Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Special Track on AI for Good

DiRe Committee : Diversity and Representation Constraints in Multiwinner Elections

Kunal Relia*
New York University
krelia@nyu.edu

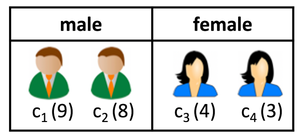
Abstract

The study of fairness in multiwinner elections focuses on settings where candidates have attributes. However, voters may also be divided into predefined populations under one or more attributes. The models that focus on candidate attributes alone may systematically under-represent smaller voter populations. Hence, we develop a model, DiRe Committee Winner Determination (DRCWD), which delineates candidate and voter attributes to select a committee by specifying diversity and representation constraints and a voting rule. We analyze its computational complexity and develop a heuristic algorithm, which finds the winning DiRe committee in under two minutes on 63% of the instances of synthetic datasets and on 100% of instances of real-world datasets. We also present an empirical analysis of feasibility and utility traded-off.

Moreover, even when the attributes of candidates and voters coincide, it is important to treat them separately as diversity does not imply representation and vice versa. This is to say that having a female candidate on the committee, for example, is different from having a candidate on the committee who is preferred by the female voters, and who themselves may or may not be female.

1 Introduction

The problem of selecting a committee from a given set of candidates arises in multiple domains; it ranges from political sciences (e.g., selecting the parliament of a country) to recommendation systems (e.g., selecting the movies to show on Netflix). Formally, given a set C of m candidates (politicians and movies, respectively), a set V of n voters (citizens and Netflix subscribers, respectively) give their ordered preferences over all candidates $c \in C$ to select a committee of size k. These preferences can be stated directly in case of parliamentary elections, or they can be derived based on input, such as when Netflix subscribers' viewing behavior is used to derive their preferences. In this paper[1], we focus on selecting a committee using direct, ordered, and complete preferences.



(a) **candidates**

Voter	State	Preferences
v_1	California	$c_1 \succ c_2 \succ c_3 \succ c_4$
v_2	California	$c_1 \succ c_2 \succ c_3 \succ c_4$
v_3	California	$c_1 \succ c_2 \succ c_3 \succ c_4$
v_4	Illinois	$c_4 \succ c_2 \succ c_3 \succ c_1$

(b) **voters**

Figure 1: (a) Candidates with "gender" attribute (and their Borda scores) and (b) voters with "state" attribute.

Which committee is selected depends on the committee selection rule, also called the multiwinner voting rule. Examples of commonly used families of rules when a complete ballot of each voter is given are Condorcet principle-based rules [Faliszewski *et al.*, 2016], approval-based voting rules [Faliszewski *et al.*, 2016; Sánchez-Fernández *et al.*, 2017], and ordinal preference ballot-based voting rules like k-Borda and β-Chamberlin-Courant (β-CC) [Elkind *et al.*, 2017; Faliszewski *et al.*, 2017]. We refer readers to Section 2.2 of [Faliszewski *et al.*, 2017] for further details. In this paper, we focus on ordinal preference-based rules.

However, these rules can create or propagate biases by systematically harming candidates coming from historically disadvantaged groups [Bredereck *et al.*, 2018; Celis *et al.*, 2018]. Hence, diversity constraints on candidate attributes were introduced to mitigate this. However, voters may be divided into predefined populations under one or more attributes, which may be different from candidate attributes. For example, voters in Figure 1b are divided into "California" and "Illinois" populations under the "state" attribute. The models that focus on candidate attributes alone may systematically under-represent smaller voter populations.

Example 1. *An election E consists of m = 4 candidates (Figure 1a) and n = 4 voters giving complete, ordered preference (Figure 1b) to select a committee of size k = 2. Candidates and voters have one attribute each (gender and state, respectively). The k-Borda[2] winning committee for each voter population is $\{c_1, c_2\}$ for California and $\{c_4, c_2\}$ for Illinois.*

*This work was supported, in part, by Julia Stoyanovich's NSF grants No. 1916647, 1934464, and 1916505.

[1] An extended version of this paper is available at [Relia, 2021].

[2] The Borda rule associates score $m - i$ with the i^{th} position, and k-Borda selects candidates with the k highest Borda scores.

Suppose that we impose a diversity constraint *that requires the committee to have at least one candidate of each gender and a* representation constraint *that requires at least one candidate from the winning committee of each state. Observe that the highest-scoring committee consists of* $\{c_1, c_2\}$ *(score = 17), which is representative but not diverse, since both candidates are male. Further, the highest-scoring diverse committee consisting of* $\{c_1, c_3\}$ *(score = 13) is not representative because it does not include any winning candidates from Illinois, the smaller state. The highest-scoring diverse* ***and*** *representative committee is* $\{c_2, c_3\}$ *(score = 12).*

This example illustrates the inevitable utility cost due to enforcing additional constraints.

In contrast to prior work in computational social choice, we incorporate voter attributes that are separate from candidate attributes. Also, our work is different from "proportional representation" [Sánchez-Fernández *et al.*, 2017; Brill *et al.*, 2018; Monroe, 1995], where the number of candidates selected in the committee from each group is proportional to the number of voters preferring that group, and from its variants such as "fair" representation [Koriyama *et al.*, 2013]. All these approaches dynamically divide the voters based on the cohesiveness of their preferences. Another related work, multi-attribute proportional representation [Lang and Skowron, 2018], couples candidate and voter attributes. An important observation we make here is that, even if the attributes of the candidates and of the voters *coincide*, it may still be important to treat them *separately*. This is because having a female candidate on the committee, for example, is different from having a candidate on the committee who is preferred by the female voters, and who themselves may or may not be female. This observation has implications on United Nations' sustainable development goals on equality.

Contributions. In this paper, we define a model that treats candidate and voter attributes separately during committee selection, and thus enables selection of the highest-scoring diverse *and* representative committee. We classify the computational complexity of using our model, study the limits of approximability, and develop a heuristic algorithm. Finally, we present an experimental evaluation using real and synthetic datasets, in which we show the efficiency of our algorithm, analyze the feasibility of committee selection and illustrate the utility trade-offs.

2 Related Work

Our work is motivated by the work on constrained multiwinner elections. Goalbase score functions, which specify an arbitrary set of logic constraints and let the score capture the number of constraints satisfied, could be used to ensure diversity [Uckelman *et al.*, 2009]. Using diversity constraints over multiple attributes in single-winner elections is NP-hard [Lang and Skowron, 2018]. Also, using diversity constraints over multiple attributes in multiwinner elections and participatory budgeting is NP-hard, which has led to approximation algorithms and matching hardness of approximation results by Bredereck *et al.* [2018] and Celis *et al.* [2018]. Finally, due to the hardness of using diversity constraints over multiple attributes in approval-based multiwinner elections [Brams, 1990], these have been formalized as integer linear programs (ILP) [Potthoff, 1990]. In contrast, Skowron *et al.* [2015] showed that ILP fails in the real world when using ranking-based proportional representation rules.

Overall, the work by Bredereck *et al.* [2018], Celis *et al.* [2018], and Lang and Skowron [2018] is closest to ours but we differ as we: (i) consider elections with predefined voter populations under one or more attributes, (ii) delineate voter and candidate attributes even when they coincide, and (iii) consider diversity *and* representation constraints.

3 Preliminaries and Notation

Multiwinner Elections. Let $E = (C, V)$ be an election consisting of a candidate set $C = \{c_1, \ldots, c_m\}$ and a voter set $V = \{v_1, \ldots, v_n\}$, where each voter $v \in V$ has a preference (ranked) list $\succ_v$ over m candidates. $\mathrm{pos}_v(c)$ denotes the position of candidate $c \in C$ in the ranking of voter $v \in V$, where the most preferred candidate has position 1 and the least preferred has position m.

Given $E = (C, V)$ and $k \in [m]$ (for $k \in \mathbb{N}^+$, $[k] = \{1, \ldots, k\}$), a multiwinner election selects a k-sized subset of candidates (or committee) W using a multiwinner voting rule $\mathtt{f}$ (discussed later) such that the score of the committee $\mathtt{f}(W)$ is the highest. Ties are broken using a pre-decided order.

Candidate Groups. The candidates have μ attributes, $A_1, \ldots, A_\mu : \mu \in \mathbb{Z}$ and $\mu \geq 0$. $\forall\ i \in [\mu]$, A_i partitions the candidates into $g_i \in [m]$ groups, $A_{(i,1)}, \ldots, A_{(i,g_i)} \subseteq C$. For example, candidates in Figure 1a have gender attribute (μ = 1) with disjoint groups, male and female (g_1 = 2). Overall, the set $\mathcal{G}$ of *all* such arbitrary and potentially non-disjoint groups is $A_{(1,1)}, \ldots, A_{(\mu,g_\mu)} \subseteq C$.

Voter Populations. The voters have π attributes, $A'_1, \ldots, A'_\pi : \pi \in \mathbb{Z}$ and $\pi \geq 0$, which may be different from the candidate attributes. $\forall\ i \in [\pi]$, A'_i partitions the voters into $p_i \in [n]$ populations, $P_{(i,1)}, \ldots, P_{(i,p_i)} \subseteq V$. For example, voters in Figure 1b have state attribute (π = 1), which has populations California and Illinois (p_1 = 2). Overall, the set $\mathcal{P}$ of *all* such predefined and potentially non-disjoint populations will be $P_{(1,1)}, \ldots, P_{(\pi,p_\pi)} \subseteq V$.

Additionally, we are given W_P, the winning committee $\forall$ $P \in \mathcal{P}$. We limit the scope of W_P to be a committee instead of a ranking of k candidates because using a committee selection rule such as CC rule does not return a complete ranking.

Committee Selection Rules. In this paper, we focus on committee selection rules $\mathtt{f}$ that are based on single-winner positional voting rules and are monotone and submodular ($\forall A \subseteq B, f(A) \leq f(B)$ and $f(B) \leq f(A) + f(B \setminus A)$).

Definition 1. ***Chamberlin–Courant (CC) rule [1983]:*** *The CC rule associates each voter with their most preferred candidate in the committee. The score of a committee is the sum of scores given by voters to their associated candidate.* β*-CC uses the Borda rule to assign scores to each voter's candidate.*

Definition 2. ***Monroe rule:*** *The Monroe rule [1995] dynamically divides the* n *voters into* π *populations based on the cohesiveness of their preferences where* $\pi = k$ *(assuming* k *divides* n*). Then, each subpopulation's most preferred candidate is selected into the* k*-sized committee. Formally,* $\forall$ $P \in \mathcal{P}$*, select* c *for* P *:* $\max_{c \in C}(\mathtt{f}_P(c))$.

A special case of submodular functions are separable functions: score of a committee W is the sum of the scores of individual candidates in the committee. Formally, $\mathtt{f}(W) = \sum_{c \in W} \mathtt{f}(c)$ [Bredereck *et al.*, 2018]. Monotone and separable rules are natural and are considered good when the goal is to shortlist a set of individually excellent candidates:

Definition 3. ***k-Borda rule*** *The k-Borda rule outputs committee of k candidates with the k highest Borda scores.*

4 DiRe Committee Model

In this section, we formally define a model to select a diverse *and* representative committee, namely the $DiRe$ committee, and compare our model to existing related models.

Definition 4. ***Unconstrained Committee Winner Determination (UCWD):*** *We are given a set C of m candidates, a set V of n voters such that each voter $v \in V$ has a preference list $\succ_v$ over m candidates, a committee selection rule* $\mathtt{f}$*, and a committee size $k \in [m]$. Let $\mathcal{W}$ denote the family of all size-k committees. The goal of UCWD is to select a committee $W \in \mathcal{W}$ that maximizes $\mathtt{f}(W)$.*

We introduce the diversity and representation constraints such that their minimum value is 1 as each candidate group and voter population deserve at least one candidate in the committee. This replicates real-world settings like the United Nations (UN) charter, which guarantees at least one representative to each member country in the UN General Assembly, independent of the country's population. Theoretically, all results in this paper hold even if the minimum value is 0.

Diversity Constraints. These constraints, denoted by $\ell^D_G \in [\min(k, |G|)]$ for each $G \in \mathcal{G}$, enforce at least ℓ^D_G candidates from the group G to be in the committee W. Formally, $\forall$ $G \in \mathcal{G}$, $|G \cap W| \geq \ell^D_G$.

Representation Constraints. These constraints, denoted by $\ell^R_P \in [1, k]$ for each voter population $P \in \mathcal{P}$, enforce at least ℓ^R_P candidates from the population P's committee W_P to be in the committee W. Formally, $\forall\, P \in \mathcal{P}$, $|W_P \cap W| \geq \ell^R_P$.

Definition 5. ***(μ, π)-DiRe Committee Feasibility ((μ, π)-DRCF):*** *We are given an instance of $E = (C, V)$, $k \in [m]$, $\mathcal{G}$ under μ attributes and $\forall\ G \in \mathcal{G}$, ℓ^D_G, and $\mathcal{P}$ under π attributes and $\forall\ P \in \mathcal{P}$, ℓ^R_P and W_P. Let $\mathcal{W}$ denote the family of all size-k committees. The (μ, π)-DRCF problem selects committees $W \in \mathcal{W}$ such that $|G \cap W| \geq \ell^D_G \; \forall\ G \in \mathcal{G}$ and $|W_P \cap W| \geq \ell^R_P \; \forall\ P \in \mathcal{P}$. These are called DiRe committees.*

If a committee selection rule $\mathtt{f}$ is also an input to the feasibility problem, we get the (μ, π, $\mathtt{f}$)-DRCWD problem:

Definition 6. ***(μ, π, f)-DiRe Committee Winner Determination ((μ, π, f)-DRCWD):*** *Given (μ, π)-DRCF and* $\mathtt{f}$*, the (μ, π,* $\mathtt{f}$*)-DRCWD selects the highest scoring $DiRe$ committee.*

Observation 1. *(μ, π,* $\mathtt{f}$*)-DRCWD is a generalized version of (μ, π)-DRCF and UCWD.*

Finally, as our model provides the flexibility to specify the diversity and representation constraints and to select the voting rule, the diverse committee problem [Bredereck *et al.*, 2018; Celis *et al.*, 2018] and the apportionment problem [Brill *et al.*, 2018; Hodge and Klima, 2018] are special cases of (μ, π, $\mathtt{f}$)-DRCWD. See [Relia, 2021] for formal formulations.

Instance of (μ, π, f)-DRCWD	Complexity
(≤ 2, 0, separable)-DRCWD	P (Lem. 1)
(≥ 3, 0, separable)-DRCWD	NP-hard (Thm. 3, Thm. 4)
(≥ 0, ≥ 1, separable)-DRCWD	NP-hard (Thm. 5, Cor. 2)
(≥ 0, ≥ 0, submodular)-DRCWD	NP-hard (Thm. 6, Cor. 3)

Table 1: The complexity of (μ, π, $\mathtt{f}$)-DRCWD (Thm. 1, Cor. 1). The value in brackets for μ and π denote that the results hold for all non-negative integers μ and π that satisfy the condition stated in the brackets. 'Lem.': Lemma. 'Thm.': Theorem. 'Cor.': Corollary.

5 Complexity Results

Selecting a committee using a submodular function such as the CC rule is NP-hard, known via reduction from exact 3-cover [Procaccia *et al.*, 2008]. Selecting a diverse committee when a candidate belongs to three groups is also NP-hard, known via reductions from 3-dimensional matching [Bredereck *et al.*, 2018] and 3-hypergraph matching [Celis *et al.*, 2018]. However, these hardness results are fragmented over several papers and they use reductions from several well-known NP-hard problems. Next, we introduce representation constraints and hardness due to its use is unknown. Hence, we provide a complete computational classification[3] of (μ, π, $\mathtt{f}$)-DRCWD using reductions from *single* well-known NP-hard problem, the vertex cover problem, in line with the approach used in [Chakraborty and Kolaitis, 2021]. All proofs can be found in the extended version of this paper [Relia, 2021].

Theorem 1. *Let μ, $\pi \in \mathbb{Z}$: μ, $\pi \geq 0$ and* $\mathtt{f}$ *be a monotone committee selection rule, then (μ, π,* $\mathtt{f}$*)-DRCWD is NP-hard.*

Corollary 1. ***Classification of Complexity.***

1. *If $\forall \mu \in \mathbb{Z} : \mu \geq 0$, $\forall \pi \in \mathbb{Z} : \pi \geq 0$, and* $\mathtt{f}$ *is a* ***submodular*** *function, then (μ, π,* $\mathtt{f}$*)-DRCWD is NP-hard.*
2. *If $\forall \mu \in [0, 2]$, $\pi = 0$, and* $\mathtt{f}$ *is a* ***separable*** *function, then (μ, π,* $\mathtt{f}$*)-DRCWD is in P.*
3. *If $\forall \mu \in \mathbb{Z} : \mu \geq 3$, $\pi = 0$, and* $\mathtt{f}$ *is a* ***separable*** *function, then (μ, π,* $\mathtt{f}$*)-DRCWD is NP-hard.*
4. *If $\forall \mu \in \mathbb{Z} : \mu \geq 0$, $\forall \pi \in \mathbb{Z} : \pi \geq 1$, and* $\mathtt{f}$ *is a* ***separable*** *function, then (μ, π,* $\mathtt{f}$*)-DRCWD is NP-hard.*

5.1 Tractable Case

Theorem 2. *[Theorem 21, Corollary 22 in full-version of Celis* et al. *[2018]] The diverse committee feasibility problem can be solved in polynomial time when μ = 2.*

Without loss of generality (W.l.o.g.), the above theorem holds when $\mu = 2$. Hence, it holds for $\mu \in [0, 2]$. Therefore, the following lemma and Corollary 1(2) follows:

Lemma 1. *If $\mu \in [0, 2]$, $\pi = 0$, and* $\mathtt{f}$ *is a monotone,* ***separable*** *function, then (μ, π,* $\mathtt{f}$*)-DRCWD is in P.*

5.2 Hardness Results

NP-hard problem used. We reduce from the vertex cover problem on 3-regular, 2-uniform[4] hypergraphs [Garey and Johnson, 1979; Alimonti and Kann, 1997].

[3] All hardness results are under the assumption P $\neq$ NP.

[4] The size of hyperedges has implications on the inapproximability results. We use 2-uniform hypergraphs for the complexity results.

Definition 7. ***Vertex Cover (VC) problem:*** *Given a graph H consisting of a set of m vertices $X = \{x_1, x_2, \ldots, x_m\}$ and a set of n edges $E = \{e_1, e_2, \ldots, e_n\}$ where each $e \in E$ connects two vertices in X, then VC $S \subseteq H$ is where $\forall\, e \in E$, $e \cap S \neq \phi$. The VC problem is to find the minimum size S.*

(μ, π, f)-DRCWD w.r.t. Diversity Constraints
When $\pi = 0$, (μ, π, f)-DRCWD is related to the diverse committee selection problem. However, the hardness of (μ, π, f)-DRCWD when $\mu \geq 3$ does not follow the hardness of the diverse committee selection problem.

More specifically, Theorem 9 of Bredereck *et al.* [2018] uses a reduction from 3-Dimensional Matching that only holds for instances when the number of groups that a candidate can belong to is *exactly* 3. Also, lower bound *and* upper bound is set to 1, which is mathematically different from our setting where we only allow lower bounds. On the other hand, Theorem 6 ("NP-hardness of feasibility: $\Delta \geq 3$"[5]) of Celis *et al.* [2018] uses two reductions: the first reduction from Δ-hypergraph matching is indeed for the case when $\Delta \geq 3$ but is limited to instances when the lower bound is 0 and the upper bound is 1, which is a trivial case in our setting. Moreover, the reduction from Δ-hypergraph matching uses a different problem as when $\Delta \neq \Delta'$, the Δ-hypergraph matching and Δ'-hypergraph matching are separate problems. The second reduction from 3-regular VC is for instances when $\Delta = 3$.

Hence, in this section, we give reductions such that our result holds $\forall \mu \in \mathbb{Z} : \mu \geq 3$, even when $\forall G \in \mathcal{G}$, $\ell^D_G = 1$. The next two theorems help us prove Corollary 1(3).

Theorem 3. *If $\forall \mu \in \mathbb{Z} : \mu \geq 3$ and μ is an odd number, $\pi = 0$, and f is a monotone, separable function, then (μ, π, f)-DRCWD is NP-hard, even when $\forall G \in \mathcal{G}$, $\ell^D_G = 1$.*

Proof Sketch. We create an instance such that $|C| = |V| = m + (m \cdot (2\mu^2 - 7\mu + 3))$ and $|W| = k + m\mu^2 - 3m\mu$. The reduction conforms to the real-world stipulations: (i) each candidate attribute $A_i, \forall i \in [\mu]$, *partitions* all the candidates into two or more groups and (ii) either no two attributes partition the candidates in the same way or if they do, the lower bounds for the two attributes are not the same. If any two attributes A_i and $A_j : i \neq j$ violates stipulation (ii), it implies that A_i and A_j are mathematically equivalent. □

Theorem 4. *If $\forall \mu \in \mathbb{Z} : \mu \geq 3$ and μ is an even number, $\pi = 0$, and f is a monotone, separable function, then (μ, π, f)-DRCWD is NP-hard, even when $\forall G \in \mathcal{G}$, $\ell^D_G = 1$.*

(μ, π, f)-DRCWD w.r.t. Representation Constraints
We now study the computational complexity of (μ, π, f)-DRCWD due to the presence of voter attributes. The following theorem helps us prove Corollary 1(4).

Theorem 5. *If $\mu = 0$, $\forall \pi \in \mathbb{Z} : \pi \geq 1$, and f is a monotone, separable function, then (μ, π, f)-DRCWD is NP-hard, even when $\forall P \in \mathcal{P}$, $\ell^R_P = 1$.*

Proof Sketch. Create an instance where $|C| = m + (n \cdot m)$, $|V| = n^2$, and $|W| = k$. The reduction conforms to the stipulations analogous to the ones for candidate attributes. □

[5] In Celis *et al.* [2018], Δ denotes "the maximum number of groups in which any candidate can be".

Instance of (μ, π)-DRCF	Limits of Approximability
(≥ 3, 0)-DRCF	$(1-\varepsilon)\cdot$ $(\ln \mu - \mathcal{O}(\ln\ln\mu))$ (Thm. 7)
(0, ≥ 1)-DRCF	$k - \varepsilon$ (Thm. 9)
(≥ 1, ≥ 1)-DRCF	$(1-\varepsilon)\cdot$ $\ln(\lvert\mathcal{G}\rvert + \lvert\mathcal{P}\rvert)$ (Thm. 8)

Table 2: Limits of approximability of (μ, π)-DRCF. The value in brackets of μ and π denote that results hold for non-negative integers μ and π that satisfy the condition. ε denotes an arbitrarily small constant such that the results are meant to hold for every such $\varepsilon > 0$.

Corollary 2. *If $\forall \mu \in \mathbb{Z} : \mu \geq 0$, $\forall \pi \in \mathbb{Z} : \pi \geq 1$, and f is a monotone, separable function, then (μ, π, f)-DRCWD is NP-hard, even when $\forall G \in \mathcal{G}$, $\ell^D_G = 1$ and $\forall P \in \mathcal{P}$, $\ell^R_P = 1$.*

(μ, π, f)-DRCWD w.r.t. Submodular Scoring Function
Chamberlin-Courant (CC) rule is a well-known monotone, submodular scoring function [Celis *et al.*, 2018], which we use for our proof. The novelty of our reduction is that it holds for determining the winning committee using CC rule that uses *any* positional scoring rule with scoring vector s $= \{s_1, \ldots, s_m\}$ such that $s_1 = s_2$, $s_m \geq 0$, and $\forall i \in [3, m-1], s_i \in \mathbb{Z} : s_i \geq s_{i+1}$ and $s_2 > s_i$.

The following statements prove Corollary 1(1).

Theorem 6. *If f is a monotone,* ***submodular*** *function, then (μ, π, f)-DRCWD is NP-hard even when $\mu = 0$ and $\pi = 0$.*

Corollary 3. *If $\forall \mu \in \mathbb{Z} : \mu \geq 0$, $\forall \pi \in \mathbb{Z} : \pi \geq 0$, and f is a monotone,* ***submodular*** *function, then (μ, π, f)-DRCWD is NP-hard, even when $\forall G \in \mathcal{G}$, $\ell^D_G = 1$ and $\forall P \in \mathcal{P}$, $\ell^R_P = 1$.*

5.3 Inapproximability

The hardness of (μ, π, f)-DRCWD is mainly due to the hardness of (μ, π)-DRCF, even when constraints are set to 1. Hence, we now study the inapproximability of (μ, π)-DRCF.

We can try to reformulate representation constraints as diversity constraints. However, this is not possible as each candidate attribute partitions *all* m candidates into groups and $\ell^D_G \in [\min(k, |G|)]$, $\forall\, G \in \mathcal{G}$. However, for representation constraints, W_P, $\forall\, P \in \mathcal{P}$, contains only k candidates and the remainder $m - k$ candidates consisting of $C \setminus W_P$, $\forall\, P \in \mathcal{P}$, may never be selected, thus making the reformulation non-trivial. Even if we relax range of $\ell^D_G, \forall\, G \in \mathcal{G}$, from $[1, \min(k, |G|)]$ to $[0, \min(k, |G|)]$, the following settings are different and we *may not* carry out reformulations:

Observation 2. *$\forall \mu \in \mathbb{Z}$ and $\forall \pi \in \mathbb{Z}$, the following settings of the (μ, π)-DRCF problem are not equivalent: (i) μ=0 and $\pi \geq 1$, (ii) $\mu \geq 3$ and $\pi = 0$, and (iii) $\mu \geq 1$ and $\pi \geq 1$.*

These settings differ as $|G|$ and $|W_P|$ have implications. For instance, using both the constraints and using only representation constraints are mathematically as different as the vertex cover problem on hypergraphs and the vertex cover problem on k-uniform hypergraphs, respectively.

Definition 8. ***(μ, π)-DRCF-size-optimization:*** *In the (μ, π)-DRCF-size-optimization problem, the aim is to find a minimum-size committee $W \subseteq C$ such that W satisfies $|G \cap W| \geq \ell^D_G\ \forall\, G \in \mathcal{G}$ and $|W_P \cap W| \geq \ell^R_P\ \forall\, P \in \mathcal{P}$.*

Theorem 7. *For $\varepsilon > 0$, $\forall \mu \in \mathbb{Z} : \mu \geq 3$, and $\pi = 0$, (μ, π)-DRCF-size-optimization problem is inapproximable within $(1-\varepsilon)\cdot$ $(\ln \mu - \mathcal{O}(\ln\ln\mu))$, even when ℓ^D_G = 1 $\forall\, G \in \mathcal{G}$.*

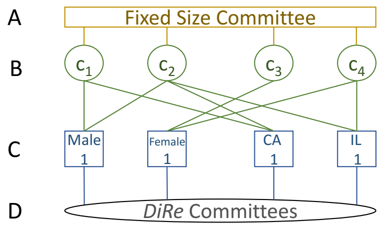


Figure 2: (μ, π)-DRCF as DiReGraph. (A) Global committee size constraint and (C) the diversity/representation constraints connected by edges with (B) the candidates and (D) the DiRe committee.

Theorem 8. *For $\varepsilon > 0$, $\forall \mu \in \mathbb{Z} : \mu \geq 1$, and $\forall \pi \in \mathbb{Z} : \pi \geq 1$, ($\mu$, π)-DRCF-size-optimization problem is inapproximable within a factor of $(1 - \varepsilon) \cdot \ln (|\mathcal{G}| + |\mathcal{P}|)$, even when $\ell^D_G = 1\ \forall\ G \in \mathcal{G}$ and $\ell^R_P = 1\ \forall\ P \in \mathcal{P}$.*

Proof Sketch. We first give a reduction from hitting set (HS) to (1, 1)-DRCF. Next, as the HS problem is equivalent to the minimum set cover problem [Niedermeier and Rossmanith, 2003], the latter's inapproximability [Dinur and Steurer, 2014] holds for our problem. □

Theorem 9. *For $\varepsilon > 0$, $\mu = 0$, and $\forall \pi \in \mathbb{Z} : \pi \geq 1$, ($\mu$, π)-DRCF-size-optimization problem, assuming the UGC, is inapproximable within $k - \varepsilon$, even when $\ell^R_P = 1\ \forall\ P \in \mathcal{P}$.*

Proof Sketch. Assuming the Unique Games Conjecture (UGC) [Khot, 2002], vertex cover problem on k-uniform hypergraphs, for any integer $k \geq 2$, is inapproximable within $k - \varepsilon$ [Bansal and Khot, 2010], which we reduce from[6]. □

6 Heuristic Algorithm

Our model is useful from the social choice theory perspective but it is computationally hard. Hence, we take a pragmatic approach to evaluate its efficiency in practice. We develop a two-stage heuristic algorithm motivated, in part, from distributed constraint satisfaction [Russell and Norvig, 2002].

6.1 DiReGraphs

We represent an instance of the (μ, π)-DRCF problem from Figure 1 as a DiReGraph (Figure 2). The quadrilaterals correspond to constraints and ellipses to candidates. Specifically, there is the global committee size constraint k (Level A), candidates (B), unary constraints ℓ^D_G and ℓ^R_P (C), and the DiRe committee (D). Edges connecting (B) and (C) depend on the candidate's membership in G or W_P. The idea is to have a "network flow" from A to D such that *all* nodes on level C are visited. Specifically, select k candidates (A) from m candidates (B) such that the in-flow τ at the unary constraint nodes (C) is at least ℓ^D_G or ℓ^R_P where $\tau = |W \cap G|$ or $\tau = |W \cap W_P|$. When $\tau \geq \ell^D_G$ or $\tau \geq \ell^R_P$, there will be a *DiRe* committee (D).

[6]For Theorem 9, we assume UGC holds as the result that showed pseudorandom sets in the Grassmann graph have near-perfect expansion completed the proof of 2-to-2 Games Conjecture [Khot *et al.*, 2018], which is a significant evidence towards proving the UGC. Moreover, GapUG($\frac{1}{2}$, ε) is NP-hard, i.e., a weaker version of the UGC holds with completeness $\frac{1}{2}$ (See [Dinur *et al.*, 2018] and "Evidence towards the Unique Games Conjecture" in [Khot *et al.*, 2018]). Without assuming UGC, the result when $\mu = 0$ and $\pi \geq 1$ will change. For $\varepsilon > 0$, the problem is inapproximable within $k - 1 - \varepsilon$ for $k \geq 3$ [Dinur *et al.*, 2005] and within $\sqrt{2} - \varepsilon$ for $k = 2$ [Khot *et al.*, 2018].

Algorithm 1 DiRe Committee Algorithm

Input:
variables $X = \{X_1, \ldots, X_{|\mathcal{G}|+|\mathcal{P}|}\}$
domain $D = (D_1, \ldots, D_{|\mathcal{G}|+|\mathcal{P}|})$: each D_i is G or W_P
constraints $S = \{S_1, \ldots, S_{|\mathcal{G}|+|\mathcal{P}|}\}$: each S_i is ℓ^D_G or ℓ^R_P
Output: set $\mathcal{W}$ of committees : $\forall W \in \mathcal{W}$, $|W \cap D_i| \geq S_i$

Create DiReGraph $DiReG$
SG = subgraph of nodes on levels B & C of $DiReG$
SCC = strongly connected components of SG
for each scc_i, $scc_j \in SCC$ **do**
  for each $X_u = \{X_i \cup X_j\} : X_i \in scc_i, X_j \in scc_j$ **do**
    if !`pairwise_feasible`(X_u, D, S) **return** *false*
  end for
end for
for each scc in SCC **do**
  X_{SCC} = list of X_i for each S_i at level C of scc
  if !`pairwise_feasible`(X_{SCC}, D, S) **return** *false*
end for
Recreate $DiReG$ using reduced D
return `heuristic_backtrack`($\{\}$, $DiReG$, X, D, S)

Example 2. ***DiReGraph:*** *Consider the election instance in Figure 1. The candidate c_2 is a male who is in winning committees of both the states, namely CA (California) and IL (Illinois). Hence, c_2 in DiReGraph (Figure 2) is connected to three sets of constraints, one each for male and two states.*

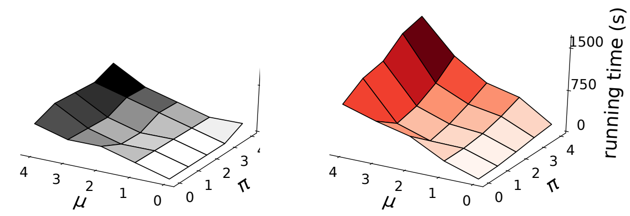
6.2 DiRe Committee Algorithm

The Algorithm 1 has two stages: (i) *preprocessing* reduces the search space used to satisfy the constraints and efficiently finds infeasible instances, and (ii) *heuristic search* of candidates decreases the number of steps needed either to find a feasible committee, or return infeasibility.

Preprocessing. Find the strongly connected components of the graph in time linear in m and n. We only do the pairwise feasibility as three-way (and higher) feasibility tests increase the computational time without improving the scope of finding infeasibility [Russell and Norvig, 2002].

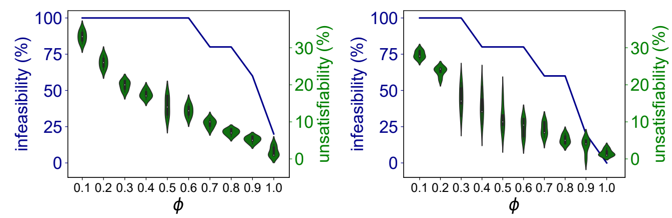
Inter-component pairwise feasibility: Select X_i, X_j corresponding to S_i, S_j on level C of DiReGraph, one each from different components of SCC. Do a pairwise feasibility check for each pair and return infeasibility if any one pair of variables can not return a valid committee. The correctness and completeness of this step are easy. If there are more constraints than the available candidates or if a pair of constraints are pairwise infeasible, then it is impossible to find a feasible solution. For *intra-component pairwise feasibility*, repeat the above procedure but within a component.

Reducing domain: If the algorithm reaches this stage, reduce the domain by removing candidates who, when selected, always return a pairwise infeasible solution.



(a) **timed out instances** (b) **mean running time**

Figure 3: Using **SynData 1**, (a) Proportion (in %) of instances that timed out at 2000 seconds and (b) mean running time of non-timed out instances. Each combination of μ and π has 10 instances.



(a) k-**Borda** (b) **Monroe**

Figure 4: Using **SynData 2**, proportion (in %) of instances that have an infeasible committee and the maximum proportion (in %) of constraints that are unsatisfiable per instance; each ϕ has 5 instances.

Heuristic Backtracking. Use depth-first search for backtracking. Specifically, choose one X_i at a time, and backtrack when X_i has no legal values left to satisfy the constraint. This technique repeatedly chooses an unassigned variable, and then tries all values in its domain, trying to find a solution. If an infeasibility is returned, traverse back by one step and move forward by trying another value.

Select unsatisfied variable: The "minimum-remaining-values" heuristic chooses the variable having the fewest legal values. It picks a variable that is most likely to cause a failure, thereby pruning the search tree as infeasibility may be returned, in turn, avoiding additional searches.

Sort most favorite candidates: This heuristic sorts the candidates in decreasing order of its out-degree. For example, the ordering of candidates (Figure 2) will be c_2 (out-degree = 3), c_1, c_4 (2), and c_3 (1). This reduces the branching factor on future choices by selecting the candidate that is involved in the largest number of constraints, which means they help satisfy the highest proportion of constraints. Next, for completeness and to get multiple DiRe committees, do a "shift-left" such that the second candidate becomes the first and so on. (See [Relia, 2021] for an example explaining the algorithm.)

7 Empirical Analysis

We assess the efficiency of the algorithm and the effect of enforcing the constraints. [Code: https://github.com/KunalRelia/DiReCommittees]

Setup. We run experiments on a Macbook Pro(R) (Docker; 2.2 GHz 6-Core Intel(R) Core i7; 2.2GHz; 16 GB RAM; MacOS Big Sur (v11.1); Python 3.7). We use previously defined voting rules and set constraints at random.

Real Data: The United Nations Resolutions dataset [Voeten, 2014] consists of 193 countries voting for 81 resolutions presented in the UN General Assembly in 2014. We select a 12-sized DiRe committee. Each candidate has two attributes and the voter has one. The Eurovision dataset [Kaggle, 2019] consists of 26 countries ranking the songs performed by each of the 10 finalist countries. We select a 5-sized DiRe committee. Each candidate has two attributes and voter has one.

SynData 1: We set k = 6, n = 100, and m = 50. We generate complete preferences using Repeated Selection Model by setting selection probability $\Pi_{i,j}$ to replicate Mallows' [Mallows, 1957] model ($\phi = 0.5$, random reference ranking σ) (Theorem 3, [Chakraborty *et al.*, 2021]) and preference probability $p(i) = 1$, $\forall i \in [m]$. We randomly divide the candidates and voters into groups and populations, respectively.

SynData 2: We use the same setting as SynData 1 except we fix μ and π to 2 and vary the cohesiveness by setting $\Pi_{i,j}$ to replicate Mallows' model's $\phi \in [0.1, 1]$, in increments of 0.1.

Efficiency of Heuristic Algorithm. All experiments in this section combine instances of k-Borda and β-CC (no significant difference in the running time; paired Student's t-test, $p > 0.05$). We present results for the Monroe rule separately.

Algorithm is efficient: Only 18.90% of 525 instances timed out at 2000 sec (vs 93.71% using a brute-force algorithm). Among the instances that did not time out, the mean running time was 566.48 sec (standard deviation (sd) = 466.66) for k-Borda and β-CC, and 724.39 sec (sd = 575.31) for Monroe (Figure 3). Sparse DiReGraphs increased the efficiency.

Comparison with ILP: Our algorithm, which (i) handles constraints and any committee selection rule and (ii) terminated in (avg) 724 sec, has a clear edge over ILP [Skowron *et al.*, 2015] as it scales linearly. Promisingly, the first committee returned by the algorithm in <120 sec was the winning DiRe committee among 63% of all instances.

Feasibility and Cost of Fairness. All experiments consider each rule separately (paired Student's t-test, $p < 0.05$).

There was a negative correlation between the maximum proportion of unsatisfied constraints and ϕ, for all the three scoring rules (mean Pearson's ρ = -0.95, p <0.05). It was easier to satisfy the constraints when the cohesiveness (ϕ) was high, which led to lower infeasibility (Figure 4).

Real Datasets. The mean ratio of utilities of constrained to unconstrained committees was 0.93 (sd = 0.08). Importantly, the algorithm always terminated in under 102 sec.

8 Conclusion

With diversity in multiwinner elections becoming necessary, we showed that **only diversity can do more harm than good** as its cost may disproportionately be paid more by historically under-represented voter population. Hence, we developed (μ, π, f)-DRCWD model, which enforces diversity **and** representation constraints. We studied its computational properties theoretically and empirically. We also delineate the candidate and voter attributes even when they coincide as just like correlation does not imply causation, we observe that **diversity does not imply representation and vice versa**. This has relevance to equality-related United Nations' sustainable development goals. (See [Relia, 2021] for full acknowledgement.)

References

[Alimonti and Kann, 1997] Paola Alimonti and Viggo Kann. Hardness of approximating problems on cubic graphs. In *CIAC*, 1997.

[Bansal and Khot, 2010] Nikhil Bansal and Subhash Khot. Inapproximability of hypergraph vertex cover and applications to scheduling problems. In *ICALP*, 2010.

[Brams, 1990] Steven J Brams. Constrained approval voting: A voting system to elect a governing board. *Interfaces*, 20(5):67–80, 1990.

[Bredereck *et al.*, 2018] Robert Bredereck, Piotr Faliszewski, Ayumi Igarashi, Martin Lackner, and Piotr Skowron. Multiwinner elections with diversity constraints. In *AAAI*, 2018.

[Brill *et al.*, 2018] Markus Brill, Jean-François Laslier, and Piotr Skowron. Multiwinner approval rules as apportionment methods. *Journal of Theoretical Politics*, 30(3):358–382, 2018.

[Celis *et al.*, 2018] L Elisa Celis, Lingxiao Huang, and Nisheeth K Vishnoi. Multiwinner voting with fairness constraints. *In IJCAI*, 2018.

[Chakraborty and Kolaitis, 2021] Vishal Chakraborty and Phokion G Kolaitis. Classifying the complexity of the possible winner problem on partial chains. In *AAMAS*, 2021.

[Chakraborty *et al.*, 2021] Vishal Chakraborty, Theo Delemazure, Benny Kimelfeld, Phokion G Kolaitis, Kunal Relia, and Julia Stoyanovich. Algorithmic techniques for necessary and possible winners. *ACM/IMS Transactions on Data Science*, 2(3):1–23, 2021.

[Chamberlin and Courant, 1983] John R Chamberlin and Paul N Courant. Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review*, 77(3):718–733, 1983.

[Dinur and Steurer, 2014] Irit Dinur and David Steurer. Analytical approach to parallel repetition. In *STOC*, 2014.

[Dinur *et al.*, 2005] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered pcp and the hardness of hypergraph vertex cover. *SIAM Journal on Computing*, 34(5):1129–1146, 2005.

[Dinur *et al.*, 2018] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In *STOC*, 2018.

[Elkind *et al.*, 2017] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. Properties of multiwinner voting rules. *Social Choice and Welfare*, 48(3), 2017.

[Faliszewski *et al.*, 2016] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Committee scoring rules: Axiomatic classification and hierarchy. In *IJCAI*, 2016.

[Faliszewski *et al.*, 2017] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice*, 74:27–47, 2017.

[Garey and Johnson, 1979] Michael R Garey and David S Johnson. *Computers and intractability*, volume 174. Freeman San Francisco, 1979.

[Hodge and Klima, 2018] Jonathan K Hodge and Richard E Klima. *The mathematics of voting and elections: a hands-on approach*. American Mathematical Soc. (30), 2018.

[Kaggle, 2019] Kaggle. *Eurovision Song Contest 1975-2019, (accessed July 22, 2020)*, 2019.

[Khot *et al.*, 2018] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *FOCS*, 2018.

[Khot, 2002] Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC*, 2002.

[Koriyama *et al.*, 2013] Yukio Koriyama, Antonin Macé, Rafael Treibich, and Jean-François Laslier. Optimal apportionment. *Journal of Political Economy*, 121(3):584–608, 2013.

[Lang and Skowron, 2018] Jérôme Lang and Piotr Skowron. Multi-attribute proportional representation. *Artificial Intelligence*, 263:74–106, 2018.

[Mallows, 1957] C. L. Mallows. Non-null ranking models. i. *Biometrika*, 44(1-2):114–130, June 1957.

[Monroe, 1995] Burt L Monroe. Fully proportional representation. *American Political Science Review*, 89(4):925–940, 1995.

[Niedermeier and Rossmanith, 2003] Rolf Niedermeier and Peter Rossmanith. An efficient fixed-parameter algorithm for 3-hitting set. *Journal of Discrete Algorithms*, 2003.

[Potthoff, 1990] R Potthoff. Use of linear programming for constrained approval voting. *Interfaces*, 20(5), 1990.

[Procaccia *et al.*, 2008] Ariel D Procaccia, Jeffrey S Rosenschein, and Aviv Zohar. On the complexity of achieving proportional representation. *Social Choice and Welfare*, 30(3):353–362, 2008.

[Relia, 2021] Kunal Relia. Dire committee: Diversity and representation constraints in multiwinner elections. *arXiv preprint arXiv:2107.07356*, 2021.

[Russell and Norvig, 2002] Stuart Russell and Peter Norvig. Artificial intelligence: A modern approach. *Pearson Education Limited*, 2002.

[Sánchez-Fernández *et al.*, 2017] Luis Sánchez-Fernández, Edith Elkind, Martin Lackner, Norberto Fernández, Jesús A Fisteus, Pablo Basanta Val, and Piotr Skowron. Proportional justified representation. In *AAAI*, 2017.

[Skowron *et al.*, 2015] Piotr Skowron, Piotr Faliszewski, and Arkadii Slinko. Achieving fully proportional representation: Approximability results. *Artificial Intelligence*, 222:67–103, 2015.

[Uckelman *et al.*, 2009] Joel Uckelman, Yann Chevaleyre, Ulle Endriss, and Jérôme Lang. Representing utility functions via weighted goals. *Mathematical Logic Quarterly*, 55(4):341–361, 2009.

[Voeten, 2014] Erik Voeten. *UN General Assembly Votes, 1946-2015, (accessed July 22, 2020)*, 2014.