisms under general distributional constraints. *American Economic Journal: Microeconomics*, 9(2):226–62, 2017.

R. Izsak, N. Talmon, and G. Woeginger. Committee selection with intraclass and interclass synergies. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2018.

Y. Kamada and F. Kojima. Efficient matching under distributional constraints: Theory and applications. *The American Economic Review*, 105(1):67–99, 2015.

F. Kojima, A. Tamura, and M. Yokoo. Designing matching mechanisms under constraints: An approach from discrete convex analysis. 2014.

C-C. Kuo, F. Glover, and K. S. Dhir. Analyzing and modeling the maximum diversity problem by zero-one programming. *Decision Sciences*, 24(6):1171–1185, 1993.

R. Kurata, N. Hamada, A. Iwasaki, and M. Yokoo. Controlled school choice with soft bounds and overlapping types. *Journal of Artificial Intelligence Research*, 58: 153–184, 2017.

J. Lang and P. Skowron. Multi-attribute proportional representation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 530–536. AAAI Conference on Artificial Intelligence (AAAI), 2016.

C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

R. Potthoff. Use of linear programming for constrained approval voting. *Interfaces*, 20(5):79–80, 1990.

T. C. Ratliff. Selecting committees. *Public Choice*, 126(3/4):343–355, 2006.

A. Straszak, M. Libura, J. Sikorski, and D. Wagner. Computer-assisted constrained approval voting. *Group Decision and Negotiation*, 2(4):375–385, 1993.