

# Size-Relaxed Committee Selection under the Chamberlin-Courant Rule

Tao Xiao

Department of Computer Science,  
Shanghai Jiao Tong University  
Shanghai, China  
xt\_1992@sjtu.edu.cn

Sujoy Sikdar

Department of Computer Science and Engineering,  
Washington University in St. Louis  
St. Louis, USA  
sujoyks@gmail.com

## ABSTRACT

The Chamberlin-Courant (CC) family of committee selection rules aim to select a committee of size  $k$  from a set of  $m$  candidates to maximize the satisfaction of  $n$  agents. The satisfaction of an agent from a committee depends only on the rank of her favorite candidate and is determined by a satisfaction function. Unfortunately, computing an optimal committee of size  $k$  is hard in general, which has led to the development of approximation algorithms that select a committee of size  $k$ , which guarantees some fraction of the optimal satisfaction. However, there is often some flexibility in the size of the committee to be selected.

In this paper, we initiate the study of *size-relaxed committee selection* for the family of CC rules. Our main results are polynomial-time algorithms to select committees of size at most  $k \cdot O(\log n)$ , whose satisfaction is guaranteed to be at least that of the optimal committee of size  $k$ , and show that this is tight. We also provide a constant-factor approximation algorithm for a class of approval ballot based CC rules.

## KEYWORDS

Chamberlin-Courant Rule; Size-Relaxed Committee Selection; Approximation Algorithms

### ACM Reference Format:

Tao Xiao and Sujoy Sikdar. 2020. Size-Relaxed Committee Selection under the Chamberlin-Courant Rule. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020)*, Auckland, New Zealand, May 9–13, 2020, IFAAMAS, 9 pages.

## 1 INTRODUCTION

Consider the problem of a university library offering a collection of journals, an airline offering a small collection of movies to customers on a flight, or a company offering a portfolio of its products [11, 22, 25, 26]. These are real-world examples of the *committee selection problem* where the goal is to select a *committee* of  $k$  candidates for a collection of  $n$  agents who have ordinal preferences over  $m > k$  candidates to maximize the *satisfaction* experienced by the agents. Often, there is some flexibility in the size of the selected committee, and guaranteeing a certain level of satisfaction is the more important consideration. Indeed, a university library may have some flexibility in the number of journals they carry, congressional committees can often be expanded, and a company

can add a few products to its portfolio, in order to provide a target satisfaction guarantee.

The Chamberlin-Courant (CC) [6] family of committee selection rules are perhaps the most popular and well-known among rules that aim at maximizing satisfaction for a committee of fixed size  $k$ . Here, the satisfaction an agent associates with a committee depends only on the position of her highest-ranked candidate who is a member of the committee. The satisfaction is measured by a *satisfaction function*, which maps each rank to a satisfaction score. A prominent example is the Borda-CC rule, where the satisfaction an agent derives from a committee is the Borda score of her highest-ranked member of the committee. In general, a *utilitarian*  $\alpha$ -CC rule is characterized by a satisfaction function  $\alpha$ , and selects a committee which maximizes total satisfaction, while an *egalitarian*  $\alpha$ -CC rule selects a committee which maximizes the satisfaction of the least satisfied agent. Unfortunately, the decision versions of selecting a committee of size  $k$  is NP-hard for utilitarian [22, 23, 27] and egalitarian [27]  $\alpha$ -CC rules.

A common approach to circumvent the computational hardness of CC rules is to fix the size of the committee at  $k$ , and compute a committee whose satisfaction is *approximately* that of the optimal committee. Lu and Boutilier [22] provide a  $(1 - \frac{1}{e})$ -approximation algorithm, and Skowron et al. [27] provide a polynomial-time approximation scheme for the Borda-CC rule, to find a committee of size  $k$  which approximates the satisfaction of the optimal committee. However, as Skowron et al. [27] argued, approximating satisfaction while fixing the size of the committee raises the concern that an agent, a candidate, or other parties, may identify a committee with higher satisfaction and demand it to be selected instead.

This concern leads to the following natural notion of *size-relaxed committee selection* introduced by Sekar et al. [25]: *Can we compute a committee of size at most  $\gamma \cdot k$  in polynomial time whose satisfaction is at least that of the optimal committee of size  $k$ ?*

Our work follows in the research agenda initiated recently by Kilgour [19], who introduces the notion of selecting committees without fixing the size of the committee. More recently, Sekar et al. [25] consider the class of Condorcet-consistent [8] committee selection rules due to [15], and provide an approximation algorithm for the Maximin rule which picks a committee of size at most  $2k$ , while guaranteeing that the selected committee meets the Maximin objective of the optimal committee of size  $k$ . However, to the best of our knowledge, nothing is known about size-relaxed committee selection for the celebrated Chamberlin-Courant rule.

(Dis)satisfaction function family	Utilitarian		Egalitarian	
	Lower bound	Upper bound	Lower bound	Upper bound
UPU $\cap$ PB	$\Omega(\log n)$ [Theorem 2]	$O(\log n)$ [Theorem 3]	$\Omega(\log n)$ [Theorem 2]	$O(\log n)$ [Theorem 4]
PU		$n$ [Folklore]		
t-approval	APX-hard [Corollary 3]	$O(\log n)$ [Theorem 3]	APX-hard [Corollary 3]	$t$ [Theorem 5]

**Table 1: Size-Relaxed committee selection.** UPU stand for the set of universally polynomially unbounded (dis)satisfaction function families, PB stands for the set of polynomially bounded (dis)satisfaction function families, PU stands for the set of polynomially unbounded (dis)satisfaction function families. UPU $\cap$ PB stands for the (dis)satisfaction function families that lies in both UPU and PB.

## 1.1 Our Contributions

We provide the first results on *size-relaxed committee* selection for the CC family of committee selection rules, to the best of our knowledge. We consider versions of the CC family of rules characterized by satisfaction and dissatisfaction functions, and with utilitarian and egalitarian objectives, and provide new algorithmic and complexity results. Table 1 summarizes our results on lower and upper bounds for the approximability of size-relaxed committee selection.

In Theorem 1, we prove an equivalence relation between the hardness of approximating satisfaction and dissatisfaction function-based versions of Chamberlin-Courant rules for utilitarian and egalitarian objectives. We show that if a satisfaction function based rule is hard to approximate, then so is the corresponding dissatisfaction function based rule hard to approximate. Our proof establishes a reduction between the problem of approximating CC rules for satisfaction and dissatisfaction functions. As we note in Remark 1, this equivalence also applies for upper bound results.

In Theorem 2, we prove that it is hard to approximate to within a factor of  $o(\log n)$ , utilitarian and egalitarian CC rules based on the class of *universally polynomially unbounded* (dis)satisfaction functions (Definition 3).

In Theorem 3, we show that the lower bound is tight for utilitarian CC rules, for the class of *polynomially bounded* (dis)satisfaction functions (Definition 4). In Theorem 4, we show that the lower bound is tight for egalitarian CC rules for universally polynomially unbounded (dis)satisfaction functions.

We also provide constant factor approximation algorithms for the important class of *t*-approval based egalitarian CC rules for fixed constant *t* in Theorem 5 and Corollary 2. Notice that there is a deep connection between *t*-approval based CC rules and the VERTEXCOVER problem. We show in Corollary 3 that for *t*-approval based CC rules both utilitarian and egalitarian rules are APX-hard.

Our results provide a classification of families of (dis)satisfaction functions based on hardness of size-relaxed committee selection. Our approximation algorithms for the upper bound results rely on a common framework involving a two-step greedy-based algorithm, wherein (Step 1) we construct an instance of the WEIGHTED-MAXIMUM-*k*-COVERAGE problem, and in (Step 2) we apply a greedy-based algorithm to the constructed instance, while simultaneously selecting a committee. We note that all our results apply even when agents may have different (dis)satisfaction functions. However, for simplicity, we will provide the proofs for the case where all agents have the same (dis)satisfaction function, and briefly explain how our proof can be extended.

## 2 RELATED WORK AND DISCUSSIONS

We first note that Kocot et al. [20] also defines the notion of polynomially bounded satisfaction functions. They call a satisfaction function family as polynomially bounded, if for any  $m \in \mathbb{N}$ , and any  $l \leq m$ ,  $\alpha^m(l) \in \text{Poly}(m)$ , where  $\alpha^m(l)$  denotes the satisfaction of satisfaction function  $\alpha^m$  at position  $l$ . Thus, Borda satisfaction function is regarded as polynomially bounded in [20]. They provide efficient algorithms for multi-goal committee selection to provide lower bound guarantees for multiple satisfaction functions simultaneously, under the restriction that the satisfaction functions are polynomially bounded. We also adopt this notion to give our upper bound results for utilitarian CC rules. Our results classify the family of Chamberlin-Courant rules based on their approximability: we provide a  $O(\log n)$ -approximation algorithm for polynomially bounded (dis)satisfaction function based CC rules, while CC rules characterized by universally polynomially unbounded (dis)satisfaction functions are LOG-APX-hard, i.e., the problem is hard to approximate within a factor of  $o(\log n)$ . For the universally polynomially bounded notion introduced in our paper, we focus on the (dis)satisfaction function family from a different point of view: we use this notion to capture those (dis)satisfaction function families that are LOG-APX-hard to approximate due to the reason that there is enough gap for us to show the reduction works. We propose this notion to pioneer the study of doing classifications on (dis)satisfaction function families.

The problem of approximate committee selection has recently attracted a lot of attention in social choice theory. Exact hardness results for the Chamberlin-Courant rule are shown in [22, 23], for *t*-approval and Borda satisfaction respectively. [27] provides hardness of approximation results for the Monroe and Chamberlin-Courant rules. [22] provides a greedy  $(1 - \frac{1}{e})$ -approximation algorithm for utilitarian Borda-CC rule, which is improved to a PTAS by [27]. [3] provides approximation algorithms for egalitarian Borda-CC, as well as hardness of approximation results for other versions. [28] shows that the Chamberlin-Courant rule for utilitarian approval satisfaction is equivalent to the maximum coverage problem. For exact parameterized complexity results of egalitarian Borda-CC, Bloc, *k*-Borda rules we refer to Table 2 in [2].

Multi-winner selection with a variable number of winners is also studied in recent years. The idea is in some sense similar: they do not fix the number of winners in an election, but to find a number of winners that make most of the voters happy. To the best of our knowledge, current work only focuses on approval ballots, and focus on the computational complexity of this problem under various approval-based voting rules [12, 14, 19].

The closest work to ours is [25], which studies Condorcet consistent committee selection rules, where for *Maximin<sup>k</sup>* they provide an approximation algorithm to find a committee of size at most  $2k$  whose score is at least that of the optimal committee of size  $k$ .

The idea of size-relaxation is also quite common in other problems beyond committee selection. [21] studies a sized-relaxed version of the facility location problem, where they allow adding more than  $k$  locations to (approximately) meet the same objective as the optimal choice of  $k$  locations. In job scheduling, usually, the problem is to design scheduling that minimizes the makespan, i.e., the minimum completion time. A relaxed version is to use more machines that maintain a given makespan, which is similar to BINPACKING [7]. As another example, in algorithmic mechanism design with  $n$  buyers, size relaxation is often studied under the name *competition complexity*, with the idea of adding more (identical) agents in a “simple” auction, achieving optimal revenue of  $n$  buyers case [4, 5, 10, 13, 24].

### 3 PRELIMINARIES

For any  $n \in \mathbb{N}$ , let  $[n] = \{1, \dots, n\}$ . An *election* is a tuple  $E = (N, C, \vec{P})$ , where  $N = [n]$  is a set of  $n$  agents,  $C = [m]$  is a set of  $m$  candidates, and  $\vec{P} = (P_i)_{i \in N}$  is a *preference profile*, where each  $P_i$  is a strict ranking over  $C$ , representing the preference of agent  $i$ . For each agent  $i$  we use  $>_i$  to denote the strict *preference order* over  $C$  w.r.t.  $P_i$ . We define  $\text{pos}_i(j)$  to be the rank of candidate  $j$  in the ranking  $P_i$ . Throughout the paper, we assume that  $m = \text{Poly}(n)$ . A *committee* of size  $k$  is any subset  $C \subseteq C$ ,  $|C| = k$  of the candidates.  $C_k$  denotes the set of all committees of size  $k$ . We use  $[m]_k$  to denote the set of all  $k$ -length increasing sequences of numbers from  $[m]$ . Given a committee  $C$  of size  $k$ , we use  $\text{pos}_i(C) \in [m]_k$  to denote the sequence obtained by sorting  $\{\text{pos}_i(j) : j \in C\}$  in increasing order, we use  $i_{C,l}$  to denote the  $l$ -th member of  $\text{pos}_i(C)$ .

**Definition 1.** A (dis)satisfaction function is a non-increasing (non-decreasing) monotonic mapping  $\alpha^m : [m] \rightarrow \mathbb{R}^+$  ( $\delta^m : [m] \rightarrow \mathbb{R}^+$ ). The value  $\alpha^m(l)$  ( $\delta^m(l)$ ) is an agent’s (dis)satisfaction from being represented by a candidate that she ranks at position  $l$ .

For example, the *Borda* satisfaction function for  $m$  candidates is defined as  $\text{Borda}^m(l) = m - l$ . The  $t$ -approval (dis)satisfaction function for  $m \geq t$  candidates is defined as  $\alpha_t^m(l) = 1$  ( $\delta_t^m(l) = 0$ ) for  $l \leq t$ , and  $\alpha_t^m(l) = 0$  ( $\delta_t^m(l) = 1$ ) otherwise.

**Definition 2.** A family of (dis)satisfaction function  $\vec{\alpha}$  ( $\vec{\delta}$ ) is an infinite-dimensional vector  $\vec{\alpha} = (\alpha^1, \alpha^2, \dots)$  ( $\vec{\delta} = (\delta^1, \delta^2, \dots)$ ) such that  $\alpha^{\kappa+1}(l+1) = \alpha^\kappa(l)$  ( $\delta^{\kappa+1}(l) = \delta^\kappa(l)$ ) holds for all  $\kappa \in \mathbb{N}$  and  $l \in [\kappa]$ .

**Definition 3.** A dissatisfaction function family  $\vec{\delta}$  is *universally polynomially bounded*, if for any constant  $c > 0$ , there is a polynomial  $p(\kappa) \in \Omega(\kappa^c)$  such that

$$\frac{\delta^\kappa(\kappa)}{\delta^\kappa(l)} \leq p(\kappa)$$

holds for every  $\kappa \in \mathbb{N}^+$ , and every  $l \in [\kappa]$  such that  $l \geq \kappa^c$  and  $\delta^\kappa(l) \neq 0$ . Otherwise,  $\vec{\delta}$  is *universally polynomially unbounded*. Similar definitions hold for satisfaction case, by replacing  $\delta^\kappa(l) \neq 0$  with  $\alpha^\kappa(l) \neq 0$  and replacing  $\frac{\delta^\kappa(\kappa)}{\delta^\kappa(l)}$  with  $\frac{\alpha^\kappa(k-l)}{\alpha^\kappa(1)}$ .

We discuss this notion through examples: for the Borda dissatisfaction function family where  $\delta^\kappa(\kappa) = \kappa - 1$  for all  $\kappa \in \mathbb{N}^+$ . Since  $\kappa - 1 \in \Omega(\kappa)$  when  $c = 1$ , Borda is a universally polynomially unbounded satisfaction function family. The  $t$ -approval satisfaction function family,  $\vec{\alpha}_t$ , is universally polynomially bounded, as  $\alpha_t^m(m) = 1 \in o(m^c)$  for any constant  $c > 0$ . The family of dissatisfaction functions, where  $\forall m \in \mathbb{N}^+$ ,  $\delta^m(l) = \log_2 l$ , is a universally polynomially bounded dissatisfaction function, since  $\delta^\kappa(\kappa)/\delta^\kappa(2) = \log_2 \kappa \in o(\kappa^c)$  for any constant  $c > 0$ .

**Definition 4.** [20] A dissatisfaction function family  $\vec{\delta}$  is *polynomially bounded*, if there exists a polynomial  $p(\cdot)$  such that

$$\frac{\delta^\kappa(\kappa)}{\delta^\kappa(l)} \leq p(\kappa)$$

holds for every  $\kappa \in \mathbb{N}^+$ , and every  $l \in [\kappa]$  such that  $\delta^\kappa(l) \neq 0$ . Otherwise,  $\vec{\delta}$  is *polynomially unbounded*. Similar definition holds for satisfaction case, by replacing  $\delta^\kappa(l) \neq 0$  with  $\alpha^\kappa(l) \neq 0$ , and replacing  $\frac{\delta^\kappa(\kappa)}{\delta^\kappa(l)}$  with  $\frac{\alpha^\kappa(l)}{\alpha^\kappa(1)}$ .

An example of a family of polynomially bounded functions is the Borda dissatisfaction function family where  $\delta^\kappa(\kappa) = \kappa - 1$  for all  $\kappa \in \mathbb{N}$ . It is easy to see that it is polynomially bounded by choosing  $p(\kappa) = \kappa - 1$ . A polynomially unbounded dissatisfaction function family is the one defined:  $\forall m \in \mathbb{N}$ ,  $\delta^m(l) = 2^l$ , since there does not exist a polynomial  $p(\cdot)$  that makes  $\delta^\kappa(\kappa)/\delta^\kappa(1) \leq p(\kappa)$  for every  $\kappa \in \mathbb{N}^+$ .

#### 3.1 Size-Relaxed Computational Problems

Given an election with  $n$  agents and  $m$  candidates, and an integer  $k < m$ , we define the size-relaxed versions of the traditional committee selection problems where the constraint on the size of the committee is relaxed, but the selected committee provides the (dis)satisfaction guarantee of the optimal committee of size  $k$ . For notation simplicity we assume that every agent has the same family of (dis)satisfaction function, and will discuss the same results for agents having different satisfaction functions after each proof.

**Definition 5.** Given an election with  $n$  agents and  $m$  candidates, and an integer  $k < m$ , we are asked to find a committee  $C^*$ ,  $|C^*| \geq k$  of small size in polynomial time such that in:

- RELAXED-UTILITARIAN- $\vec{\alpha}$ -CC with satisfaction function family  $\vec{\alpha}$ ,  $\sum_{i \in [n]} \alpha^m(i_{C^*,1}) \geq \max_{C \in C_k} \sum_{i \in [n]} \alpha^m(i_{C,1})$ .
- RELAXED-EGALITARIAN- $\vec{\alpha}$ -CC with satisfaction function family  $\vec{\alpha}$ ,  $\min_{i \in [n]} \alpha^m(i_{C^*,1}) \geq \max_{C \in C_k} \min_{i \in [n]} \alpha^m(i_{C,1})$ .
- RELAXED-UTILITARIAN- $\vec{\delta}$ -CC with dissatisfaction function family  $\vec{\delta}$ ,  $\sum_{i \in [n]} \delta^m(i_{C^*,1}) \leq \min_{C \in C_k} \sum_{i \in [n]} \delta^m(i_{C,1})$ .
- RELAXED-EGALITARIAN- $\vec{\delta}$ -CC with dissatisfaction function family  $\vec{\delta}$ ,  $\max_{i \in [n]} \delta^m(i_{C^*,1}) \leq \min_{C \in C_k} \max_{i \in [n]} \delta^m(i_{C,1})$ .

If we are allowed to use exponential time, then by brutal-force one can find a committee  $C^*$  of size  $k$  that satisfy the above properties. To measure the “small” here, we use the notion of approximation ratio.

**Definition 6.** We say that algorithm  $\mathcal{A}$  is an  $r$ -approximation algorithm for committee selection problems in Definition 5, if  $\mathcal{A}$  can find a feasible committee  $C^*$  of such that  $|C^*| \leq rk$ .

We now recall the definition of the following NP-hard problems that will be used in this paper.

**Definition 7 (DOMINATINGSET).** Given a graph  $G = (V, E)$ , where  $V$  is a vertex set,  $E$  is an edge set, and  $k$  is a positive integer, we are asked whether there exists a vertex set  $W \subset V$  such that  $|W| \leq k$ , and for each vertex  $v \in V$ , there is a vertex  $w \in W$  such that  $w, v$  are adjacent. We use  $I = (V, E, k)$  to denote an instance of DOMINATINGSET.

**Definition 8 (SETCOVER).** Given a ground set  $U = [n]$ , a family  $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$  of subsets of  $U$ , and a positive integer  $k$ , we are asked whether there exists a subset  $J$  of  $[m]$  such that  $|J| \leq k$ , and  $\bigcup_{j \in J} F_j = U$ . We use  $I = (U, \mathcal{F}, k)$  to denote an instance of SETCOVER.

**Definition 9 (WEIGHTED-MAXIMUM- $k$ -COVERAGE).** Given a ground set  $U = [n]$ , a weight function  $w : [n] \rightarrow \mathbb{R}^+$ , a family  $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$  of subsets of  $U$ , and a positive integer  $k$ , we are asked to find a subset  $J \subseteq [m]$  of size  $k$ , such that  $w(\bigcup_{I \in J} F_I) := \sum_{i \in \bigcup_{I \in J} F_I} w(i)$  is maximized. We note that our weight function  $w$  is additive, thus by defining weights for each element in the ground set gives a full characterization of a weight function. We use  $I = (U, w, \mathcal{F}, k)$  to denote an instance of WEIGHTED-MAXIMUM- $k$ -COVERAGE.

## 4 LOWER BOUNDS ON EFFICIENT APPROXIMATION OF SIZE-RELAXED COMMITTEE SELECTION

We start by proving an equivalence in the hardness of approximating satisfaction and dissatisfaction based Chamberlin-Courant rules in Theorem 1. We provide the proof for the egalitarian objectives only in the interest of space. The proof for the utilitarian objective is similar.

**Theorem 1.** *For any integer  $m$ , and any dissatisfaction function family  $\vec{\delta}$ , if there is no algorithm that achieves  $r$ -approximation for RELAXED-EGALITARIAN- $\vec{\delta}$ -CC, then there is no algorithm that achieves  $r$ -approximation for RELAXED-EGALITARIAN- $\vec{\alpha}$ -CC, where  $\alpha^m(\cdot) : \alpha^m(i) = \delta^m(m) - \delta^m(i)$  for all  $i \in [m]$ .*

**PROOF.** By construction we know that both problems share the same candidate set. For any candidate set  $K$  that has a dissatisfaction of less than  $\min_{C \in C_k} \max_{i \in [n]} \delta^m(i_{C,1})$  for RELAXED-EGALITARIAN- $\vec{\delta}$ -CC, we know that the same candidate set  $K$  has a satisfaction of greater than  $\max_{C \in C_k} \min_{i \in [n]} \alpha^m(i_{C,1})$  for RELAXED-EGALITARIAN- $\vec{\alpha}$ -CC.

We prove by contradiction. Suppose there is an algorithm  $\mathcal{A}$  for RELAXED-EGALITARIAN- $\vec{\alpha}$ -CC that achieves  $r$  approximation, with output candidate set  $\hat{K}$ . By the definition of approximation, we know that  $|\hat{K}| \leq kr$ . By the above argument, choosing  $\hat{K}$  for RELAXED-EGALITARIAN- $\vec{\delta}$ -CC has a dissatisfaction of less than  $\min_{C \in C_k} \max_{i \in [n]} \delta^m(i_{C,1})$ . This gives an algorithm that achieves  $r$  approximation for RELAXED-EGALITARIAN- $\vec{\delta}$ -CC, a contradiction. This concludes the proof.  $\square$

*Remark 1.* It is not hard to see that the same equivalence holds for upper bound results. We will use this in Section 4.

In light of Theorem 1 and its implications for both lower bound and upper bound results, we will only consider either satisfaction or dissatisfaction based Chamberlin-Courant rules. Unless otherwise stated, we will prove for either rule, and the same results also apply to the other rule.

*Hardness of Approximation for Size-Relaxed Committee Selection.* We prove an  $o(\log n)$  lower bound for the approximability of size-relaxed committee selection in Theorem 2 by a reduction from DOMINATINGSET. For convenience, we prove the results for dissatisfaction functions. Results for satisfaction functions follow from Theorem 1.

**Theorem 2.** *For any dissatisfaction function family  $\vec{\delta}$  that is universally polynomially unbounded, there is no polynomial time algorithm that has an approximation ratio of  $o(\log n)$  unless  $P = NP$  for:*

- RELAXED-UTILITARIAN- $\vec{\delta}$ -CC,
- RELAXED-EGALITARIAN- $\vec{\delta}$ -CC.

To prove this theorem, we first introduce the following lemma.

**Lemma 1.** *For any dissatisfaction function family  $\vec{\delta}$  that is universally polynomially unbounded, for large enough  $n \in \mathbb{N}$ , there exists an integer  $r \in \text{Poly}(n)$  such that  $\delta^r(r) \geq n\delta^r(n)$ .*

**PROOF.** By the definition of universally polynomially unbounded, we know that there exists a constant  $c > 0$  such that for any polynomial  $p(\kappa) \in \Omega(\kappa^c)$ , there exists  $\kappa \in \mathbb{N}$  and  $l \geq \kappa^c$  such that  $\delta^\kappa(\kappa)/\delta^\kappa(l) \geq p(\kappa)$ . Let  $n = \lfloor p(\kappa) \rfloor = \lfloor \kappa^c \rfloor$ , we know that  $\delta^\kappa(\kappa) \geq n\delta^\kappa(n)$  and  $\kappa \in \text{poly}(n)$ . Take  $r = \lfloor \kappa + 1 \rfloor$  gives the desired integer  $r$ .  $\square$

Now we prove Theorem 2.

**PROOF.** We first prove for RELAXED-UTILITARIAN- $\vec{\delta}$ -CC. The high-level idea is to establish a reduction from DOMINATINGSET, and show that the reduction is approximation preserving.

*Reduction:* Let  $I = (V, E, k)$  be an instance of DOMINATINGSET. We construct an instance  $J = (\mathcal{N}, C, \vec{P})$  of RELAXED-UTILITARIAN- $\vec{\delta}$ -CC as follows:

Let  $\mathcal{N} = V$  be the agents, and the set of candidates  $C = \bigcup_{j=1}^n C_j$  where each  $C_j$  corresponds to vertex  $j$  in  $V$  and contains exactly  $r$  candidates, where  $r$  is the smallest value for which  $\delta^m(r) \geq n\delta^m(n)$  (the existence of  $r$  follows from Lemma 1). For each  $C_j$  we pick one candidate  $c_j$  as representative of  $C_j$ . Denote by  $\Gamma^+(i)$  the set of vertices that is either  $i$  itself, or adjacent to  $i$  in graph  $G$ . For each agent  $i \in \mathcal{N}$ , we set the preference  $>_i$  as follows:

$$c_{\Gamma^+(i)} >_i C_{\Gamma^+(i)} - c_{\Gamma^+(i)} >_i C - C_{\Gamma^+(i)},$$

where  $c_{\Gamma^+(i)}$  includes all candidate  $c_j$  such that  $j \in \Gamma^+(i)$  and  $C_{\Gamma^+(i)}$  includes all candidate set  $C_j$  such that  $j \in \Gamma^+(i)$ . The ordering of candidates in each part can be set arbitrarily.

Let  $l = \min_{C \in C_k} \sum_{i \in [n]} \delta^m(i_{C,1})$ , the minimum utilitarian dissatisfaction with  $k$  candidates. We claim that  $l \leq n\delta^m(n)$  if and only if  $I$  is a YES instance of DOMINATINGSET.

( $\Leftarrow$ ) If  $I$  a YES instance of DOMINATINGSET, then there is a dominating set  $K$  of size  $k$ . Thus for each  $i \in V$  there is a  $j \in K$  such that  $i$  and  $j$  are the same or adjacent. Let  $T = \{c_j | j \in K\}$  be our chosen candidate set of size  $k$ . By the above argument, for each

agent  $i \in \mathcal{N}$ , there is a  $j \in T$  s.t.  $pos_i(j) \leq n$ . This gives  $k$  candidates such that the total dissatisfaction is at most  $n \cdot \delta^m(n)$ .

( $\Rightarrow$ ) If  $l \leq n\delta^m(n)$ , we prove by contradiction that  $I$  is a YES instance of DOMINATINGSET. If not, then there is no dominating set of size  $k$ , thus we know that for any size  $k$  subset  $K$  of  $V$ , there exists an  $i \in V$  such that for any vertex  $j \in K$ ,  $i$  and  $j$  are neither the same nor adjacent. This implies that for instance  $J$ , by choosing any  $k$  candidates  $T$ , we know that there is an agent  $i$  such that all candidates in  $T$  get a dissatisfaction of at least  $\delta^m(r)$  (since they are all in  $C - C_{\Gamma^+(i)}$ ). Thus, the egalitarian dissatisfaction for choosing any  $k$  candidates is at least  $\delta^m(r) > \delta^m(n)$ , which contradicts. This gives the reduction which implies this problem is NP-hard. The approximation preserving property directly follows from Lemma 2. This concludes the proof.  $\square$

The following lemma shows that the reduction is approximation preserving.

**Lemma 2.** *If there is an  $o(\log n)$ -approximation algorithm for RELAXED-UTILITARIAN- $\bar{\delta}$ -CC for universally polynomially unbounded dissatisfaction function family  $\bar{\delta}$ , then there is an  $o(\log n)$  approximation algorithm for DOMINATINGSET.*

PROOF. Suppose that algorithm  $\mathcal{A}$  has approximation ratio of  $o(\log n)$ . For a DOMINATINGSET instance  $I = (V, E, k)$ , let  $k^*$  be the smallest  $k$  that makes  $I$  a YES instance. let  $J = (\mathcal{N}, C, \bar{P})$  be the constructed election instance in the proof of Theorem 2. Note that in the construction we do not need to know the value of  $k$ . We run algorithm  $\mathcal{A}$  with  $k$  enumerating from 1 to  $n$ , and check if the output candidate set has a total dissatisfaction of less than  $n\delta^m(n)$ . If for  $\hat{k}$  the condition holds, we end our algorithm with a candidate set  $\hat{C}$  of size  $\hat{k} \cdot o(\log n)$ . Let  $T = \{j \mid |\hat{C} \cap C_j| \geq 1\}$ . We claim that  $T$  is a dominating set of  $I$ .

If it is not the case, then there exists a vertex  $i \in V$  such that for any vertex  $j \in T$ ,  $i$  and  $j$  are neither the same, nor adjacent. This means that agent  $i$ 's dissatisfaction, when choosing candidate set  $\hat{C}$ , is at least  $\delta^m(r) \geq n\delta^m(n)$ , a contradiction.

Next we prove that  $\hat{k} \leq k^*$ . If not, then we have already run algorithm  $\mathcal{A}$  with  $k^*$ . We know from above proof of Theorem 2 that it is possible to choose  $k^*$  candidates such that the total dissatisfaction is less than  $n\delta^m(n)$ , and algorithm should in round  $k^*$  output a candidate set of  $k^*o(\log n)$  with dissatisfaction  $\leq n\delta^m(n)$ , which means that our algorithm should stop at round  $k^*$ , a contradiction.

By now, we end up constructing an algorithm that chooses at most  $k^* \cdot o(\log n)$  vertices that form a dominating set, which gives us an  $o(\log n)$  approximation algorithm for DOMINATINGSET problem. This concludes the proof.  $\square$

Since the dominating set problem is LOG-APX-hard [1], this concludes the proof.

For RELAXED-EGALITARIAN- $\bar{\delta}$ -CC, we know that by universally polynomially unbounded property, there exists an  $r'$  such that  $\delta^m(n) < \delta^m(r')$ . We apply the same reduction but with  $r$  replace by  $r'$ . We denote by  $l = \min_{C \in C_k} \max_{i \in [n]} \delta^m(i, C, 1)$  the minimum egalitarian dissatisfaction with  $k$  candidates. We claim that  $l \leq \delta^m(n)$  if and only if  $I$  is a YES instance of DOMINATINGSET.

( $\Leftarrow$ ) If there is a dominating set  $K$  of size  $k$ , then there exist  $k$  candidates such that the total egalitarian dissatisfaction is at most  $\delta^m(n)$  since for each agent one will contribute a total dissatisfaction of at most  $\delta^m(n)$ .

( $\Rightarrow$ ) If  $l \leq \delta^m(n)$ , we prove by contradiction that  $I$  is a YES instance of DOMINATINGSET. If not, then there is no dominating set of size  $k$ , thus there exists an  $i \in V$  such that for any vertex  $j \in K$ ,  $i$  and  $j$  are neither the same nor adjacent. This implies that for our election instance, by choosing any  $k$  candidates  $T$ , we know that there is an agent  $i$  such that all candidates in  $T$  get a dissatisfaction of at least  $\delta^m(r)$  (since they are all in  $C - C_{\Gamma^+(i)}$ ). Thus the egalitarian dissatisfaction for choosing any  $k$  candidates is at least  $\delta^m(r) > \delta^m(n)$ , which contradicts. This gives the reduction which implies this problem is NP-hard. The approximation preserving property directly follows from Lemma 2. This concludes the proof.  $\square$

*Remark 2.* Although the proof seems to rely on the condition that  $m \geq n$ , this is not necessary: one can choose appropriate  $n' = \Omega(n^c)$  for a constant  $c > 0$ , and make the reduction work. Since  $o(\log n') = o(\log n)$ , we can still conclude the LOG-APX-hardness.

*Remark 3.* This theorem also holds for the cases that the dissatisfaction families among agents are different: let  $\bar{\delta}_i$  be the dissatisfaction function family for agent  $i$ , we just need to replace the condition  $\delta^m(r) \geq n\delta^m(n)$  by  $\min_i \delta_i^m(r) \geq \sum_i \delta_i^m(n)$  (which can also be derived from Lemma 1), and the proof also applies.

## 5 APPROXIMATION ALGORITHMS FOR SIZE-RELAXED COMMITTEE SELECTION

In this section, we give upper bounds results for the size-relaxed committee selection problem. At a high level, all of our approximation algorithms involve the following two main steps: given an instance of committee selection, (Step 1) construct an instance of WEIGHTED-MAXIMUM- $k$ -COVERAGE, and (Step 2) run a greedy/LP rounding based algorithm on the WEIGHTED-MAXIMUM- $k$ -COVERAGE instance, and transform the solution to a solution for the committee selection instance. Throughout this section we will prove upper bounds for satisfaction based Chamberlin-Courant rules for convenience. The results for dissatisfaction functions follow from Theorem 1 and Remark 1.

### 5.1 Utilitarian Satisfaction

We first illustrate our approach through our algorithm for RELAXED-UTILITARIAN- $\bar{\alpha}$ -CC, where given an instance with election  $E = (\mathcal{N}, C, \bar{P})$ , we proceed in two steps:

Step 1. Construct an instance  $I = (U, w, F, k)$  of WEIGHTED-MAXIMUM- $k$ -COVERAGE by applying Algorithm 1 to  $E = (\mathcal{N}, C, \bar{P})$ , where (i) there is an element  $a_{i,j}$  in ground set  $U$ , for each agent  $i \in \mathcal{N}$ , and candidate  $j \in C$ , with weight  $\alpha^m(pos_i(j)) - \alpha^m(pos_i(j) + 1)$ , and (ii) there is a set  $F_j = \{a_{i,\hat{j}} : i \in \mathcal{N}, \hat{j} \in C, pos_i(j) > pos_i(\hat{j})\}$ , for every candidate  $j \in C$  in  $F$ . W.l.o.g. assume  $\alpha^m(m+1) = 0$ .

Step 2. Algorithm 2 greedily picks the set  $F_{j^*}$  with maximum marginal increase in the weight of covered elements for instance  $I$ , and adds candidate  $j^*$  to the committee.

Example 1 shows the key idea behind our approach: the utilitarian satisfaction of a committee equals to the total weight covered by the corresponding subsets for the WEIGHTED-MAXIMUM- $k$ -COVERAGE instance.

*Example 1.* Consider the following simple multi-winner selection instance  $E = (\mathcal{N}, \mathcal{C}, \vec{P})$  with Borda satisfaction. We let  $\mathcal{N} = \{1, 2, 3, 4\}$ ,  $\mathcal{C} = \{a, b, c, d\}$ . Our objective is to look for the satisfaction of  $k = 2$  candidates. For preference profile, we define the preference order for each  $i$  as follows:

- 1 :  $a > b > c > d$
- 2 :  $a > c > b > d$
- 3 :  $b > c > d > a$
- 4 :  $d > a > b > c$

It is not hard to see by choosing  $a$  and  $b$  one can get the maximum Borda satisfaction of  $3 + 3 + 3 + 2 = 11$ . Consider a WEIGHTED-MAXIMUM- $k$ -COVERAGE instance  $I = (U, w, F, 2)$ , with  $U = \{a_{i,j}\}_{4 \times 4}$ ,  $F = \{F_a, F_b, F_c, F_d\}$ ,  $w(\{a_{i,j}\}) = 1$  for any  $i$  and  $j < 4$ ,  $w(\{a_{i,j}\}) = 0$  for any  $i$  and  $j = 4$ .

$$\begin{aligned} F_a &= \{a_{1,1}, \dots, a_{1,4}, a_{2,1}, \dots, a_{2,4}, a_{3,4}, a_{4,2}, a_{4,3}, a_{4,4}\} \\ F_b &= \{a_{1,2}, a_{1,3}, a_{1,4}, a_{2,3}, a_{2,4}, a_{3,1}, \dots, a_{3,4}, a_{4,3}, a_{4,4}\} \\ F_c &= \{a_{1,3}, a_{1,4}, a_{2,2}, a_{2,3}, a_{2,4}, a_{3,2}, a_{3,3}, a_{3,4}, a_{4,4}\} \\ F_d &= \{a_{1,4}, a_{2,4}, a_{3,3}, a_{3,4}, a_{4,1}, a_{4,2}, a_{4,3}, a_{4,4}\} \end{aligned}$$

We can see that by choosing  $F_a$  and  $F_b$  we get the weighted-maximum-2-coverage, with cover size equal to 11. This is exactly maximum utilitarian satisfaction.  $\square$

---

**Algorithm 1** SC( $E, k$ ) Election to Cover.

---

**Input:** An election  $E = (\mathcal{N}, \mathcal{C}, \vec{P})$ , positive integer  $k$

**Output:** A weighted-maximum- $k$ -coverage  $I = (U, w, F, k)$

- 1: Let  $U \leftarrow \{a_{i,j}\}_{n \times m}$ ,  $F = \{F_1, F_2, \dots, F_m\}$
  - 2: **for**  $i = 1, \dots, n$  **do**
  - 3:   **for**  $l = 1, \dots, m$  **do**
  - 4:     Let  $w(\{a_{i,l}\}) = \alpha^m(l) - \alpha^m(l+1)$
  - 5: **for**  $j = 1, \dots, m$  **do**
  - 6:    $F_j \leftarrow \{a_{i,l} | i \in [n], l \in [m], l \geq \text{pos}_i(j)\}$
  - 7: **return**  $(U, w, F, k)$
- 

---

**Algorithm 2** RELAXED-UTILITARIAN- $\vec{\alpha}$ -CC

---

**Input:** An election  $E = (\mathcal{N}, \mathcal{C}, \vec{P})$ , positive integer  $k$

**Output:** A set of candidates  $K$

- 1:  $K = \emptyset, L = [m], V = \emptyset$
  - 2:  $(U, w, F, k) \leftarrow \text{SC}(E, k)$
  - 3:  $\gamma = \max_{i_1, i_2 \in [n], l_1, l_2 \in [m]} \frac{w(\{a_{i_1, l_1}\})}{w(\{a_{i_2, l_2}\})}$
  - 4: **for**  $l = 1, \dots, k \cdot (\log mn \gamma + 1)$  **do**
  - 5:   Pick  $j \in L$  s.t.  $w(F_j - V)$  is maximized
  - 6:    $V \leftarrow F_j \cup V, K \leftarrow K \cup j, L \leftarrow L - j$
  - 7: **return**  $K$
- 

Our main result in Theorem 3 proves a  $O(\log n)$  upper bound by applying Algorithm 1 and 2. The proof relies on Lemma 3, where we prove an  $O(k \log n)$  upper bound for size-relaxed version of WEIGHTED-MAXIMUM- $k$ -COVERAGE, under the restriction that the

ratio of any two weights is bounded. In the classical WEIGHTED-MAXIMUM- $k$ -COVERAGE problem, we are given a ground set  $U$ , whose elements are associated with weights, a family  $F$  of subsets of  $U$ , and a positive integer  $k$ , and we are asked to pick  $k$  members of  $F$  whose cover has the maximum total weight. In the size-relaxed version, we are asked to select at most  $\gamma \cdot k$  members of  $F$  such that the weight covered is at least that of the optimal set of  $k$  members. Although the size-relaxed version of WEIGHTED-MAXIMUM- $k$ -COVERAGE does not formally appear in any reference to the best of our knowledge, it can be derived from Lemma 3.14 of [16].

**Theorem 3.** For polynomially bounded satisfaction function family  $\vec{\alpha}$ , Algorithm 2 guarantees an approximation ratio of  $O(\log n)$  to RELAXED-UTILITARIAN- $\vec{\alpha}$ -CC in polynomial time.

The key step in the proof is Lemma 3, where we show that Algorithm 2 is a  $O(\log n)$ -approximation algorithm for the size-relaxed version of WEIGHTED-MAXIMUM- $k$ -COVERAGE, where given a WEIGHTED-MAXIMUM- $k$ -COVERAGE instance  $I = (U, w, F, k)$ , in the size-relaxed version we are asked to find a subset  $\hat{J}$  of  $F$  such that the total weight of elements covered in  $\hat{J}$  is at least the total weight covered by the optimal solution  $J^*$  to WEIGHTED-MAXIMUM- $k$ -COVERAGE.

**Lemma 3.** Let  $I^*$  be the total weight of the optimal WEIGHTED-MAXIMUM- $k$ -COVERAGE, if there exists a polynomial  $p(\cdot)$  such that:

$$w_{\max} := \max_{i,j \in [n]} \frac{w(i)}{w(j)} \in O(p(n)),$$

then there is a polynomial time algorithm that selects  $O(k \log n)$  elements with coverage of at least  $I^*$ .

We claim that the greedy algorithm, which in each round, picks the subset with highest incremental weight (breaking ties arbitrarily), has this property.

Let  $L_j$  be the set that by running greedy algorithm for  $j$  rounds, the elements covered in ground set. Let  $M_k$  be the set that by choosing  $k$  subsets in  $\mathcal{F}$ , the elements covered in ground set with maximum weight (weighted maximum  $k$  coverage). We extend the definition of weight to subset of  $[n]$  such that for any subset  $S \subseteq [n]$ ,  $w(S) = \sum_{i \in S} w(i)$ . We show that  $w(L_{k \log(n \cdot w_{\max}) + 1}) \geq w(M_k)$ .

Let  $OPT_k = w(M_k)$ , and  $a_i$  and  $b_i$  be the total weight of elements covered in  $M_k$  and  $\bar{M}_k = U - M_k$  respectively in round  $i$  by greedy algorithm. We know from Lemma 3.14 of [16], that for all  $i \geq 1$ ,

$$\sum_{j=1}^i (a_j + b_j) \geq OPT_k - OPT_k \left(1 - \frac{1}{k}\right)^i. \quad (1)$$

Now we start to prove Lemma 3.

**PROOF.** We know from (1) that by running the greedy algorithm for  $k \log(n \cdot w_{\max})$  rounds, the number of elements covered by greedy algorithm is at least

$$\begin{aligned} &OPT_k - OPT_k \left(1 - \frac{1}{k}\right)^{k \log(n \cdot w_{\max})} \\ &= OPT_k \left(1 - \frac{1}{n \cdot w_{\max}}\right). \end{aligned}$$

Since for any  $i \in [n]$ ,  $OPT_k \leq n \cdot w_{\max} \cdot w(i)$ , by picking another set which contribute an increment of at least 1 element, greedy algorithm covers a total weight of at least  $OPT_k$ . Thus  $w(L_k \log(n \cdot w_{\max}) + 1) \geq w(M_k)$ . This concludes the proof of this lemma.  $\square$

We are now ready to prove Theorem 3.

**PROOF OF THEOREM 3.** It is not hard to see that the weight of coverage by any  $k$  subsets in  $I$  exactly equals the satisfaction of the corresponding size  $k$  committee. In particular, the optimal  $k$  subsets with maximum weighted coverage corresponds to the optimal size  $k$  committee. By Lemma 3, Algorithm 2 which greedily chooses  $F_j$  that maximizes incremental weight (equals to size in this case) will output a set  $K$  such that  $|K| = O(k \log mn)$ , while covering at least the size of the maximum coverage of  $k$  subsets chosen from  $F$ . Since we are focusing on polynomially bounded satisfaction function family, thus  $m$  and  $\gamma$  are in  $Poly(n)$ , we conclude the proof.  $\square$

Theorem 3 also works for the case that satisfaction between agents are different, but we omit the proof in the interest of space.

**Corollary 1.** Let  $\vec{\alpha}_i$  be the satisfaction function family for agent  $i$ .  $\forall i$ ,  $\vec{\alpha}_i$  satisfy polynomial gap among weights property. By replacing weights with  $w(\{a_{i,l}\}) = \alpha_i^m(l) - \alpha_i^m(l+1)$  in Algorithm 2, the modified algorithm guarantees an approximation ratio of  $O(\log n)$  to RELAXED-UTILITARIAN- $\vec{\alpha}$ -CC in polynomial time.

*Remark 4.* It turns out this algorithm does not work with the family of unbounded satisfaction functions. The following counterexample shows why this happens.

*Example 2.* Let  $m = O(n^2)$ , let  $\alpha^m(i) = \sum_{j=i}^n \frac{1}{n^{n-j}}$  for  $i \leq n$ , and  $\alpha^m(i) = 0$  for  $i > n$ . Candidate 1 appears in each vote at rank  $n$ , candidate 2 appears in each vote but the last one at rank  $n-1$ , with rank in the last voter behind  $n$ , candidate 3 appears in each vote but the first two, at rank  $n-2$ , with rank in the first two voters behind  $n, \dots$ , candidate  $n/4$  appears in each vote but the first (or maybe last, depending on if  $n/4$  is odd)  $n/4$  voters at rank  $n - n/4$ . Candidate Tom appears in the first half of voters at rank  $n/2$ , and the second half of voters at rank behind  $n$ . Candidate Jerry appears in the first half of voter at rank behind  $n$ , and the second half of voters at rank  $n/2$ . For the other candidates, it ranks on one voter less than  $n$ , and on the other voters behind  $n$ .

One can check that in this example, the greedy algorithm will start choosing from candidate 1 to  $n/4$  in the first  $n/4$  rounds. Still, they end up having less satisfaction than directly choosing Tom and Jerry. Thus the greedy algorithm does not have a satisfaction guarantee of  $O(\log n)$  here: the gap is already  $\Omega(n)$ . It is clear to see that by choosing at most  $n$  candidates, one can achieve optimal satisfaction.  $\square$

## 5.2 Egalitarian Satisfaction

Here we show upper bound results for *any* satisfaction functions. For the egalitarian objective, we design similar algorithms but based on SETCOVER, thus without weights. We note that, SETCOVER is a special case of WEIGHTED-MAXIMUM-K-COVERAGE where all weights are set to be 1, and  $k$  is set to be the value of optimal  $k^*$  that covers the ground set.

---

### Algorithm 3 RELAXED-EGALITARIAN- $\vec{\alpha}$ -CC

---

**Input:** An election  $E = (\mathcal{N}, C, \vec{P})$ , positive integer  $k$

**Output:** A set of candidates  $K$

```

1:  $K = \emptyset, L = [m], V = \emptyset$ 
2:  $(U, F, k) \leftarrow SC(E, k)$ 
3: for  $t = 1, \dots, m$  do
4:    $S_t \leftarrow \{a_{i,k} \mid k \in [m], k > m - t\}, K_t = \emptyset$ 
5:   for  $l = 1, \dots, k(\log mn + 1)$  do
6:     Pick  $j \in L$  s.t.  $|(F_j \cup (V \cap S_t)) - (V \cap S_t)|$  is maximized
7:      $V \leftarrow F_j \cup V, K_t \leftarrow K_t \cup j, L \leftarrow L - j$ 
8:   if  $S_t \subseteq V$  then
9:      $K \leftarrow K_t$ 
10: return  $K$ 

```

---

The high level idea of this algorithm is this: we enumerate over all possible egalitarian satisfaction (polynomial many), each egalitarian satisfaction  $l_t$  corresponds to a subset  $S_t$  of ground set, and we run the traditional greedy algorithm for  $O(k \log n)$  rounds with the target to cover  $S_t$ . If in round  $t$ , the greedy algorithm succeed covering  $S_t$ , then we update chosen candidate set  $K$ . We prove in Theorem 4 that this algorithm is a polynomial time algorithm that gives  $O(\log n)$  approximation.

**Theorem 4.** Algorithm 3 guarantees an approximation ratio of  $O(\log n)$  to RELAXED-EGALITARIAN- $\vec{\alpha}$ -CC in polynomial time.

**PROOF.** Let  $\alpha^m(l) = \max_{C \in C_k} \min_{i \in [n]} \alpha^m(i_{C,1})$ , if we can choose  $k$  subsets in  $F$  that cover all elements in  $\{a_{i,j} \mid j > m - l\}$ , then we can get an egalitarian satisfaction of  $\alpha^m(l)$ .

Recall that our algorithm enumerates all possible target satisfaction. We first claim that when the algorithm moves to the step checking egalitarian satisfaction of  $\alpha^m(l)$ , running the greedy algorithm for  $k(\log mn + 1)$  times  $S_l$  will be covered. This is because there exists  $k$  subsets that covers  $S_l$ , so  $(S_l, F, k)$  is a YES instance of SETCOVER. Lemma 3 tells us the above greedy algorithm will output a set cover within  $k(\log mn + 1)$  rounds. So the above algorithm will update  $K$  with the approximation guarantee and satisfaction guarantee. Later updates in this algorithm will only give egalitarian satisfaction higher than  $\alpha^m(l)$ . Thus, the candidate set outputted by Algorithm 3 has at least optimal egalitarian satisfaction by  $k$  candidates, with  $O(\log mn) = O(\log n)$  approximation ratio. The algorithm runs in  $m \cdot k \cdot (\log mn + 1)$  steps, as  $m = poly(n)$  this algorithm runs in polynomial time. This concludes the proof.  $\square$

*t-Approval.* Algorithm 3 also works for the class of  $t$ -approval satisfaction functions. As mentioned in Remark 3, if  $t = \Omega(Poly(n))$ , then size-relaxed committee problem is LOG-APX-hard. However, in practice,  $t$  is usually very small, say, a constant. We prove in Theorem 5 that there is a  $t$  approximation algorithm for  $t$ -approval satisfaction.

**Theorem 5.** Algorithm 4 guarantees an approximation ratio of  $O(t)$  to RELAXED-EGALITARIAN- $\vec{\alpha}_t$ -CC in polynomial time.

First, we give some intuition behind our algorithm and the proof. Notice that for the  $t$ -approval satisfaction, there are  $t$  candidates with satisfaction of 1 for each voter. Therefore, the corresponding SETCOVER instances are restricted to the case where each element

in  $U$  appears at most  $t$  times in the subsets of  $\mathcal{F}$ . [17] provides a  $t$ -approximation algorithm for this case using LP rounding, which we refer to as LOWFREQ, for low frequency set cover. Also, in the egalitarian case, the optimal satisfaction can only be 0 or 1, which we exploit in our algorithm.

**PROOF.** To solve this, we only need to solve the following special instance of set cover:  $U = [n]$ ,  $F = \{F_1, F_2, \dots, F_m\}$  where each  $F_l = \{i | \text{pos}_i(l) \leq t\}$ . By definition, to get a satisfaction of 1 on a specific agent  $i$ , we need to select a candidate  $j$  with  $\text{pos}_i(j) \leq t$ . This is by construction equivalent to cover  $i \in U$ . Notice that in the egalitarian case, optimal satisfaction can only be 0 or 1. For the former case, we return  $\emptyset$ . For the latter case, we need to cover all  $i \in U$ . Thus it is equivalent to cover all elements in  $U$ . One can directly apply the set cover algorithm with a frequency of  $t$  as in Algorithm 4.  $\square$

---

**Algorithm 4** RELAXED-EGALITARIAN- $\vec{\alpha}_t$ -CC
 

---

**Input:** An election  $E = (N, C, \vec{P})$ , positive integers  $k, t$

**Output:** A set of candidates  $K$

```

1: Let  $U \leftarrow [n]$ ,  $F = \{F_1, F_2, \dots, F_m\}$ 
2: for  $l = 1, \dots, m$  do
3:    $F_l \leftarrow \{i | i \in N, \text{pos}_i(l) \leq t\}$ 
4:  $K \leftarrow \text{LOWFREQ}(U, F, k)$ 
5: if  $|K| < k \cdot t$  then
6:   return  $K$ 
7: else
8:   return  $\emptyset$ 

```

---

By combining techniques from Theorem 4 and Theorem 5, there is an  $O(t)$  approximation algorithm for any  $\vec{\alpha}$  which assigns a positive (dis)satisfaction to at most  $t$  positions, i.e. for any  $\vec{\alpha}$ , where  $\alpha^m(l) = 0$  for any  $l > t$ : just enumerate possible satisfaction as Algorithm 3 does, and cover the corresponding target set as Algorithm 4 does.

**Corollary 2.** For any  $\vec{\alpha}$  which assigns a positive (dis)satisfaction to at most  $t$  positions, i.e. for any  $\vec{\alpha}$ , where  $\alpha^m(l) = 0$  for any  $l > t$ , there is an algorithm that guarantees an approximation ratio of  $O(\log n)$  to RELAXED-EGALITARIAN- $\vec{\alpha}$ -CC in polynomial time.

Also, notice that when  $t = 2$ , the size-relaxed problem exactly becomes VERTEXCOVER[18]. Note that the VERTEXCOVER problem is APX-hard to solve [9], thus, for  $t$ -approval our size-relaxed objective is APX-hard.

**Corollary 3.** For  $t$ -approval (dis)satisfaction function family, it is APX-hard to solve

- RELAXED-UTILITARIAN- $\vec{\alpha}^t$ -CC,  
 - RELAXED-EGALITARIAN- $\vec{\alpha}^t$ -CC.

## 6 DISCUSSION

Throughout the paper, we assume that  $m = O(\text{Poly}(n))$ , which is often the case in practice. Here we discuss the theoretical reason we make this assumption. First, if  $m$  is too large, then our algorithm has  $O(\log(mn))$  approximation, which can be greater than  $n$ . In this case, we should choose  $n$  candidates that rank top of each agent

as the committee. Second, our approach of proving lower bounds for approximating CC rules through a lower bound preserving reduction from a hard to approximate problem does not provide an immediate way forward to prove lower bounds based on both  $m$  and  $n$ , because we are not able to identify a suitable known problem.

In our paper, we tackle the size-relaxed committee selection problem by classifying (dis)satisfaction function families, and study each cases. This classification makes it convenient for us to derive clean results. But still, it is not clear if there is an approximation algorithm for polynomially unbounded satisfaction function with relatively good guarantee. A similar question on size-relaxed version of weighted-maximum- $k$ -coverage which is fundamental, is also open to the best of our knowledge. We believe new insights are required.

## ACKNOWLEDGMENTS

SS was supported by NSF #1453542 and #1716333 when the work was done.

## REFERENCES

- [1] Noga Alon, Dana Moshkovitz, and Shmuel Safra. 2006. Algorithmic construction of sets for  $k$ -restrictions. *ACM Transactions on Algorithms (TALG)* 2, 2 (2006), 153–177.
- [2] Haris Aziz, Piotr Faliszewski, Bernard Grofman, Arkadii Slinko, and Nimrod Talmon. 2018. Egalitarian committee scoring rules. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. AAAI Press, 56–62.
- [3] Nadja Betzler, Arkadii Slinko, and Johannes Uhlmann. 2013. On the Computation of Fully Proportional Representation. *Journal of Artificial Intelligence Research* 47, 1 (2013), 475–519.
- [4] Hedyeh Beyhaghi and S Matthew Weinberg. 2018. Optimal (and Benchmark-Optimal) Competition Complexity for Additive Buyers over Independent Items. *arXiv preprint arXiv:1812.01794* (2018).
- [5] Jeremy Bulow and Paul Klemperer. 1996. Auctions Versus Negotiations. *The American Economic Review* (1996), 180–194.
- [6] John R. Chamberlin and Paul N. Courant. 1983. Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule. *The American Political Science Review* 77, 3 (1983), 718–733.
- [7] EG Coffman Jr, MR Garey, and DS Johnson. 1996. Approximation algorithms for bin packing: a survey. In *Approximation algorithms for NP-hard problems*. PWS Publishing Co., 46–93.
- [8] Marquis de Condorcet. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L'Imprimerie Royale.
- [9] Irit Dinur and Samuel Safra. 2005. On the hardness of approximating minimum vertex cover. *Annals of mathematics* (2005), 439–485.
- [10] Alon Eden, Michal Feldman, Ophir Friedler, Inbal Talgam-Cohen, and S Matthew Weinberg. 2017. The Competition Complexity of Auctions: A Bulow-Klemperer Result for Multi-Dimensional Bidders. In *Proceedings of the 2017 ACM Conference on Economics and Computation*. ACM, 343–343.
- [11] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. 2017. Properties of multiwinner voting rules. *Social Choice and Welfare* 48, 3 (2017), 599–632.
- [12] Piotr Faliszewski, Arkadii Slinko, and Nimrod Talmon. 2017. The complexity of multiwinner voting rules with variable number of winners. *arXiv preprint arXiv:1711.06641* (2017).
- [13] Michal Feldman, Ophir Friedler, and Aviad Rubinstein. 2018. 99% revenue via enhanced competition. In *Proceedings of the 2018 ACM Conference on Economics and Computation*. ACM, 443–460.
- [14] Peter C Fishburn and Aleksandar Pecek. 2004. Approval voting for committees: Threshold approaches. *Techn Rept* (2004).
- [15] William V Gehrlein. 1985. The Condorcet criterion and committee selection. *Mathematical Social Sciences* 10, 3 (1985), 199–209.
- [16] Dorit S Hochba. 1997. Approximation algorithms for NP-hard problems. *ACM Sigact News* 28, 2 (1997), 40–52.
- [17] Dorit S Hochbaum. 1982. Approximation algorithms for the set covering and vertex cover problems. *SIAM Journal on computing* 11, 3 (1982), 555–556.
- [18] Richard M Karp. 1972. Reducibility among combinatorial problems. In *Complexity of computer computations*. Springer, 85–103.
- [19] D Marc Kilgour. 2016. Approval elections with a variable number of winners. *Theory and Decision* 81, 2 (2016), 199–211.
- [20] Maciej Kocot, Anna Kolonko, Edith Elkind, Piotr Faliszewski, and Nimrod Talmon. 2019. Multigoal committee selection. In *Proceedings of the 28th International Joint*

- Conference on Artificial Intelligence*. AAAI Press, 385–391.
- [21] Jyh-Han Lin and Jeffrey Scott Vitter. 1992.  $\epsilon$ -Approximations with minimum packing constraint violation. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*. ACM, 771–782.
- [22] Tyler Lu and Craig Boutilier. 2011. Budgeted Social Choice: From Consensus to Personalized Decision Making. In *Twenty-Second International Joint Conference on Artificial Intelligence*. AAAI Press.
- [23] Ariel D Procaccia, Jeffrey S Rosenschein, and Aviv Zohar. 2008. On the complexity of achieving proportional representation. *Social Choice and Welfare* 30, 3 (2008), 353–362.
- [24] Tim Roughgarden, Inbal Talgam-Cohen, and Qiqi Yan. 2015. *Robust auctions for revenue via enhanced competition*. Technical Report. Working paper.
- [25] Shreyas Sekar, Sujoy Sikdar, and Lirong Xia. 2017. Condorcet consistent bundling with social choice. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 33–41.
- [26] Piotr Skowron, Piotr Faliszewski, and Jerome Lang. 2016. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artificial Intelligence* 241 (2016), 191–216.
- [27] Piotr Skowron, Piotr Faliszewski, and Arkadii Slinko. 2015. Achieving fully proportional representation: Approximability results. *Artificial Intelligence* 222 (2015), 67–103.
- [28] Piotr Krzysztof Skowron and Piotr Faliszewski. 2015. Fully proportional representation with approval ballots: approximating the maxcover problem with bounded frequencies in FPT time. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.