Conference Program Design with Single-Peaked and Single-Crossing Preferences

Dimitris Fotakis[1], Laurent Gourvès[2], and Jérôme Monnot[2]

[1] School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Athens, Greece
[2] Université Paris-Dauphine, PSL Research University, CNRS UMR [7243], LAMSADE, 75016 Paris, France
fotakis@cs.ntua.gr , laurent.gourves@dauphine.fr , jerome.monnot@dauphine.fr

Abstract. We consider the *Conference Program Design* (CPD) problem, a multi-round generalization of (the maximization versions of) q-Facility Location and the Chamberlin-Courant multi-winner election, introduced by (Caragiannis, Gourvès and Monnot, IJCAI 2016). CPD asks for the selection of kq different items and their assignment to k disjoint sets of size q each. The agents derive utility only from their best item in each set, and we want to maximize the total utility derived by all agents from all sets. Given that CPD is **NP**-hard for general utilities, we focus on utility functions that are either single-peaked or single-crossing. For general single-peaked utilities, we show that CPD is solvable in polynomial time and that Percentile Mechanisms are truthful. If the agent utilities are determined by the distances of the agents to the items in the unit interval, we show that a Percentile Mechanism achieves an approximation ratio 1/3, if $q = 1$, and at least $(2q - 3)/(2q - 1)$, for any $q \geq 2$. On the negative side, we show that a generalization of CPD, where some items must be assigned to specific sets in the solution, is **NP**-hard for dichotomous single-peaked preferences. For single-crossing preferences, we present a dynamic programming exact algorithm that runs in polynomial time if k is constant.

1 Introduction

Many problems in Social Choice deal with selecting q items (or candidates), from a given set of m items, based on the preferences of n agents. In more than a few, each agent derives utility from his best item in the solution and the objective is to maximize the total utility of the agents, sometimes also considering incentive compatibility constraints.

An instance of this general setting is the classical *q-Facility Location* problem, where we want to place q facilities in a metric space, based on the locations suggested by n agents. Each agent uses his nearest facility in the solution and the objective is to minimize the total distance of the agents to their facilities. Facility Location is a fundamental optimization problem and has played a key role in the field of Approximation Algorithms (see e.g., [29]). In Social Choice, the relevant literature mostly focuses on strategic agents with single-peaked preferences over the possible facility locations. The goal is to characterize the class of truthful mechanisms and to determine the best approximation ratio achievable by truthful mechanisms when the agent preferences are determined by agent distances on the real line (see e.g., [15, 22, 24, 28] and the references therein).

A different specimen of the same setting appears in the context of multi-winner elections. In the model introduced by Chamberlin and Courant [7], we want to form a committee by selecting q representatives, from a set of m candidates, so as to minimize the committee's "misrepresentation" with respect to a set of n agents. Similarly to Facility Location, each agent is associated with the committee member that represents him best, and we want to minimize the total "misrepresentation cost" of the agents. The winner determination problem for the multi-winner election of Chamberlin-Courant has received significant attention recently, with **NP**-hardness results and approximation algorithms for general agent preferences and polynomial-time algorithms for restricted preferences, such as single-peaked or single-crossing (see e.g., [2, 6, 23, 27] and the references therein).

In this work, we study the *Conference Program Design* problem, which was recently introduced by Caragiannis *et al.* [6] and can be regarded as a generalization of (the maximization versions of) q-Facility Location and the Chamberlin-Courant election. An instance of Conference Program Design, or CPD in short, consists of a set of m items $X = \{x_1, \dots, x_m\}$, a set of n agents $L = \{1, \dots, n\}$, each with a utility function $u_\ell : X \to \mathbb{R}_{\geq 0}$, and two positive integers k and q. A feasible solution $\mathcal{S} = \{S_1, \dots, S_k\}$ is a collection of k pairwise disjoint subsets of X (or *slots*) such that each slot S_i contains at most q items. The agents derive utility only from their most preferred item in each slot and have additive utilities for different slots. Hence, the utility of an agent ℓ for a solution $\mathcal{S} = \{S_1, \dots, S_k\}$ is $u_\ell(\mathcal{S}) = \sum_{i=1}^{k} \max_{x \in S_i} u_\ell(x)$. The social objective is to maximize the total utility of all agents, which is $U(\mathcal{S}) = \sum_{\ell \in L} u_\ell(\mathcal{S})$ for any given solution $\mathcal{S} = \{S_1, \dots, S_k\}$. We should underline that although a larger total utility may be achieved by assigning some items to multiple slots, we require that the slots $S_1, \dots, S_k$ should be pairwise disjoint.

Example 1. We consider 5 items $\{x_1, x_2, x_3, x_4, x_5\}$, 3 agents and $k = q = 2$. The utility functions of the agents are $u_1 = (4, 3, 5, 1, 2)$, $u_2 = (1, 2, 3, 9, 2)$ and $u_3 = (6, 1, 4, 0, 7)$ (the i-th coordinate in u_j denotes the utility of agent j for item x_i). The total utility of the solution $\mathcal{S} = (\{x_1, x_2\}, \{x_3, x_4\})$ is $U(\mathcal{S}) = (u_1(x_1) + u_1(x_3)) + (u_2(x_2) + u_2(x_4)) + (u_3(x_1) + u_3(x_3)) = 30$. ♦

The name of Conference Program Design is motivated from the possibility of regarding each item as a conference talk. The conference has q parallel sessions and k time slots. During each slot S_i, at most q talks are given and each agent can attend only one of them. We assume that every agent attends his most preferred talk, i.e., the talk that maximizes his utility, in each slot. In addition to the natural connection with the design of multi-session conference programs, CPD should be regarded as an abstraction of multi-round multi-winner elections, where the set of winners in different rounds must be disjoint, each agent is represented by his most preferred winner in each round, and the utility functions of the agents are additive with respect to their representatives in different rounds (see also [6] for further discussion on the motivation of CPD).

Previous Work. Apart from the fact that it is a maximization problem, CPD incorporates both q-Facility Location and the Chamberlin-Courant election (for $k = 1$): each item is a facility/candidate and the utilities are the opposite of the distance/misrepresentation costs. Since the multi-winner election of Chamberlin-Courant and q-Facility Location [16, 23] are known to be **NP**-hard for general cost functions, CPD is also **NP**-hard for general utilities. Interestingly, Caragiannis *et al.* [6] proved that CPD remains **NP**-hard (and inapproximable up to given constant factors) in the special case where agent utilities are either 0 or 1 (a.k.a. *uniformly dichotomous* preferences), all items fit in the solution, i.e., $m = kq$, and either $k = 2$ or $q = 3$. The only case where CPD is known to be polynomially solvable is for $q = 2$, by a reduction to maximum matching. Based on a natural Integer Linear Programming formulation (see also Section A.1), Caragiannis *et al.* [6] obtained polynomial-time approximation algorithms for CPD with general utilities, with ratios $1 - 1/e$, if q is a constant, and $1/e - 1/e^2$, if q is part of the input.

On the other hand, many positive results are known for natural (and practically interesting) special cases of q-Facility Location and of the Chamberlin-Courant election, especially for the line metric and for single-peaked or single-crossing preferences. Specifically, q-Facility Location on the line is polynomially solvable, by a simple dynamic programming algorithm. Recently, using Linear Programming techniques, Hajiaghayi *et al.* [18] extended this result to the *fault tolerant* version of q-Facility Location, where each agent must connect to k different facilities. Moreover, there has been a significant recent interest in the approximability of q-Facility Location on the line by truthful mechanisms. As for deterministic mechanisms, the Median Mechanism is optimal for $q = 1$ [22, 24], the 2-Extremes Mechanism achieves a best possible approximation ratio of $n - 2$ for $q = 2$ [15, 24], and the Percentile Mechanisms comprise the only known general class of truthful deterministic mechanisms [28] for all $q \geq 2$, but their worst-case approximation ratio cannot be bounded in terms of n and q [15] (all these mechanisms are actually known to be group strategyproof). Moreover, there are randomized truthful mechanisms for q-Facility Location on the line with a constant approximation ratio for $q \in \{2, n-1\}$ and an approximation ratio of n for any $q \in \{3, \dots, n-2\}$ [14]. However, all these results on the approximability of Facility Location by truthful mechanisms are about cost minimization and assume that a facility can be placed at any point on the real line. So they are not directly relevant for CPD, where we want to maximize the total utility and the item locations are restricted by the input. In a recent work, Feldman *et al.* [13] characterized the approximability of 1-Facility Location on the line metric (by truthful or non-truthful, deterministic or randomized mechanisms) when the potential facility locations are restricted by the input.

Similarly to the line metric for Facility Location, for the election of Chamberlin-Courant, it is reasonable to assume that the agent preferences over the candidates is consistent with a placement of the candidates on a societal axis. In fact, the line metric is a special case of two popular concepts in structured preferences in Social Choice literature, namely *single-peaked* and *single-crossing* preferences [4, 21, 25] (see Section 2 for their formal definition). Recent work presents polynomial-time exact algorithms for the winner determination problem of the Chamberlin-Courant election when the agent preferences are either single-peaked [2] or single-crossing [27].

Contribution and Techniques. Motivated by the many interesting positive results for q-Facility Location and for the Chamberlin-Courant election when the agent preferences either are determined by the line metric or are single-peaked or single-crossing, we investigate here the algorithmic properties of the Conference Program Design problem for such preferences. We give an almost complete picture for CPD with single-peaked preferences and show that CPD with single-crossing preferences is polynomially solvable if the number of slots k is constant.

An interesting observation is that for single-peaked utility functions, the best k items of any agent occupy consecutive positions on the societal axis (Proposition 1). Therefore, for any set M of items, $|M| \leq kq$, a simple greedy assignment of the items to slots ensures that every agent can derive utility from his best k papers in M. This observation allows us to focus on the item selection aspect of CPD for single-peaked preferences (recall that finding an optimal assignment of kq items to slots is **NP**-hard for uniformly dichotomous utilities [6]). Combining this observation with a generalization of the Linear Programming approach of [18], in Section 3, we show that CPD can be solved in polynomial-time for general single-peaked utility functions (Theorem 1).

In sections 4 and 5, we study the approximability of CPD with single-peaked preferences by truthful mechanisms. To achieve truthfulness, we exploit the idea of Percentile Mechanisms, which are known to be group strategyproof for agents with single-peaked preferences [28, Theorem 1]. To optimize the approximation guarantees, we apply Percentile Mechanisms to the set of all tuples consisting of k consecutive items on the societal axis. Interestingly, we show that the extension of any single-peaked utility function on items to a utility function on tuples of k consecutive items is also single-peaked (Lemma 1). Consequently, this variant of Percentile Mechanisms is truthful (Theorem 2, we can also show that it is group strategyproof).

As for the approximation ratio of Percentile Mechanisms, we consider the special case of linear preferences where the items and the agents lie in the unit interval $[0, 1]$ and the utility of an agent ℓ located at v_ℓ for an item j located at x_j is $u_\ell(x_j) = 1 - |v_\ell - x_j|$. The restriction to the unit interval is wlog., since all our results hold for any interval length B, provided that the utility functions are $u_\ell(x_j) = B - |x_j - v_\ell|$. We first observe that if $k = q = 1$, the optimal solution is not truthful and any deterministic truthful mechanism must have an approximation ratio at most $5/7$. For $q = 1$ and any $k \geq 1$, we show that the approximation ratio of the $1/2$-Percentile Mechanism is $1/3$ (Lemma 3 and Lemma 4). For any $q \geq 2$ and $k \geq 1$, we show that if the number of agents is a multiple of q, the approximation ratio of the $(\frac{1}{2q}, \frac{3}{2q}, \dots, \frac{2q-1}{2q})$-Percentile Mechanism is at least $(2q-3)/(2q-1)$ (Theorem 3) and at most $(2q-1)/(2q-1/q)$. Interestingly, the approximation ratio tends to 1, as q

increases. If the number of agents n is not a multiple of q, we obtain a slightly weaker approximation ratio of $(2q-3)/(2q-1) - O(q/n)$ (Theorem 4).

To the best of our knowledge, this is the first analysis of the approximation ratio of Percentile Mechanisms for linear preferences (note their approximation ratio for cost minimization problems is unbounded). As for the proof technique, for the general case where $q \geq 2$, we introduce the notion of the *width* of a subset of agents, which allows us to bound the approximation ratio for the entire set of agents by analyzing independently the approximation ratio of non-overlapping groups with n/q agents each (see Lemma 5 and the proof of Theorem 3). The proof technique can be applied to any quasilinear nonnegative utility function of the form $u_\ell(x_j) = 1 - f(|x_j - v_\ell|)$, where $f : [0,1] \to [0,1]$ is any nondecreasing function of the distance. The approximation ratio can be derived using the same approach and depends on the steepness of f.

Nevertheless, single-peaked preferences are not enough to make CPD polynomially solvable if some items need to be assigned to specific slots. Using a reduction from PreColoring Extension, which is known to be **NP**-complete in unit interval graphs [20], we show that this generalization of CPD is **NP**-hard if the agent utilities are single-peaked and either 0 or 1 for each item (a.k.a. *dichotomous single-peaked* preferences, see Theorem 5, in Section 6).

Finally, in Section 7, we extend the dynamic programming approach applied in [27] to the Chamberlin-Courant election with single-crossing preferences and show that CPD with single-crossing preferences can be solved in $O(m(nq)^{k+1})$ time (Theorem 6). An interesting open question is whether CPD with single-crossing preferences is polynomially solvable if k is part of the input.

2 Notation and Preliminaries

CPD is formally introduced in Section 1. In this section, we introduce some additional notation and terminology and some basic facts used in this work. For any integer $p \geq 1$, we denote $[p] = \{1, \ldots, p\}$. We write $x \succ_\ell x'$ to express that an agent ℓ prefers item x to item x', which happens iff $u_\ell(x) > u_\ell(x')$. In such cases, we sometimes write that $\succ_\ell$ is the preference order induced by the utility function u_ℓ. In case of ties (in the utility functions or in the selections made by an algorithm), we always break them in an arbitrary fixed deterministic way.

The best paper of an agent ℓ in a set of items $Y \subseteq X$ is Y's most valuable item to agent ℓ, i.e., $\arg\max_{y \in Y} u_\ell(y)$. We define the second, ..., the k-th best item of an agent ℓ in Y similarly. Given a set of items $Y \subseteq X$, and assuming the case where $k = 1$, i.e., where each agent uses a single item, we let $u_\ell(Y) = \max_{y \in Y}\{u_\ell(y)\}$ denote the maximum utility of an agent ℓ for his best item in Y, and let $U(Y) = \sum_{\ell=1}^n u_\ell(Y)$ denote the total utility derived by the agents from Y. Similarly, we let $U(x) = \sum_{\ell=1}^n u_\ell(x)$ denote the total utility derived by the agents from an item $x \in X$.

Conference Program Design with Item Preselection. In Section 6, we consider a natural generalization of CPD, where a specified subset of items $X' \subseteq X$ must appear in the final solution and the assignment of the items in X' to slots is fully specified by the input. We call this variant *Conference Program Design with Item Preselection*, or PRE-CPD, in short.

More formally, in addition to the input of CPD, the input of PRE-CPD includes a subset $X' \subseteq X$ of items and a mapping $g : X' \to [k]$. A solution $\mathcal{S}$ is a collection of k disjoint subsets $S_1, \ldots, S_k$ of X, such that each S_i contains at most q items and $g^{-1}(i) = \{x \in X' : g(x) = i\} \subseteq S_i$. In particular, we assume $|g^{-1}(i)| \leq q$. Thus, CPD corresponds to PRE-CPD with $X' = \emptyset$.

Approximation Ratio. An algorithm achieves an *approximation ratio* of $\rho \in (0,1]$, if for any instance I of CPD, the solution $\mathcal{S}$ computed by the algorithm satisfies $U(\mathcal{S}) \geq \rho\, U(\mathcal{S}^*)$, where $\mathcal{S}^*$ denotes the optimal solution to instance I.

Truthfulness. A mechanism A for CPD is *truthful* (or *strategyproof*) if no agent can increase his utility from the outcome of A by misreporting her utility function. Formally, for any pair of instances I and I' that differ in the utility function of any single agent ℓ, with u_ℓ denoting ℓ's utility in I, we have that $u_\ell(\mathcal{S}) \geq u_\ell(\mathcal{S}')$, where $\mathcal{S}$ (resp. $\mathcal{S}'$) is the solution of A on instance I (resp. I').

A mechanism A for CPD is *group strategyproof* if no coalition of agents can simultaneously increase their utility from the outcome of A by misreporting their utility functions. Formally, for any pair of instances I and I' that differ in the utility functions of any nonempty subset $L' \subseteq L$ of agents, with $u_1, \dots, u_n$ denoting the utility functions in I, there exists an agent $\ell \in L'$ so that $u_\ell(\mathcal{S}) \geq u_\ell(\mathcal{S}')$, where $\mathcal{S}$ (resp. $\mathcal{S}'$) denotes the solution computed by A on instance I (resp. I').

Single-Peaked Preferences. A societal axis is a linear order $\sqsupset$ over X. We say that an agent's preference order $\succ$ is *consistent* with $\sqsupset$, if for each three items $x_a, x_b, x_c \in X$, we have that $((x_a \sqsupset x_b \sqsupset x_c) \vee (x_c \sqsupset x_b \sqsupset x_a)) \Rightarrow (x_a \succ x_b \Rightarrow x_b \succ x_c)$. We say that a utility function u_ℓ of an agent $\ell \in L$ is *single-peaked* wrt axis $\sqsupset$ [4], if the preference order $\succ_\ell$ induced by u_ℓ is consistent with $\sqsupset$. An instance of CPD is single-peaked (or has single-peaked utilities or preferences) with respect to axis $\sqsupset$, if the utility functions u_ℓ of all agents $\ell \in L$ are single-peaked wrt axis $\sqsupset$. An instance of CPD is single-peaked if it is single-peaked with respect to some societal axis.

Note that it is possible to check in polynomial time whether a set of utility functions $u_1, \dots, u_n$ is single-peaked [1, 12]. For instances of CPD with single-peaked utilities wrt axis $\sqsupset$, we always index the items according to $\sqsupset$, i.e., we have that $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_m$. We sometimes abuse the notation slightly and use $x_i \sqsupseteq x_j$ to denote that either x_i precedes x_j in $\sqsupset$ or $x_i = x_j$.

For instances of CPD with single-peaked preferences wrt some axis $\sqsupset$, we say that two items x_i and x_j are consecutive, if there is no other item x' such that $x_i \sqsupset x' \sqsupset x_j$ or $x_j \sqsupset x' \sqsupset x_i$. This definition naturally extends to any number of items.

For example, let us consider 4 items x_1, x_2, x_3, x_4 and 5 agents with the following preferences:

$1 : x_1 \succ_1 x_2 \succ_1 x_3 \succ_1 x_4$	$2 : x_2 \succ_2 x_1 \succ_2 x_3 \succ_2 x_4$	$3 : x_2 \succ_3 x_3 \succ_3 x_1 \succ_3 x_4$
$4 : x_3 \succ_4 x_2 \succ_4 x_4 \succ_4 x_1$	$5 : x_3 \succ_5 x_4 \succ_5 x_2 \succ_5 x_1$	

This set of preferences is single peaked wrt the societal axis $x_1 \sqsupset x_2 \sqsupset x_3 \sqsupset x_4$. In this example, the items e.g., x_1, x_2 and x_3 are consecutive.

Optimal Item Allocation for Single-Peaked Preferences. For instances with single-peaked preferences wrt axis $\sqsupset$, we can allocate any set of items M, $|M| = kq$, to slots $S_1, \dots, S_k$ in a greedy manner, so that each slot gets q items and the utility of each agent ℓ is $\max_{S \subseteq M, |S|=k} \sum_{x \in S} u_\ell(x)$, i.e., equal to the maximum utility that agent ℓ can derive from the items in M. Specifically, we arrange the items in M according to $\sqsupset$, so that $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_{kq}$, and let each slot $S_i = \{x_i, x_{i+k}, \dots, x_{i+(q-1)k}\}$. This allocation ensures that any k items consecutive in $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_{kq}$ are assigned to k different slots. The following shows that for single-peaked preferences, the best k items of any agent are consecutive in the societal axis (see Section A.2 for the proof).

Proposition 1. *Let X be a set of m items arranges as $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_m$, according to the societal axis $\sqsupset$, and let $u : X \to \mathbb{R}_{\geq 0}$ be any utility function that is single-peaked wrt $\sqsupset$. Then, for any $k \in [m]$, the maximum utility obtained from k items in X is achieved by considering k consecutive items in $\sqsupset$, i.e.,* $\max_{S \subseteq X, |S|=k} \sum_{x_p \in S} u(x_p) = \max_{x_j \in X} \sum_{p=j}^{k+j-1} u(x_p)$.

Therefore, for instances with single-peaked utilities, we can assume a greedy allocation of the items to slots and focus on the item selection aspect of CPD. Hence, for such instances, given any set of items $M \subseteq X$, with $|M| \leq kq$, we avoid referring to any particular allocation and let $u_\ell(M) = \max_{x_j \in M} \sum_{p=j}^{k+j-1} u_\ell(x_p)$ denote the maximum utility of an agent ℓ for the items in M and

$U(M) = \sum_{\ell=1}^{n} u_\ell(M)$ denote the maximum total utility that the agents can get from M. Moreover, for such instances, we assume that $|X| > kq$, since otherwise, CPD is easily solvable.

Linear Preferences. An interesting special case of single-peaked preferences are *linear preferences* (or *linear utilities*), where both the items and the agents lie in $[0, 1]$ and the utility of an agent ℓ for an item j is a linear decreasing function of their distance. For such instances, we assume that the items are located at $0 \le x_1 < x_2 < \cdots < x_m \le 1$ and the agents are located $0 \le v_1 \le v_2 \le \cdots \le v_n \le 1$. The utility of an agent ℓ for an item x_j is $u_\ell(x_j) = 1 - |x_j - v_\ell|$, i.e., the length of the interval minus the distance of v_ℓ to x_j. Namely, the utility of ℓ for j is equal to the opposite of the distance of v_ℓ to x_j, to which we add the length of the interval, so that the utility is nonnegative.

Single-Crossing Preferences. The notion of *single-crossing* preferences comes from Mirrlees [21] and Roberts [25]. A profile of preferences is single-crossing if there exists an ordering of the agents, say $\pi : [n] \to L$, such that for every pair of items $x_i, x_j \in X$, either all the agents rank x_i and x_j in the same way, or there is an index $t_{ij} \in \{1, \ldots n\}$ such that agents $\pi(1)$ to $\pi(t_{ij})$ all agree to rank x_i and x_j in the same way, and agents $\pi(t_{ij} + 1)$ to $\pi(n)$ all agree to rank x_i and x_j in the opposite way. So, either all the agents agree on the relative positions of two given items, or there is a dichotomy L_1, $L \setminus L_1$ such that both L_1 and $L \setminus L_1$ contain consecutive agents with respect to ordering π. We can check whether a preference profile is single-crossing in polynomial time [9].

For example, let us consider 4 items x_1, x_2, x_3, x_4 and 5 agents with the following preferences:

$1 : x_1 \succ_1 x_2 \succ_1 x_3 \succ_1 x_4$	$2 : x_1 \succ_2 x_2 \succ_2 x_4 \succ_2 x_3$	$3 : x_1 \succ_3 x_4 \succ_3 x_2 \succ_3 x_3$
$4 : x_4 \succ_4 x_1 \succ_4 x_2 \succ_4 x_3$	$5 : x_4 \succ_5 x_1 \succ_5 x_3 \succ_5 x_2$	

This preference profile is single-crossing, with π being the identity permutation of the agents.

3 Conference Program Design with Single-Peaked Preferences

Next, we consider CPD with single-peaked preferences and show that it can be solved in polynomial time. By Proposition 1, we can assume an optimal greedy allocation of the selected items to the slots and focus on the item selection aspect of CPD. Moreover, we can assume that $|X| > kq$.

Theorem 1. *CPD with single-peaked preferences is solvable in polynomial time.*

Proof (sketch). The proof extends the Linear Programming approach used in [18, Section 3] to show that the fault-tolerant version of q-Facility Location can be solved in polynomial-time on the line metric. Since we focus on item selection, we consider a simplified Integer Linear Programming formulation of CPD, where the slots $S_1, \ldots, S_k$ are not explicitly taken into account.

$$
\begin{aligned}
\text{(SLP)} \qquad \text{maximize} \quad & \sum_{\ell \in N} \sum_{x_i \in X} z_{\ell i} \cdot u_\ell(x_i) \\
\text{subject to:} \quad & y_i - z_{\ell i} \ge 0, \ \forall \ell \in N, x_i \in X && (1) \\
& \sum_{x_i \in X} z_{\ell i} \le k, \ \forall \ell \in N && (2) \\
& \sum_{x_i \in X} y_i \le k \cdot q && (3) \\
& y_i, z_{\ell i} \in \{0, 1\}, \ \forall \ell \in N, x_i \in X
\end{aligned}
$$

In (SLP), each variable y_i indicates whether an item x_i is included in the solution and each variable $z_{\ell i}$ indicates whether an agent ℓ derives utility from an item x_i. Constraint (1) ensures that an agent ℓ derives utility from an item x_i only if x_i is included in the solution. Constraint (2) ensures that every agent should derive utility from at most k items. Constraint (3) ensures that at most kq items should be selected in the solution.

The optimum of (SLP) is equal to the optimal total utility. Let us denote by (R-SLP) the relaxation of (SLP) where the constraints $y_i, z_{\ell i} \in \{0, 1\}$ are replaced by $y_i, z_{\ell i} \in [0, 1]$. Thus, the optimal value of (R-SLP) is no less than the value of (SLP). We solve (R-SLP) and let $X' = \{x_i \in X : \exists \ell \in L \text{ with } z_{\ell i} > 0\}$. We say that the *usage* of an item $x_i \in X'$ by an agent ℓ is *full* when $z_{\ell i} = y_i$, *null* when $z_{\ell i} = 0$ and *intermediate* when $0 < z_{\ell i} < y_i$. In Section A.3, we show that:

Claim 1 *Fix an agent ℓ and two consecutive items $x_a, x_b \in X'$ such that $x_a \succ_\ell x_b$. For that agent, the intermediate or null usage of x_a implies a null usage of x_b.*

Claim 1 implies that the items of X' for which a given agent has a non-null usage are consecutive. Moreover, within this set of consecutive items of interest for the agent, only the two extreme items can be used in an intermediate way. Using this observation, we modify X' as done in [18, Section 3], in the context of fault tolerant q-Facility Location. The crucial observation is that we can write a new Linear Program (FLP) for the modified instance such that (i) the optimum of (FLP) is as good as the optimum of (R-SLP); and (ii) (FLP) satisfies the *consecutive ones* property, and thus, has an integral optimal solution (see Section A.4 for a detailed description of the transformation and for the precise formulation of (FLP)).

The algorithm consists of solving (R-SLP) to get $(\boldsymbol{y}, \boldsymbol{z})$ and X'. Then, we split the items of X' according to $(\boldsymbol{y}, \boldsymbol{z})$ in order to get a new set of items X''. Next, we solve (FLP) and obtain an integral optimal solution of total utility equal to the total utility of the optimal solution of (R-SLP) (and of (SLP)). Obtaining an optimal selection of kq items from the solution of (FLP), we allocate the selected items using the optimal greedy allocation described in Section 2. □

4 A Truthful Mechanism for Single-Peaked Preferences

In this section, we present a truthful mechanism for CPD with single-peaked preferences. Given a set X of m items arranged as $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_m$ on axis $\sqsupset$, we consider the set $\mathcal{X} = \{C_1, \ldots, C_{m-k+1}\}$ of k-tuples of consecutive items, where $C_i = (x_i, \ldots, x_{k+i-1})$ for each $i \in [m-k+1]$. These k-tuples can be naturally arranged on $\sqsupset$, as $C_1 \sqsupset C_2 \sqsupset \cdots \sqsupset C_{m-k+1}$, according to their first coordinate.

For each agent ℓ and each k-tuple C_i, we let $\bar{u}_\ell(C_i) = \sum_{j=i}^{k+i-1} u_\ell(x_j)$ be the utility of agent ℓ for the items in C_i. Naturally, utilities $\bar{u}_\ell(C_1), \ldots, \bar{u}_\ell(C_{m-k+1})$ define the preference relation of agent ℓ on the set $\mathcal{X}$ of k-tuples of consecutive items. Hence, we extend the utility function u_ℓ of each agent ℓ on the set of items X to a utility function $\bar{u}_\ell$ on the set $\mathcal{X}$ of k-tuples of consecutive items. In Lemma 1, we show that if u_ℓ is single-peaked on X, $\bar{u}_\ell$ is single-peaked on $\mathcal{X}$.

Percentile Mechanism. In an $(\alpha_1, \ldots, \alpha_q)$-*Percentile Mechanism* for CPD, with $0 \leq \alpha_1 < \cdots < \alpha_q \leq 1$, each agent ℓ casts a vote for his best k-tuple $C^\ell = \arg\max_{C \in \mathcal{X}}\{\bar{u}_\ell(C)\}$. For each k-tuple C_i, we let $\mathrm{cnt}(C_i)$ denote the number of agents voting for C_i, i.e., $\mathrm{cnt}(C_i) = \{\ell \in L : C_i = C^\ell\}$. The mechanism selects q tuples from $\mathcal{X}$. For each $j \in [q]$, the k-tuple $C_i \in \mathcal{X}$ is selected as the j-th tuple of the $(\alpha_1, \ldots, \alpha_q)$-Percentile Mechanism if $\sum_{p=1}^{i-1} \mathrm{cnt}(C_p) < \alpha_j n \leq \sum_{p=1}^{i} \mathrm{cnt}(C_p)$.

Let $C(1), \ldots, C(q)$ be the k-tuples selected by the Percentile Mechanism, in the order of selection. We note that $C(1) \sqsupset \cdots \sqsupset C(q)$. Let also $M = \bigcup_{j=1}^{q} C(j)$ be the set of items selected by the mechanism. It may be $|M| \leq kq$, since $C(1), \ldots, C(q)$ do not need to be disjoint. The items in M are assigned greedily to slots so that if $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_{|M|}$ denote the items in M, each slot S_i consists of the items $x_i, x_{i+k}, x_{i+2k}, \ldots$. By Proposition 1 and by the discussion in Section 2, this greedy allocation is optimal and ensures that the utility $u_\ell(M)$ of each agent ℓ from the outcome of the mechanism is best possible. ♦

We next show that Percentile Mechanisms are truthful. Namely, for any instance, each agent ℓ maximizes his utility from the outcome of the mechanism by voting for his best k-tuple $C^\ell \in \mathcal{X}$.

Theorem 2. *For any tuple $(\alpha_1, \dots, \alpha_q)$, with $0 \le \alpha_1 < \cdots < \alpha_q \le 1$, the $(\alpha_1, \dots, \alpha_q)$-Percentile Mechanism is truthful for the Conference Program Design problem with single-peaked preferences.*

We first show that if a utility function u is single-peaked wrt $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_m$, its extension $\bar{u}$ on $\mathcal{X}$ is single-peaked wrt $C_1 \sqsupset C_2 \sqsupset \cdots \sqsupset C_{m-k+1}$.

Lemma 1. *Let $u : X \to \mathbb{R}_{\ge 0}$ be a single-peaked utility function wrt $x_1 \sqsupset x_2 \sqsupset \cdots \sqsupset x_m$ and let $\bar{u} : \mathcal{X} \to \mathbb{R}_{\ge 0}$ be its extension to the set $\mathcal{X} = \{C_1, \dots, C_{m-k+1}\}$ of k-tuples of consecutive items, where $\bar{u}(C_i) = \sum_{j=i}^{k+i-1} u(x_j)$ for each $C_i \in \mathcal{X}$. Then, $\bar{u}$ is single-peaked wrt $C_1 \sqsupset C_2 \sqsupset \cdots \sqsupset C_{m-k+1}$.*

Proof (sketch). We show that for any three k-tuples C_a, C_b, C_c such that either $C_a \sqsupset C_b \sqsupset C_c$ or $C_c \sqsupset C_b \sqsupset C_a$, if $\bar{u}(C_a) > \bar{u}(C_b)$, then $\bar{u}(C_b) \ge \bar{u}(C_c)$. We assume that $C_a \sqsupset C_b \sqsupset C_c$ (the other case is symmetric) and that $C_a = (x_1, \dots, x_k)$, $C_b = (x_\beta, \dots, x_{k+\beta-1})$, for some integer $\beta \ge 2$, and $C_c = (x_\gamma, \dots, x_{k+\gamma-1})$, for some integer $\beta + 1 \le \gamma \le m - k + 1$.

We need to show that for any choice of u, β and γ, $\bar{u}(C_a) > \bar{u}(C_b)$ implies that $\bar{u}(C_b) \ge \bar{u}(C_c)$. We let x_{j^*} be the item in X with maximum utility $u(x_{j^*})$, i.e., the peak of the utility function u, and consider three different cases depending on whether $j^* \le \beta$, $j^* \ge k+\beta-1$, or $\beta+1 \le j^* \le k+\beta-2$.

If $j^* \le \beta$, the single-peakedness of u implies that for all $j \in [k]$, $u(x_{\beta+j-1}) \ge u(x_{\gamma+j-1})$, because $x_{j^*} \sqsupseteq x_{\beta+j-1} \sqsupset x_{\gamma+j-1}$. Hence, we conclude that $\bar{u}(C_b) \ge \bar{u}(C_c)$.

We next show that $j^* \ge k + \beta - 1$ implies $\bar{u}(C_b) \ge \bar{u}(C_a)$, and thus, this case cannot occur. Specifically, if $j^* \ge k + \beta - 1$, the single-peakedness of u implies that for all $j \in [k]$, $u(x_{\beta+j-1}) \ge u(x_j)$, because $x_j \sqsupset x_{\beta+j-1} \sqsupseteq x_{j^*}$. Hence we obtain that $\bar{u}(C_b) \ge \bar{u}(C_a)$.

The most interesting case is where $\beta + 1 \le j^* \le k + \beta - 2$. Then, we need to compare the utility of some items in C_b with index less than j^* with some items in C_c with index greater than j^* (note that this comparison cannot be made directly for general single-peaked utilities). We do so by exploiting an item x_ℓ in C_a whose utility is no less than the utility of the rightmost items in C_b. Intuitively, such an item should exist because each item included in C_a but not in C_b has utility no greater than the utility of each item in $\{x_\beta, \dots, x_{j^*}\}$. Hence $\bar{u}(C_a) > \bar{u}(C_b)$ can hold only if some item in $C_a \setminus C_b$ has utility greater then the utility than the rightmost item in C_b. Although the idea is simple, the proof is a bit technical and can be found in Section A.5. □

We can now conclude the proof of Theorem 2. The greedy allocation of the items in M to slots ensures that all agents get a maximum utility from M. Thus, they do not have any incentive to manipulate the greedy assignment. We can also show that the agents cannot change the item selected by the mechanism in their favor. The intuition is the same as the intuition in the proofs that Generalized Median and Percentile Mechanisms [22, 28] are truthful for agents with single-peaked preferences. If an agent ℓ lies and votes for a k-tuple C' on the left (resp. on the right) of C^ℓ, this could only push a k-tuple $C(j)$ selected by the mechanism further on the left (resp. on the right) of C^ℓ. Since agent ℓ has single-peaked preferences over $\mathcal{X}$, such a change is not profitable for him. In Section A.6, we give a proof of this claim for completeness and because in our setting, an agent can use items from more than one k-tuples selected by the mechanism (instead of using a single outcome in [22, 28]). In fact, working as in the proof of [28, Theorem 1], we can show that Percentile Mechanisms for CPD with single-peaked preferences are group strategyproof.

5 The Approximation Ratio of Percentile Mechanisms for Linear Preferences

Next, we consider the special case of linear preferences. The items are located at $0 \le x_1 < x_2 < \cdots < x_m \le 1$ and the agents are located at $0 \le v_1 \le v_2 \le \cdots \le v_n \le 1$. The utility of an agent ℓ for an item x_j is $u_\ell(x_j) = 1 - |x_j - v_\ell|$. We always assume that $m > kq$, otherwise a greedy assignment of the items to slots is optimal.

The Approximation Ratio for Selecting a Single Item. We start with the case where $k = q = 1$, where we want to select a single item that maximizes the total utility of the agents. In contrast to 1-Facility Location, where the Median Mechanism is optimal (see e.g., [24]), the approximation ratio for this special case of CPD is $1/3$ and we can show that any deterministic truthful mechanism has approximation ratio at most $5/7$ (see Section A.7).

In the 1/2-Percentile Mechanism, each agent ℓ votes for his best item, i.e. for the item x_j that minimizes $|v_\ell - x_j|$. We recall that $\mathrm{cnt}(x_j)$ is the number of agents that vote for x_j. Then, the 1/2-Percentile Mechanism selects the item x_i that satisfies $\sum_{j=1}^{i-1} \mathrm{cnt}(x_j) < n/2 \le \sum_{j=1}^{i} \mathrm{cnt}(x_j)$. So, because it cannot select the location of the median agent, as the Median Mechanism does for 1-Facility Location [24], the 1/2-Percentile Mechanism selects the item closest to the location of the median agent (see Section A.8 for a proof of the following).

Claim 2 *Let $0 \le x_1 < \cdots < x_m \le 1$ be m items and $0 \le v_1 \le \cdots \le v_n \le 1$ be n agent locations in $[0, 1]$, and let v_{med} be the location of the median agent. Then, the 1/2-Percentile Mechanism selects the item $x_i = \arg\min_{j\in[m]} |x_j - v_{\mathrm{med}}|$, i.e., the item that minimizes the distance to v_{med}.*

The analysis of the approximation ratio uses the following (see Section A.9 for the proof).

Lemma 2. *Let $0 \le v_1 \le \cdots \le v_n \le 1$ be n agent locations in $[0, 1]$ and let v_{med} be the location of the median agent. For any items $z, y \in [0, 1]$ with $|v_{\mathrm{med}} - z| \le |v_{\mathrm{med}} - y|$, $U(y) \le 3\,U(z)$.*

Using Lemma 2, we can now determine the approximation ratio for the case where $k = q = 1$.

Lemma 3. *If $k = q = 1$, the 1/2-Percentile Mechanism achieves an approximation ratio of $1/3$.*

Proof. For the lower bound on the approximation ratio, we apply Lemma 2 with the item selected by the mechanism as z and the item selected by the optimal solution as y. Claim 2 implies that that $|v_{\mathrm{med}} - z| \le |v_{\mathrm{med}} - y|$, where v_{med} denotes the location of the median agent. Then, Lemma 2 immediately implies that the approximation ratio of the 1/2-Percentile Mechanism is at least $1/3$.

To conclude the proof, we present a class of instances where the mechanism has an approximation ratio of $1/3 + \epsilon$, for any $\epsilon > 0$ (see also the instance in the proof of Proposition 2). Such instances consist of $n/2$ agents located at $1/2 - \varepsilon$, where $\varepsilon > 0$ is arbitrarily small, and $n/2$ agents located at 1, and of 2 items, one at 0 and the other at 1. The optimal solution selects the item at 1 and has a total utility of $3n/4 - n\varepsilon/2$. The 1/2-Percentile Mechanism selects the item at 0 and has a total utility of $n/4 + n\varepsilon/2$. □

The Approximation Ratio for Singleton Slots. We now use Lemma 2 and show that for $q = 1$ and any $k \ge 1$, the approximation ratio of the 1/2-Percentile Mechanism is 3. In this case, each agent ℓ votes for his best k-tuple of consecutive items.The mechanism selects the k-tuple C_i that satisfies $\sum_{j=1}^{i-1} \mathrm{cnt}(C_j) < n/2 \le \sum_{j=1}^{i} \mathrm{cnt}(C_j)$. As in Claim 2, the k-tuple C_i selected by the mechanism is best k-tuple of the median agent, i.e., $C_i = \arg\max_{C_j \in \mathcal{X}} \{\bar{u}_{\lfloor (n+1)/2 \rfloor}(C_j)\}$.

Lemma 4. *If $q = 1$, for any $k \ge 2$, the 1/2-Percentile Mechanism on k-tuples of consecutive items achieves an approximation ratio of $1/3$.*

Proof. Let $Z = \{z_1, \dots, z_k\}$ be the set of items selected by the mechanism and let Y be the optimal set of k items. Since the items $z_1, \dots, z_k$ are consecutive in $[0,1]$ and correspond to the k items closest to the location v_{med} of the median agent (see also Proposition 1), we can arrange the items in Y as $y_1, \dots, y_k$ so that for each $j \in [k]$, $|z_j - v_{\text{med}}| < |y_j - v_{\text{med}}|$. Hence, Lemma 2 implies that for each pair of items z_j and y_j, $U(y_j) \le 3\,U(z_j)$. Since the optimal utility is $U(Y) = \sum_{j=1}^{k} U(y_j)$ and the mechanism's utility is $U(Z) = \sum_{j=1}^{k} U(z_j)$, the approximation ratio is at least $1/3$.

Moreover, for any $k \ge 2$, there instances where the mechanism has an approximation ratio of $1/3 + \epsilon$, for any $\epsilon > 0$. To see this, we generalize the tight example in the proof of Lemma 3. We consider the same agent locations and $2k$ items, k of them are essentially collocated at 0 and k of them are essentially collocated at 1. It is not hard to verify that the approximation ratio of the 1/2-Percentile Mechanism for this class of instances can be arbitrarily close to 1/3. □

The Approximation Ratio for the General Case. We proceed to bound the approximation of the $(\frac{1}{2q}, \frac{3}{2q}, \dots, \frac{2q-1}{2q})$-Percentile Mechanism for agents with linear preferences. The tight example in the proof of Lemma 3 shows that the distances $|y - z|$ and $v_n - v_1$ essentially determine the approximation ratio. This motivates us to introduce the notion of the width for a subset of agents.

We let L be a set of n agents with locations $0 \le v_1 \le \dots \le v_n \le 1$ in $[0,1]$, let $z \in [0,1]$ be an item, and let $Y \subseteq [0,1]$ be a nonempty set of items. Assuming that L, z and Y are fixed, we denote $y_l = \arg\max_{y \in Y \cup \{z\}} u_1(y)$ and $y_r = \arg\max_{y \in Y \cup \{z\}} u_n(y)$ the leftmost and the rightmost items in $Y \cup \{z\}$ used by some agent in L. The *width* $\beta(L, z, Y)$ (or simply β when L, z and Y are clear from the context) of the agent set L with respect to the item z and to the set Y is defined as:

$$\beta(L, z, Y) = \begin{cases} 0 & \text{if } Y \cap [y_l, y_r] \subseteq \{z\} \\ \max\{v_n - \min\{z, v_1\}, \max\{z, v_n\} - v_1\} & \text{otherwise} \end{cases}$$

Namely, if the only useful item in $Y \cup \{z\}$ is z, the width is 0. Otherwise, the width of L is either $v_n - v_1$, if $z \in [v_1, v_n]$, or $v_n - z$, if $z < v_1$, or $z - v_1$, if $z > v_n$. We next show that when a set of agents L is partitioned into groups that occupy non-overlapping intervals in $[0,1]$, we can bound the approximation ratio for L by analyzing independently the approximation ratio in each group (the proof can be found in Section A.10). This is the main intuition behind the proof of Theorem 3.

Lemma 5. *Let L be a set of n agents, located at $0 \le v_1 \le \dots \le v_n \le 1$ and partitioned into groups $L^1, \dots, L^q$, where each group consists of agents at consecutive locations. For any $j \in [q]$, let v^j_{med} be the location of the median agent in group L^j, and for any set Z of items, let $z^j = \arg\min_{z \in Z} |v^j_{\text{med}} - z|$. For any set Z with at most q items and any set Y of items with $|v^j_{\text{med}} - z^j| \le \min_{y \in Y} |v^j_{\text{med}} - y|$, let β^j denote the width of group L^j wrt z^j and Y. Then, $\sum_{j=1}^{q} \beta^j \le 2$.*

The following lemma determines the approximation ratio for a group of agents L that use the same item z, as a function of the width β. In addition to the conclusion of the lemma, part of its proof (in particular (4) and (16), in Section A.11), play an important role in the proof of Theorem 3.

Lemma 6. *Let L be a set of n agents located at $0 \le v_1 \le \dots \le v_n \le 1$ and let v_{med} be the location of the median agent in L. For any item z and any set of items Y such that $|v_{\text{med}} - z| \le \min_{y \in Y} |v_{\text{med}} - y|$, $U(Y) \le \frac{4-\beta}{4-3\beta}\,U(z)$, where $\beta \in [0,1]$ is the width of L with respect to z and Y.*

Proof (sketch). We use integer division by 2, in order to deal with both even and odd $n = |L|$. Since we are only interested in the ratio of $U(Y)/U(z)$, we restrict our attention to the set of useful items (for the agents in L) in Y. Specifically, we assume that $Y = (Y \cap [y_l, y_r]) \cup \{z\}$, i.e., that Y is restricted to its items in $[y_l, y_r]$ and includes z. In case where $Y = \{z\}$, the lemma holds trivially, because $\beta = 0$ and $U(Y) = U(z)$. So, from now on, we assume that $\{z\} \subset Y$.

We let $y = \arg\min_{y' \in Y \setminus \{z\}} |v_{\text{med}} - y'|$ and consider the case where $z < y$ (the case where $z > y$ is symmetric). So, $\beta = \max\{v_n - v_1, v_n - z\}$ (if $z > y$, $\beta = \max\{v_n - v_1, z - v_1\}$. We denote $\delta = (y - z)/2$. We distinguish two cases depending on whether $z \geq v_1$ or $z < v_1$ (if $z > y$, we distinguish two cases depending on whether $z \leq v_n$ or $z > v_n$).

We first consider the case where $z \geq v_1$. For convenience, we denote $\gamma = z - v_1$. In this case, $\beta = v_n - v_1$. Wlog., we assume that $y \leq v_n$ and that $\gamma + 2\delta \leq \beta$. (These inequalities can be enforced if we add to Y an artificial item located at v_n, which does not change the value of β, can only increase $U(Y)$ and does not change $U(z)$). We let n_1 be the number of agents located in $[v_1, z)$, n_2 (resp. L_2) denote the number (resp. the set) of agents located in $[z, z + \delta]$, and n_3 denote the number of agents in $(z + \delta, v_n]$. Since $z < y$ and $|v_{\text{med}} - z| \leq |v_{\text{med}} - y|$, the median agent is located in $[v_1, z + \delta]$. Therefore, $n_3 \leq n/2$. Moreover, we assume that $n_1 \leq n/2$ (i.e., we assume that $v_{\text{med}} \geq z$). Otherwise, the median agent is located on the left of z and this case is similar to the case where $y < z$ (again, we may add to Y an artificial item located at v_1). We have that

$$U(Y) \leq n - \sum_{j \in L_2} (z - v_j),$$

because each agent $j \in L_2$ has utility at most $1 - (z - v_j)$ for his favorite item in Y, while all the remaining agents have utility at most 1 for y. Similarly,

$$U(z) \geq n - n_1\gamma - \sum_{j \in L_2} (z - v_j) - n_3(\beta - \gamma),$$

because n_1 agents have utility at least $1 - (z - v_1) = 1 - \gamma$ for z, each agent $j \in L_2$ has utility $1 - (z - v_j)$ for z, and n_3 agents have utility at least $1 - (v_n - z) = 1 - (\beta - \gamma)$ for z.

We next show that for any $\alpha \geq 1$,

$$\alpha U(z) + \sum_{j \in L_2} (z - v_j) \geq \alpha n - \tfrac{3\alpha - 1}{4}\beta n \tag{4}$$

Then, using $\alpha = (4 - \beta)/(4 - 3\beta)$, we obtain that $(3\alpha - 1)\beta/4 = \alpha - 1$. Combining this equation with (4), we conclude that $\frac{4-\beta}{4-3\beta}U(z) \geq n - \sum_{j \in L_2}(z - v_j) \geq U(Y)$.

A detailed proof of (4) and the analysis for the case where $z < v_1$ can be found in Section A.11. The analysis for the case where $z < v_1$ follows exactly the same steps, but it is considerably simpler, since we have $n_1 = \gamma = 0$ in this case. □

We proceed to lower bound the approximation ratio in case where the number of agents is a multiple of q. The proof is based on the analysis in the proof of Lemma 6 and on Lemma 5.

Theorem 3. *For any integers $k \geq 1$ and $q \geq 2$, any number of items $m > qk$ and qn agents, for any integer $n \geq 1$, the approximation ratio of the $(\frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{2q-1}{2q})$-Percentile Mechanism for CPD instances with linear preferences is at least $(2q - 3)/(2q - 1)$.*

Proof. We denote $0 \leq v_1 \leq v_2 \leq \cdots \leq v_{qn} \leq 1$ the agent locatins. We partition the agents into q groups with n consecutive agents each. Specifically, for every $j \in [q]$, we let L^j consist of the agents $(j-1)n + 1, \ldots, jn$. Working as in the proof of Claim 2, we can show that for each $j \in [q]$, the $(\frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{2q-1}{2q})$-Percentile Mechanism selects the best k-tuple $C(j)$ of the median agent of L^j. By Proposition 1, each selected tuple $C(j)$ consists of the best k items of the median agent of L^j.

We let $M = \bigcup_{j=1}^{q} C(j)$ be the set of items selected by the mechanism. The utility is

$$U(M) = \sum_{j=1}^{q} \sum_{\ell \in L^j} u_\ell(M) \geq \sum_{j=1}^{q} \sum_{\ell \in L^j} \sum_{x_i \in C(j)} u_\ell(x_i)$$

For the proof, we assume that the agents in each group L^j use only the items in $C(j)$ and that the utility of each group L^j is $\sum_{\ell \in L^j} \sum_{x_i \in C(j)} u_\ell(x_i)$. Hence, the utility of the agents in L^j does not depend on the scheduling of the items in $C(j)$.

We let Y denote the set of items selected by the optimal solution and let $Y_i = \{y_i^1, \ldots, y_i^q\}$, with $0 \le y_i^1 < \cdots < y_k^j \le 1$, be the set of items assigned to each slot $i \in [k]$. We analyze the utility in each slot i separately. To argue about the approximation ratio for each slot i, we need an appropriate assignment of the items in the selected k-tuple $C(j)$, so that we can use the same approach as in the proof of Lemma 6 (this assignment does not change the utility that the agents in L^j get from $C(j)$, it just facilitates the proof). Since each $C(j)$ consists of the k items closest to the median location v^j_{med} in L^j and since these k items are consecutive in $[0, 1]$, we can schedule the items in $C(j)$ so that the distance of the item z_i^j scheduled in slot i to v^j_{med} is no greater than the distance of v^j_{med} to the nearest item in Y_i. Specifically, if the item in Y closest to v^j_{med} appears in slot i_1, we assign the item in $C(j)$ closest to v^j_{med} to slot i_1. Next, if the item in $Y \setminus Y_{i_1}$ closest to v^j_{med} appears in slot i_2, we assign the second closest item to v^j_{med} in $C(j)$ to slot i_2. In general, for each $p = 1, \ldots, k$, if the item in $Y \setminus (Y_{i_1} \cup \cdots \cup Y_{i_{p-1}})$ closest to v^j_{med} appears in slot i_p, we assign the p-th closest item to v^j_{med} in $C(j)$ to slot i_p. We let z_i^j denote the item in $C(j)$ assigned to slot i by this procedure. Since $C(j)$ contains the k items closest to v^j_{med}, we have that for each slot $i \in [k]$, $|z_i^j - v^j_{\text{med}}| \le \min_{y \in Y_i} |y - v^j_{\text{med}}|$.

After all items in M are assigned to slots, we consider the total utility of the agents for the set $Z_i = \{z_i^1, \ldots, z_i^q\}$, where $0 \le z_i^1 \le \cdots \le z_i^q \le 1$. Since each z_i^j belongs to $C(j)$, $z_i^j = \arg\min_{z \in Z_i} |v^j_{\text{med}} - z|$, i.e., z_i^j is the item in Z_i closest to v^j_{med}. Moreover, $|Z_i|$ may be less than q, i.e., some z_i^j's may correspond to the same actual item, since the k-tuples selected by the mechanism do not need to be disjoint. We next bound the ratio of the mechanism's utility $U(Z_i) = \sum_{j=1}^q \sum_{\ell \in L^j} u_\ell(z_i^j)$ for slot i to the optimal utility $U(Y_i) = \sum_{j=1}^q \sum_{\ell \in L^j} \max_{y \in Y_i}\{u_\ell(y)\}$ for slot i. At this point, we use the analysis of Lemma 6 (and refer to its proof).

For each $j \in [q]$, we let $U_j(z_i^j) = \sum_{\ell \in L^j} u_\ell(z_i^j)$ be the utility of the agents in group L^j for item z_i^j and let $U_j(Y_i) = \sum_{\ell \in L^j} \max_{y \in Y_i}\{u_\ell(y)\}$ be the utility of the agents in L^j for their best item in Y_i. By the assignment of the items in $C(j)$ to slots, we have that $|v^j_{\text{med}} - z_i^j| \le \min_{y \in Y_i} |y - v^j_{\text{med}}|$. We let β_i^j denote the width of the group L^j with respect to the item z_i^j and to the set of items Y_i (or simply, the width of group L^j for slot i). From the proof of Lemma 6, we know that if $\beta_i^j = 0$, then $U_j(z_i^j) = U_j(Y_i)$. So, we can assume wlog. that $\beta_i^j > 0$, for all groups L^j. Then, applying (4) and (16), from the proof of Lemma 6, we obtain that for any $\alpha \ge 1$,

$$\alpha U_j(z_i^j) \ge \alpha n - \frac{(3\alpha - 1)\beta_i^j n}{4} - \sum_{j \in L_2^j} |z - v_j|\,, \tag{5}$$

where L_2^j denotes the set of agents $\ell \in L^j$ so that $|z_i^j - v_\ell| \le \min_{y \in Y_i \setminus \{z_i^j\}} |y - v_\ell|$. Moreover, from the proof of Lemma 6, we have that

$$U(Y_i) = \sum_{j=1}^q U_j(Y_i) \le qn - \sum_{j=1}^q \sum_{j \in L_2^j} |z - v_j| \tag{6}$$

Summing up (5) over all $j \in [q]$, we obtain that

$$\alpha U(Z_i) \ge \alpha qn - \frac{(3\alpha - 1)n}{4} \sum_{j=1}^q \beta_i^j - \sum_{j=1}^q \sum_{j \in L_2^j} |z - v_j| \tag{7}$$

We observe that Lemma 5, with agent groups $L^1, \ldots, L^q$ and item sets $Z = Z_i$ and $Y = Y_i$, implies that $\sum_{j=1}^{q} \beta_i^j \leq 2$, for every slot $i \in [k]$. Using this bound on the sum of the widths and inequalities (7), (5) and (6), we obtain that for $\alpha = (2q-1)/(2q-3)$,

$$\alpha U(Z_i) \geq \alpha q n - \frac{(3\alpha - 1)n}{2} - \sum_{j=1}^{q} \sum_{j \in L_2^j} |z - v_j| = qn - \sum_{j=1}^{q} \sum_{j \in L_2^j} |z - v_j| \geq U(Y_i) \tag{8}$$

For the equality, we observe that for $\alpha = (2q-1)/(2q-3)$, $\alpha q - \frac{3\alpha - 1}{2} = q$. Summing up (8) over all slots $i \in [k]$, we obtain that

$$\frac{2q-1}{2q-3} U(M) \geq \frac{2q-1}{2q-3} \sum_{i=1}^{k} U(Z_i) \geq U(Y)\,,$$

which concludes the proof of the theorem. ☐

There are instances with nq agents and $k = 1$ where the approximation ratio of the Percentile Mechanism tends to $(2q-1)/(2q-1/q)$. E.g., for some odd integer $n \geq 3$, consider an instance with $(n-1)/2$ agents at 0, $n-1$ agents at each point i/q, $i = 1, \ldots, q-1$, $(n-1)/2$ agents at 1, and a single agent at each point $(2i+1)/(2q)$, $i = 0, \ldots, q-1$. We have $2q$ items located at points i/q, $i = 1, \ldots, q$, and at points $(2i+1)/(2q)$, $i = 0, \ldots, q-1$. The optimal solution is to select the items at i/q, for a total utility of roughly $(n-1)q - (n-1)/(2q)$. The $(\frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{2q-1}{2q})$-Percentile Mechanism selects the items at $(2i+1)/(2q)$, $i = 0, \ldots, q-1$, for a total utility of roughly $nq - (n-1)/2$. So, as n increases, the approximation ratio tends to $(2q-1)/(2q-1/q)$.

If the number of agents is not a multiple of q, we obtain a slightly weaker approximation ratio. The proof is similar to the proof of Theorem 3 and can be found in Section A.12.

Theorem 4. *For any $k \geq 1$ and $q \geq 2$, any $m > qk$ and any number of agents $|L| \geq q+1$, the approximation ratio of the $(\frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{2q-1}{2q})$-Percentile Mechanism for CPD instances with linear preferences is at least* $(2q-3-3q/|L|)/(2q-1-q/|L|) = (2q-3)/(2q-1) - O(q/|L|)$.

6 Conference Program Design with Item Preselection

In this section, we show that despite CPD being polynomially solvable for single-peaked preferences (Theorem 1), PRE-CPD is **NP**-hard for single-peaked preferences, even with the additional restriction of dichotomous preferences.

The agent preferences are *dichotomous* if each agent "likes" a subset of items and "dislikes" the remaining ones. This induces a preorder with two indifference classes for every agent. In a general setting, this implies $u_i(x) \in \{a, b\}$ where a, b are two non negative reals satisfying $a < b$. It is called *approval-based utility* when $a = 0$ and $b = 1$. Dichotomous preferences have recently attracted some attention from the community of Computational Social Choice, especially in the case of committee selection rules for voters [10, 11] or in judgment aggregation [8] even if the preferences also satisfy other restrictions like single-peakedness. In our setting, dealing with approval utilities is not restrictive with respect to algorithmic complexity issues, after a rescaling. So, we can assume that $u_\ell(x) \in \{0, 1\}$ for $\ell \in L$ and $x \in X$.

The preferences are *dichotomous single-peaked* if they are both single-peaked and dichotomous. In an equivalent view, the items are located one a line and each agent $\ell \in L$ corresponds to a closed interval I_ℓ of this line, where $u_\ell(x) = 1$ if $x \in I_\ell$ and $u_\ell(x) = 0$ otherwise. This is also known as *Voter Interval* in Voting Theory [10] or *Single-Plateauedness* in majority judgments [8].

In Section A.13, we prove the main result of this section.

Theorem 5. PRE-*CPD is* **NP**-*hard for dichotomous single-peaked preferences.*

7 Conference Program Design with Single-Crossing Preferences

In this section, we consider CPD with single-crossing preferences. Wlog., we assume that the preference profile is single-crossing for the identity permutation of the agents and that agent 1 prefers x_i to x_j if and only if $i < j$.

We extend the dynamic programming approach applied in [27] to the multi-winner election of Chamberlin-Courant. In particular, our dynamic programming algorithm is based on the contiguous blocks property of the optimal solution of Chamberlin-Courant with single-crossing preferences [27, Lemma 5], which directly extends to CPD. For a slot S_j of a solution to CPD such that S_j contains an item x_i, let $L(j,i)$ denote the set of agents who consider x_i as their best item in S_j. The *contiguous blocks property* for CPD states that for every $j \in [k]$ and $x_i \in S_j$, either $L(j,i) = \emptyset$ or there are two indices, t_{ji} and t'_{ji}, such that $t_{ji} \leq t'_{ji}$ and $L(j,i) = \{t_{ji}, t_{ji}+1, \ldots, t'_{ji}\}$. Moreover, for each $i < i'$ such that $L(j,i) \neq \emptyset$ and $L(j,i') \neq \emptyset$, it holds that $t'_{ji} < t_{ji'}$. In words, the contiguous blocks property states that an item is considered as the most preferred in a slot by a set of consecutive agents and that such sets of agents who prefer different items of the same slot do not overlap with each other. The proof that the contiguous blocks property holds for the optimal solution of CPD with single-crossing preferences is obtained by applying the proof of [27, Lemma 5] for each slot separately. The following summarizes the main result of this section.

Theorem 6. *A dynamic programming algorithm solves every single-crossing instance of CPD in* $O(m(nq)^{k+1})$ *time.*

Proof. Let $U(j,(i_1,t_1),\ldots,(i_k,t_k))$ be the maximum total utility if we use items from set $X_j = \{x_1,\ldots,x_j\}$ only, and in each slot S_p, only the agents $1,\ldots,i_p$ are taken into account and only t_p items are used. The function U is defined for all $j = 0,\ldots,m$ and for all $(i_p,t_p) \in \{0,\ldots,n\} \times \{0,\ldots,q\}$ such that $t_1 + \cdots + t_k \leq j$. For $j = 0$, $X_0 = \emptyset$.

We start with $U(0,(i_1,t_1),\ldots,(i_k,t_k)) = 0$, for all pairs $(i_1,t_1),\ldots,(i_k,t_k)$. For each $j \geq 0$, the next item x_{j+1} either is not selected (provided that $t_1 + \cdots + t_k \leq j$), in which case $U(j+1,(i_1,t_1),\ldots,(i_k,t_k)) = U(j,(i_1,t_1),\ldots,(i_k,t_k))$, or it is assigned to some slot S_p, in which case

$$U(j+1,(i_1,t_1),\ldots,(i_k,t_k)) = \max_{0\leq \ell \leq i_p}\left\{U(j,(i_1,t_1),\ldots,(\ell,t_p-1),\ldots,(i_k,t_k)) + \sum_{\nu=\ell+1}^{i_p} u_\nu(x_{j+1})\right\}$$

Therefore, for each $j \geq 0$ and each fixed $(i_1,t_1),\ldots,(i_k,t_k)$, such that $t_1 + \cdots + t_k \leq j+1$, $U(j+1,(i_1,t_1),\ldots,(i_k,t_k))$ can be defined recursively as follows:

$$\max\left\{\begin{array}{l} U(j,(i_1,t_1),\ldots,(i_k,t_k)) \\ \max_{1\leq p\leq k}\max_{0\leq \ell\leq i_p}\left\{U(j,\ldots,(\ell,t_p-1),\ldots) + \sum_{\nu=\ell+1}^{i_p} u_\nu(x_{j+1})\right\}\end{array}\right\}$$

in case where $t_1 + \cdots + t_k \leq j$, or

$$\max_{1\leq p\leq k}\max_{0\leq \ell\leq i_p}\left\{U(j,\ldots,(\ell,t_p-1),\ldots) + \sum_{\nu=\ell+1}^{i_p} u_\nu(x_{j+1})\right\}$$

in case where $t_1 + \cdots + t_k = j+1$. The optimal solution is given by $U(m,(n,k),\ldots,(n,k))$ (we implicitly assume here that $m \geq kq$). The number of values that we need to compute is $O(m(nq)^k)$ and the total running time is $O(m(nq)^{k+1})$. In Section A.14, we provide some more details on how we compute these values. □

References

1. J. Bartholdi III and M. Trick. Stable Matching with Preferences Derived from a Psychological Model. *Operations Research Letters*, 5(4):165-169, 1986.
2. N. Betzler, A. Slinko, and J. Uhlmann. On the Computation of Fully Proportional Representation. *Journal of Artificial Intelligence Research (JAIR)*, 47:475-519, 2013.
3. M. Biró, M. Hujter, and Z.Tuza. Precoloring extension. I. Interval graphs. *Discrete Mathematics*, 100(1-3):267-279, 1992.
4. D. Black. *The Theory of Committees and Elections*. Cambridge University Press, 1958.
5. A. Brandstadt, V.B. Le, and J.P. Spinrad. *Graph Classes, a Survey*. SIAM Monographs on Discrete Mathematics and Applications, 1999.
6. I. Caragiannis, L. Gourvès, and J. Monnot. Achieving proportional representation in conference programs. In *Proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)*, 2016.
7. J.R. Chamberlin and P.N. Courant. Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule. *The American Political Science Review*, 77(3):718-733, 1983.
8. F. Dietrich and C. List. Majority voting on restricted domains. *Journal of Economic Theory*, 145(2):512-543, 2010.
9. E. Elkind, P. Faliszewski, and A.M. Slinko. Clone structures in voters' preferences. In *Proc. of the 13th ACM Conference on Electronic Commerce (EC 2012)*, pp. 496-513, 2012.
10. E. Elkind and M. Lackner. Structure in Dichotomous Preferences. In *Proc. of the 24th International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pp. 2019-2025, 2015.
11. G. Erdélyi, M. Lackner, and A. Pfandler. Manipulation of k-Approval in Nearly Single-Peaked Electorates. In *Proc. of the 4th International Conference on Algorithmic Decision Theory (ADT 2015)*, LNCS 9346, pp. 71-85, 2015.
12. B. Escoffier, J. Lang, and M. Öztürk. Single-Peaked Consistency and its Complexity. In *Proc. of the 18th European Conference on Artificial Intelligence (ECAI 2008)*, pp. 366-370, 2008.
13. M. Feldman, A. Fiat, and I. Golomb. On Voting and Facility Location. In *Proc. of the 17th ACM Conference on Economics and Computation (EC 2016)*, pp. 269-286, 2016.
14. D. Fotakis and C. Tzamos. Strategyproof facility location for concave cost functions. In *Proc. of the 14th ACM Conference on Electronic Commerce (EC 2013)*, pp. 435-452, 2013.
15. D. Fotakis and C. Tzamos. On the Power of Deterministic Mechanisms for Facility Location Games. *ACM Transactions on Economics and Computation*, 2(4):15:1-15:37, 2014.
16. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1979.
17. M. C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Annals of Discrete Mathematics, Vol 57, Elsevier, 2004.
18. M. Hajiaghayi, W. Hu, J. Li, S. Li, and B. Saha. A Constant Factor Approximation Algorithm for Fault-Tolerant k-Median. In *Proc. of the 25th ACM-SIAM Symposium on Discrete Algorithms (SODA 2014)*, pp. 1-12, 2014.
19. P. J. Looges and S. Olariu. Optimal greedy algorithms for indifference graphs. *Computers and Mathematics with Applications*, 25(7):15-25, 1993.
20. Dániel Marx. Precoloring extension on unit interval graphs. *Discrete Applied Mathematics*, 154(6):995-1002, 2006.
21. J. Mirrlees. An Exploration in the Theory of Optimal Income Taxation. *Review of Economic Studies*, 38:175-208, 1971.
22. H. Moulin. On strategy-proofness and single-peakedness. *Public Choice*, 35:437-455, 1980.
23. A. Procaccia, J. Rosenschein, and A. Zohar. On the complexity of achieving proportional representation. *Social Choice and Welfare*, 30(3):353-362, 2008.
24. A.D. Procaccia and M. Tennenholtz. Approximate mechanism design without money. *ACM Transactions on Economics and Computation*, 1(4):18, 2013.
25. K.W.S. Roberts. Voting over income tax schedules. *Journal of Public Economics*, 8(3):329-340, 1977.
26. A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Springer-Verlag, 2003.
27. P. Skowron, L. Yu, P. Faliszewski, and E. Elkind. The complexity of fully proportional representation for single-crossing electorates. *Theoretical Computer Science*, 569:43-57, 2015.
28. X. Sui, C. Boutilier, and T. Sandholm. Analysis and Optimization of Multi-Dimensional Percentile Mechanisms. In *Proc. of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, 2013.
29. D.P. Williamson and D.B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, 2011.

A Appendix

A.1 An Integer Linear Programming Formulation of Conference Program Design

In [6], it was shown that the Conference Program Design problem can be formulated as an Integer Linear Program, denoted by (ILP), where $y_{ij} = 1$, if item $x_i \in X$ is assigned to slot S_j, and $y_{ij} = 0$, otherwise, and $z_{\ell,i,j} = 1$, if x_i is the item of S_j from which agent ℓ derives his utility.

$$\text{(ILP)} \qquad \text{maximize} \sum_{\ell \in N} \sum_{j \in [k]} \sum_{x_i \in X} z_{\ell,i,j} \cdot u_\ell(x_i)$$

$$\text{subject to:} \sum_{x_i \in X} z_{\ell,i,j} = 1, \ \forall \ell \in N, j \in [k] \tag{9}$$

$$y_{ij} - z_{\ell,i,j} \geq 0, \ \forall \ell \in N, j \in [k], x_i \in X \tag{10}$$

$$\sum_{j \in [k]} y_{ij} \leq 1, \ \forall x_i \in X \tag{11}$$

$$\sum_{x_i \in X} y_{ij} \leq q, \ \forall j \in [k] \tag{12}$$

$$y_{ij}, z_{\ell,i,j} \in \{0,1\}, \ \forall \ell \in N, j \in [k], x_i \in X$$

Constraint (9) ensures that agent ℓ should derive his utility from a single item in each slot S_j. Constraint (10) ensures that an agent ℓ can derive utility only from an item x_i which is assigned to slot S_j. Constraint (11) ensures that each item x_i should appear in at most one slot S_j, and constraint (12) ensures that each slot S_j should contain at most q items.

A.2 The Proof of Proposition 1

We observe that the subsets $S \subseteq X$, with $|S| \leq k$, forms a uniform matroid of rank k. Obtaining the maximum utility from k items in X is equivalent to selecting a basis of the uniform matroid with maximum total utility. The greedy algorithm, which selects the k best items in X according to u, finds a subset of k items in X of maximum utility. Since the utility function u is single-peaked with respect to $\sqsupset$, the k best items in X are consecutive. □

A.3 The Proof of Claim 1

If the usage of x_b is not null then $z_{\ell b} > 0$. If the usage of a is intermediate or null then $z_{\ell a} < y_a$. Let $\delta = \min\{z_{\ell b}, y_a - z_{\ell a}\}$ and note that $\delta > 0$. We can increase the utility of agent ℓ by simultaneously decreasing $z_{\ell b}$ and increasing $z_{\ell a}$ by δ. This modification has no impact on the utility of the other agents, so we get a contradiction with the optimality of (y, z) for (R-SLP). □

A.4 Missing Details from the Proof of Theorem 1

We recall that Claim 1 implies that the items of X' for which a given agent has a non-null usage are consecutive. Moreover, within this set of consecutive items of interest for the agent, only the two extreme items can be used in an intermediate way.

Using this observation, we now modify X' as done in [18, Section 3], in the context of fault tolerant q-Facility Location. Specifically, every item $x_i \in X'$ for which no agent has an intermediate usage is renamed as f_i^1 and we let $\mathsf{fic}(x_i) := \{f_i^1\}$. Every item $x_i \in X'$ for which at least one agent has an intermediate usage is replaced by a set $\mathsf{fic}(x_i)$ of new items. Let left(i) and right(i) denote the sets of agents placed to the left and to the right of x_i on the societal axis, respectively. Numbers $\{z_{\ell i}\}_{\ell \in \text{left}(i)}$ and $\{y_i - z_{\ell i}\}_{\ell \in \text{right}(i)}$ split the interval $[0, y_i]$ into several pieces, say $d(i)$ pieces, and for each piece, we create a fictitious item with fractional value equal to the length of that piece. Thus x_i is replaced by a set of $d(i)$ items $\mathsf{fic}(x_i) := \{f_i^1, \ldots, f_i^{d(i)}\}$ whereas y_i is split in $d(i)$ positive values $(y_i^1, \ldots, y_i^{d(i)})$ such that $y_i^1 + \ldots + y_i^{d(i)} = y_i$.

Now the new set of items X'' is defined as $\bigcup_{x_i \in X'} \mathsf{fic}(x_i)$. Each agent ℓ is associated with a minimal subset X''_ℓ of X'' satisfying:

- if $z_{\ell,i} = 0$ then $X''_\ell \cap \mathsf{fic}(x_i) = \emptyset$
- if $z_{\ell,i} = y_i$ then $X''_\ell \supseteq \mathsf{fic}(x_i)$
- if $0 < z_{\ell,i} < y_i$ and $\ell \in \text{left}(i)$ then $X''_\ell \cap \mathsf{fic}(x_i) = \{f_i^1, \ldots, f_i^{g(\ell)}\}$ where $y_i^1 + \ldots + y_i^{g(\ell)} = z_{\ell,i}$
- if $0 < z_{\ell,i} < y_i$ and $\ell \in \text{right}(i)$ then $X''_\ell \cap \mathsf{fic}(x_i) = \{f_i^{g(\ell)}, \ldots, f_i^{d(i)}\}$ where $y_i^{g(\ell)} + \ldots + y_i^{d(i)} = z_{\ell,i}$

We now consider the following Linear Program, where the variable t_i^j is associated with the item f_i^j of X''.

$$
\begin{aligned}
\text{(FLP)} \qquad & \text{maximize} \sum_{\ell \in N} \sum_{f_i^j \in X''_\ell} t_i^j \cdot u_\ell(x_i) \\
& \text{subject to:} \sum_{f_i^j \in X''_\ell} t_i^j \leq k, \ \forall \ell \in N && (13) \\
& \sum_{f_i^j \in \mathsf{fic}(x_i)} t_i^j \leq 1, \ \forall x_i \in X' && (14) \\
& \sum_{f_i^j \in X''} t_i^j \leq k \cdot q && (15) \\
& t_i^j \in [0,1], \ \forall f_i^j \in X''
\end{aligned}
$$

Constraint (13) ensures that every agent should derive his utility from at most k items within the set of items from which he was deriving utility in (SLP). Constraint (14) ensures that at most one fictitious item should be kept per original item. Constraint (15) ensures that at most kq items should be selected in the solution.

The optimum of (FLP) is as good as the optimum of (R-SLP). Moreover, (FLP) has an integral optimal solution $\boldsymbol{t}$, because it satisfies the so-called consecutive ones property. Namely, in each row of the constraint matrix, the 1s appear in consecutive positions. Indeed, in (FLP) the items of $\mathsf{fic}(x_i)$ for every $x_i \in X'$, or the items of X''_ℓ for every $\ell \in L$, are consecutive. Matrices with the consecutive ones property are known to be totally unimodular [26], and thus, (FLP) has an integral optimal solution $\boldsymbol{t}$.

Overall, the algorithm consists of first solving (R-SLP) to get $(\boldsymbol{y}, vecz)$ and X'. Then, we split the items of X' according to $(\boldsymbol{y}, \boldsymbol{z})$ in order to get a new set of items X''. Next, we solve (FLP) and obtain an integral optimal solution $\boldsymbol{t}$ of total utility equal to the total utility of the optimal solution of (R-SLP) (and of (SLP)). Hence, we retrieve an optimal set of at most kq items by selecting x_i if and only if there some j such that $t_i^j = 1$. Finally, we allocate the selected items using the optimal greedy allocation described in Section 2. □

A.5 Missing Details from the Proof of Lemma 1

The most interesting case is where $\beta+1 \le j^* \le k+\beta-2$. In this case, we need to compare the utility of some items in C_b with index less than j^* with some items in C_c with index greater than j^*. We do so by exploiting an item x_ℓ in C_a whose utility is no less than the utility of the rightmost items in C_b. Specifically, we show that there is an index ℓ, $1 \le \ell \le \min\{k, \beta-1, k+\beta-j^*-1\}$, so that $u(x_\ell) > u(x_{k+\beta-\ell})$. Intuitively, since C_b includes the item of maximum utility x_{j^*} but $\bar{u}(C_a) > \bar{u}(C_b)$, there is some item in C_a which is not included in C_b and has more utility than the corresponding item on the right part of C_b. To prove this claim, we first show that there is such an index ℓ with $1 \le \ell \le k$. Otherwise, for all $j \in [k]$, it would be $u(x_j) \le u(x_{k+\beta-j})$. Hence, $\bar{u}(C_a) \le \bar{u}(C_b)$, which contradicts the hypothesis that $\bar{u}(C_a) > \bar{u}(C_b)$. To show that $\ell \le \beta - 1$, we observe that if $\beta \le k$, i.e., if both C_a and C_b include items $x_\beta, \dots, x_k$, then $\sum_{j=\beta}^{k} u(x_j) = \sum_{j=\beta}^{k} u(x_{k+\beta-j})$, because the same terms $u(x_\beta), \dots, u(x_k)$ appear in both sums. Therefore, to avoid reaching the contradiction where $\bar{u}(C_a) \le \bar{u}(C_b)$, there must be a pair of items x_ℓ, included in C_a but not in C_b, and $x_{k+\beta-\ell}$, included in C_b but not in C_a, such that $u(x_\ell) > u(x_{k+\beta-\ell})$. To show that $\ell \le k+\beta-j^*-1$, we observe that if $j^* \le k$, i.e., if the item x_{j^*} of maximum utility is included in C_a, then $\ell \le \beta - 1$ implies that $\ell \le k+\beta-j^*-1$. If $j^* > k$, for all $j = 0, \dots, j^*-\beta$, $u(x_{j^*-j}) \ge u(x_{k-j})$, because $x_{k-j} \sqsupset x_{j^*-j} \sqsupseteq x_{j^*}$. Therefore, $\sum_{j=0}^{j^*-\beta} u(x_{j^*-j}) \ge \sum_{j=0}^{j^*-\beta} u(x_{k-j})$, where the sum on the lhs accounts for the utilities of $j^*-\beta+1$ items in C_b and the sum on the rhs accounts for the utilities of $j^*-\beta+1$ items in C_a. Hence, to avoid reaching the contradiction where $\bar{u}(C_b) \ge \bar{u}(C_a)$, there must be a pair of items $x_\ell \in C_a$ and $x_{k+\beta-\ell} \in C_b$, which are not included in the two sums. Hence, it must be that $\ell \le k-(j^*-\beta)-1$.

Since $1 \le \ell \le k+\beta-j^*-1$, $j^*+1 \le k+\beta-\ell \le k+\beta-1$. So, the item $x_{k+\beta-\ell}$ is included in C_b and satisfies $x_{j^*} \sqsupset x_{k+\beta-\ell}$. Then, for any index j, with $\beta \le j \le j^*$, $u(x_j) \ge u(x_\ell) > u(x_{k+\beta-\ell})$, where the first inequality holds because $\ell \le \beta-1 < j \le j^*$ and thus, $x_\ell \sqsupset x_j \sqsupseteq x_{j^*}$, and the second inequality holds by the choice of index ℓ. Therefore, $u(x_j) > u(x_{k+\beta-\ell})$, for all j with $\beta \le j \le j^*-1$, i.e., for any item x_j in C_b with $x_j \sqsupset x_{j^*}$.

We are now ready to complete the proof of the lemma by showing that in case where $\beta+1 \le j^* \le k+\beta-2$, that $\bar{u}(C_b) \ge \bar{u}(C_c)$. For brevity, we only discuss here the case where $j^* < \gamma \le k+\beta-1$ (in all other cases, the desired inequality can be derived by a straightforward adaptation of the same argument). Since the items $x_\gamma, \dots, x_{k+\beta-1}$ are included in both C_b and C_c, they contribute the same utility to both $\bar{u}(C_b)$ and $\bar{u}(C_c)$. Moreover, we observe that for any index $j = 0, \dots, \gamma-1-j^*$, $u(x_{j^*+j}) \ge u(x_{k+\beta+j})$, because $x_{j^*} \sqsupseteq x_{j^*+j} \sqsupset x_{k+\beta+j}$. Finally, we observe that for any index $j = \beta, \dots, j^*-1$, $u(x_j) > u(x_{k+\beta-\ell}) \ge u(x_{k+\gamma-j^*+j})$, where the first inequality was shown in the previous paragraph and the second inequality holds because $x_{j^*} \sqsupset x_{k+\beta-\ell} \sqsupseteq x_{k+\gamma-j^*+j}$, since $k+\beta-\ell \le k+\beta+\gamma-j^*$. Therefore, we have found a one-to-one mapping of the items in C_b to the items in C_c so that the utility of each item in C_b is no less than the utility of the corresponding item in C_c. This implies that $\bar{u}(C_b) \ge \bar{u}(C_c)$ and concludes the proof of the lemma. □

A.6 Missing Details from the Proof of Theorem 2

To conclude the proof of Theorem 2, we fix the votes of all agents, except for a fixed agent ℓ. To reach a contradiction, we assume if agent ℓ votes for his best k-tuple C^ℓ, the set of items selected by the $(\alpha_1, \dots, \alpha_q)$-Percentile Mechanism is M, while if ℓ votes for a different k-tuple C', the outcome of the mechanism is M', with $u_\ell(M) < u_\ell(M')$.

Let $C(1) \sqsupset \cdots \sqsupset C(j-1) \sqsupset C(j) \sqsupset \cdots \sqsupset C(q)$ be the k-tuples selected by the mechanism when agent ℓ votes for his true best k-tuple C^ℓ. We assume that C^ℓ is different from all the selected k-tuples, since if C^ℓ is selected, agent ℓ has no incentive to lie about his best k-tuple. We let

$C(j-1) \sqsupset C^\ell \sqsupset C(j)$ and assume that $j \le q$ (i.e., C^ℓ is on the left of at least one k-tuple $C(j)$ selected by the mechanism). We consider the case where agent ℓ lies by voting for a k-tuple C' on the right of C^ℓ, i.e., $C^\ell \sqsupset C'$ (the case where $C' \sqsupset C^\ell$ is symmetric). As a result of ℓ's false declaration, the Percentile Mechanism selects some sets $C'(j), \dots, C'(q)$, where $C(j) \sqsupseteq C'(j), \dots, C(q) \sqsupseteq C'(q)$. Since we assume that agent ℓ improves his utility by reporting C' instead of C^ℓ, there must be at least one index p, $j \le p \le q$, so that $C(p)$ and $C'(p)$ are different. For simplicity, we assume that the smallest such index is j and that $C(j) \neq C'(j)$ (the same argument applies to the smallest index p such that $C(p) \neq C'(p)$).

We first observe that since u_ℓ is single-peaked, the best k-tuple C^ℓ of agent ℓ includes the item x^* of maximum utility in u_ℓ. Since $C^\ell \sqsupset C(j)$, x^* lies on the left of the rightmost item in $C(j)$. Therefore, if the rightmost item used by ℓ in M is on the left of the rightmost item of $C(j)$, agent ℓ cannot increase his utility by forcing the mechanism to select $C'(j)$ instead of $C(j)$, because $C'(j)$ just includes some items further on the right than the items used by ℓ. So, the rightmost item used by ℓ in M must coincide with the rightmost item of $C(j)$. By Proposition 1, the best k items of ℓ in M are consecutive. So, $C(j)$ consists of the best k items of ℓ in M. Since $C(j) \sqsupset C'(j)$ and ℓ prefers $C'(j)$ to $C(j)$, agent ℓ prefers using some items on the right of the rightmost item of $C(j)$ to using the leftmost items of $C(j)$. Therefore, there exists a $C''(j)$, $C(j) \sqsupset C''(j) \sqsupseteq C'(j)$ (note that it may be $C''(j) = C'(j)$) so that $\bar{u}_\ell(C''(j)) > \bar{u}_\ell(C(j))$. This contradicts the fact that the utility function $\bar{u}_\ell$ is single-peaked (Lemma 1). More specifically, since agent ℓ prefers $C'(j)$ to $C(j)$, it must be $\bar{u}_\ell(C^\ell) > \bar{u}_\ell(C(j))$. Moreover, $C^\ell \sqsupset C(j) \sqsupset C''(j)$ and $\bar{u}_\ell(C''(j)) > \bar{u}_\ell(C(j))$, which is a contradiction to Lemma 1.

In case where $C(q) \sqsupset C^\ell$ (resp. where $C^\ell \sqsupset C(1)$), the k-tuples selected by the mechanism do not change if agent ℓ votes for a k-tuple C' with $C(q) \sqsupset C'$ (resp. with $C' \sqsupset C(1)$). This concludes the proof of Theorem 2. □

Working as in the proof of [28, Theorem 1], we can strengthen the proof of Theorem 2 and show that Percentile Mechanisms for CPD with single-peaked preferences are group strategyproof. The crucial observation is that an agent ℓ could lie and improve his utility from the outcome M of the mechanism only if ℓ's best k items in M are included in a single k-tuple $C(j)$ (which is selected by the mechanism when ℓ reports C^ℓ truthfully) and there is a different k-tuple C' of total utility $\bar{u}_\ell(C') > \bar{u}_\ell(C(j))$, either on the left or on the right of $C(j)$. Thus, even though in our setting, an agent ℓ can use items from more than one k-tuples, the proof of group strategyproofness (and the proof of truthfulness above) boils down to the q-Facility Location setting analyzed in [28, Theorem 1]. We omit further details from this extended abstract.

A.7 On the Approximation Ratio of Deterministic Truthful Mechanisms

Proposition 2. *Any deterministic truthful mechanism for the Conference Program Design with linear preferences and with $k = q = 1$ has approximation ratio at most $5/7$.*

Proof. We assume that there exists a deterministic truthful mechanism with approximation ratio strictly greater than $5/7$ and reach a contradiction. We consider an instance with 2 items, located at $x_1 = 0$ and $x_2 = 1$, and 2 agents, located at $v_1 = 1/6$ and $v_2 = 2/3$. The optimum is to select item $x_1 = 0$, for a total utility of $7/6$. Selecting item $x_2 = 1$, which is preferred by agent 2, gives a total utility of $5/6$. Therefore, any deterministic mechanism with an approximation ratio greater than $5/7$ selects item $x_1 = 0$ for this instance.

If agent 2 reports $v_2' = 1$, the optimum is to select item $x_2 = 1$, for a total utility of $7/6$. Selecting item $x_1 = 0$ now gives a total utility of $5/6$. So, any deterministic mechanism with an approximation ratio greater than $5/7$ selects item $x_2 = 1$ for this instance. Since agent 2 can enforce item x_2 to such a mechanism, the mechanism is not truthful. □

A.8 The Proof of Claim 2

Let x_l (resp. x_r) be the item closest to v_{med} on the left, with $x_l \le v_{\text{med}}$ (resp. on the right, with $x_r > v_{\text{med}}$). Then, $\sum_{x_j \le x_r} \text{cnt}(x_j) \ge n/2$, because all agents with index at most $(n+1)/2$ vote for an item in $\{x_1, \dots, x_r\}$. In fact, if $|x_l - v_{\text{med}}| \le |x_r - v_{\text{med}}|$, all agents with index at most $(n+1)/2$ vote for an item in $\{x_1, \dots, x_l\}$, and x_l is selected by the mechanism. Otherwise, the median agent votes for x_r, which implies that $\sum_{x_j \le x_l} \text{cnt}(x_j) < n/2 \le \sum_{x_j \le x_r} \text{cnt}(x_j)$. So, in this case, the mechanism selects the item x_r. In both cases, the 1/2-Percentile Mechanism selects the item closest to the location of the median agent v_{med}. □

A.9 The Proof of Lemma 2

We use integer division by 2, in order to deal with both even and odd n. We consider the case where $z < y$ (the case where $z > y$ is symmetric, while if $z = y$, $U(z) = U(y)$). We denote $\delta = (y - z)/2 > 0$.

Since $|v_{\text{med}} - z| \le |v_{\text{med}} - y|$, $v_1 \le \cdots \le v_{\text{med}} \le z + \delta$. Let n_1 denote the number of agents located in $[0, z)$, n_2 denote the number of agents located in $[z, z+\delta]$, and n_3 denote the number of agents in $(z+\delta, 1]$. Since the median agent is located in $[0, z+\delta]$, $n_3 \le n/2$. Moreover, we assume that $n_1 \le n/2$, because if $n_1 \ge (n+1)/2$, $U(z) \le U(y)$, since z would be located in $[v_{\text{med}}, y]$. Then,

$$U(y) \le n - n_2\delta\,,$$

because at least n_2 agents are at distance at least δ to y. So these agents have utility at most $1 - \delta$ for y, while the remaining agents have utility at most 1 for y. Similarly,

$$U(z) \ge n - n_1 z - n_2\delta - n_3(1-z)\,,$$

because n_1 agents have utility at least $1 - z$ for z, n_2 agents have utility at least $1 - \delta$ for z, and n_3 agents have utility at least z for z.

Therefore, to conclude the proof of the lemma, it suffices to show that

$$2n_2\delta + 3n_1 z + 3n_3(1-z) \le 2n\,.$$

Next, we derive a sequence of upper bounds, which shows that the lhs of the inequality above is at most $2n$. Specifically, we observe that

$$\begin{aligned}
2n_2\delta + 3n_1 z + 3n_3(1-z) &\le n_2(1-z) + 3n_1 z + 3n_3(1-z)\\
&\le n(1-z) + 3n_1 z + 2n_3(1-z)\\
&\le n(1-z) + (3n/2)z + n(1-z) \le 2n
\end{aligned}$$

The first inequality holds because $z + 2\delta \le 1$. For the second inequality, we use that $n_2 + n_3 \le n$. The third inequality follows from $n_1 \le n/2$ and $n_3 \le n/2$. □

A.10 The Proof of Lemma 5

We let v_l^j (resp. v_r^j) denote the leftmost (resp. rightmost) agent in each group L^j. For convenience, we introduce the notion of the *active interval* of each group L^j (with respect to the item z^j and to the set Y). The active interval of a group L^j is empty if $\beta_j = 0$. Otherwise, the active interval of L_j is either $[v_l^j, v_r^j]$, if $z^j \in [v_l^j, v_r^j]$, or $[z^j, v_r^j]$, if $z^j < v_l^j$, or $[v_l^j, z^j]$, if $z^j > v_r^j$. We observe that the length of the active interval of each group L^j is equal to β^j (assuming that the length of the empty interval is 0). Moreover, we observe that if $z^j \notin [v_l^j, v_r^j]$, the active interval of a group L^j cannot extend beyond the item z^j (on the left, if $z^j < v_l^j$, or on the right if $z^j > v_r^j$), where z^j is the item in Z closest to the median location v_{med}^j of the agents in L^j. In the following, we say that a point x is *included strictly* in an interval $[a, b]$ if $x \in (a, b)$.

To prove the lemma, we show that each point $x \in [0, 1]$ is included strictly in the active interval of at most two groups of agents L^j and L^{j+1} with consecutive indices. Since the width of the active interval of each L^j is equal to β^j, this implies that $\sum_{j=1}^{q} \beta^j \leq 2$.

We fix some index j, assume that the active interval of L^j is $[v_l^j, v_r^j]$, and consider any point x in the active interval of L^j (a similar reasoning also applies to the case where the active interval of L^j is either $[z^j, v_r^j]$ or $[v_l^j, z^j]$. We have to show that x is included strictly in the active interval of at most one group $L^{j'}$ different from L^j. If $z^j \in [v_l^j, v_r^j]$ and $x \leq z^j$ (the case where $x \geq z^j$ is symmetric), the left endpoint of the active interval of any group $L^{j'}$, with $j' > j$, should be on the right of z^j, because $z^j \leq z^{j'}$. Therefore, x is not included strictly in any of these active intervals.

We examine now the right endpoints of the active intervals of the groups $L^{j'}$, with $j' < j$. If the right endpoint of the active interval of L^{j-1} is at some point x', with $v_r^{j-1} \leq x' \leq x$, it must be $z^{j-1} \leq x$. Hence, in this case, x is not included strictly in any of the active intervals of groups $L^{j'}$, with $j' < j$. If the right endpoint of the active interval of L^{j-1} is at some point $x'' > x$, we show that x is included strictly in at most one of the active intervals of $L^1, \ldots, L^{j-1}$. For sake of contradiction, let us assume that x is included strictly in the active intervals of both L^{j-2} and L^{j-1}. Since the active interval of L^{j-2} extends on the right of x, there is no item $z \in Z_i$ with $v_l^{j-2} \leq z \leq x$. Hence, the item z^{j-2}, i.e., the item in Z closest to the median location v_{med}^{j-2} of L^{j-2}, is located at some point $z' \in [x, z^j]$. Since $|v_{\text{med}}^{j-2} - z^{j-2}| \leq \min_{y \in Y} |v_{\text{med}}^{j-2} - y|$, the item in Y closest to v_{med}^{j-2} is located at some point $y' \geq z'$ or at some point y' on the right of v_l^{j-2} with $|v_{\text{med}}^{j-2} - z^{j-2}| \leq |v_{\text{med}}^{j-2} - y'|$. Moreover, $z^{j-2} = z^{j-1} = z'$, i.e., the point in Z closest to the median locations v_{med}^{j-2} and v_{med}^{j-1} is z'. Therefore the item in $Y \cup \{z^{j-1}\}$ closest to the leftmost and to the rightmost agent of L^{j-1} is z^{j-1}. Hence, $\beta^{j-1} = 0$, because $y_l = y_r = z^{j-1}$ for $Y \cup \{z^{j-1}\}$ and for L^{j-1}. As a result, the active interval of L^{j-1} is empty and does not include x. A straightforward generalization of the same argument implies that if x is included strictly in the active interval of some group $L^{j'}$, $j' < j - 1$, the active intervals of all groups $L^{j'+1}, \ldots, L^{j-1}$ are empty and do not include x. □

A.11 Missing Details from the Proof of Lemma 6

To establish (4), it suffices to show that for any $\alpha \geq 1$,

$$\alpha n_1 \gamma + (\alpha - 1) \sum_{j \in L_2} (z - v_j) + \alpha n_3 (\beta - \gamma) \leq \tfrac{3\alpha - 1}{4} \beta n$$

Since for each $j \in L_2$, $z - v_j \leq \delta$, we have that $\sum_{j \in L_2} (z - v_j) \leq n_2 \delta$. Therefore, since $\delta \leq (\beta - \gamma)/2$, we obtain that:

$$\alpha n_1 \gamma + (\alpha - 1) \sum_{j \in L_2} (z - v_j) + \alpha n_3 (\beta - \gamma) \leq \alpha n_1 \gamma + \tfrac{\alpha - 1}{2} n_2 (\beta - \gamma) + \alpha n_3 (\beta - \gamma)$$

Bounding the rhs of the previous inequality from above, we obtain that:

$$\begin{aligned} \alpha n_1 \gamma + \tfrac{\alpha - 1}{2} n_2 (\beta - \gamma) + \alpha n_3 (\beta - \gamma) &\leq \alpha n_1 \gamma + \tfrac{\alpha - 1}{2} (\beta - \gamma) n + \tfrac{\alpha + 1}{2} n_3 (\beta - \gamma) \\ &\leq \tfrac{(3\alpha - 1)(\beta - \gamma) + 2\alpha\gamma}{4} n \leq \tfrac{3\alpha - 1}{4} \beta n \end{aligned}$$

For the first inequality, we use that $n_2 + n_3 \leq n$. For the second inequality, we use that $n_1 \leq n/2$ and that $n_3 \leq n/2$. The last inequality holds because $\alpha \geq 1$.

Next, we consider the case where $z < v_1$. In this case, $\beta = v_n - z$. Since $v_{\text{med}} \geq v_1$, and $|z - v_{\text{med}}| \leq |y - v_{\text{med}}|$, we get that $y \geq z + \delta \geq v_1$. As in the former case, we assume wlog. that $y \leq v_n$ and that $2\delta \leq \beta$. Now, we let n_2 (resp. L_2) denote the number (resp. the set) of agents located in $[v_1, z_1 + \delta]$ and n_3 denote the number of agents in $(z + \delta, v_n]$. Again, the median agent is located in $[v_1, z + \delta]$ and $n_3 \leq n/2$. Working as in the previous case, we obtain that

$$U(Y) \leq n - \sum_{j \in L_2} (v_j - z)\,,$$

because each agent $j \in L_2$ has utility at most $1 - (v_j - z)$ for his favorite item in Y, while all the remaining agents have utility at most 1 for y. Similarly,

$$U(z) \geq n - \sum_{j \in L_2} (v_j - z) - n_3\beta\,,$$

because each agent $j \in L_2$ has utility $1-(v_j-z)$ for z and n_3 agents have utility at least $1-(v_n-z) = 1 - \beta$ for z.

We next show that for any $\alpha \geq 1$,

$$\alpha U(z) + \sum_{j \in L_2} (v_j - z) \geq \alpha n - \tfrac{3\alpha-1}{4}\beta n \tag{16}$$

As in the previous case, using $\alpha = (4-\beta)/(4-3\beta)$, we obtain that $(3\alpha - 1)\beta/4 = \alpha - 1$ which, combined with (16), implies that $\frac{4-\beta}{4-3\beta}U(z) \geq n - \sum_{j \in L_2}(v_j - z) \geq U(Y)$.

The proof of (16) is essentially identical to the proof of (4), if we use $n_1 = \gamma = 0$. We give the proof for completeness. As in the proof of (4), it suffices to show that for any $\alpha \geq 1$,

$$(\alpha - 1) \sum_{j \in L_2} (v_j - z) + \alpha n_3(\beta - \gamma) \leq \tfrac{3\alpha-1}{4}\beta n$$

Since for each $j \in L_2$, $v_j - z \leq \delta$, we have that $\sum_{j \in L_2}(v_j - z) \leq n_2\delta$. Therefore, since $\delta \leq \beta/2$, we obtain that:

$$(\alpha - 1) \sum_{j \in L_2} (v_j - z) + \alpha n_3\beta \leq \tfrac{\alpha-1}{2} n_2\beta + \alpha n_3\beta$$

Bounding the rhs of the previous inequality from above, we obtain that:

$$\tfrac{\alpha-1}{2} n_2\beta + \alpha n_3\beta \leq \tfrac{\alpha-1}{2}\beta n + \tfrac{\alpha+1}{2} n_3\beta \leq \tfrac{3\alpha-1}{4}\beta n$$

The first inequality follows from $n_2 + n_3 \leq n$ and the second inequality holds because $n_3 \leq n/2$. □

A.12 The Proof of Theorem 4

We proceed as in the proof of Theorem 3, until (7), with the only difference that each group N^j has either $\lceil |L|/q \rceil$ or $\lfloor |L|/q \rfloor$ agents. Thus, we obtain the following weaker form of (8),

$$\alpha U(Z_i) \geq \alpha|L| - \frac{(3\alpha - 1)|L|}{2q} - \frac{3\alpha - 1}{2} - \sum_{j=1}^{q} \sum_{j \in L_2^j} |z - v_j| = |L| - \sum_{j=1}^{q} \sum_{j \in L_2^j} |z - v_j| \geq U(Y_i)\,,$$

where for the equality we need to use $\alpha = (2q - 1 - q/|L|)/(2q - 3 - 3q/|L|)$. The lower bound on the approximation ratio is obtained by summing the inequality above for all slots $i \in [k]$. □

A.13 The Proof of Theorem 5

Required Graph Theoretic Notions. Before proving Theorem 5, we need to introduce some graph theoretic notions. Given a simple graph $G = (V, E)$, we denote by $n(G)$ and $m(G)$ the number of vertices and edges of G, respectively. The *closed neighborhood* of vertex v, denoted $N_G[v]$, is the set of adjacent vertices including v and the *closed degree* is its size, i.e. $d_G[v] = |N_G[v]|$. A *proper k-coloring* of G is a mapping c from V to $[k]$ such that $c(u) \neq c(v)$ as soon as $[u, v] \in E$. The problem of finding a proper k-coloring with smallest k (Coloring in short) is **NP**-hard in general [16], but is solvable in polynomial-time for some graph classes like chordal and interval (see [17] for more). An *interval graph* is the intersection graph of a family of intervals on the real line. It is a *proper interval graph* if there is an interval representation in which no interval properly contains another. In the same way, it is a *unit interval graph* if there is an interval representation of it in which all the intervals have the same length. It is well known that the last two classes of graphs coincide [17, 5]. A *perfect elimination order* (peo in short) of a graph $G = (V, E)$ is an ordering σ of the vertices V, i.e. σ is a bijection from $[n(G)]$ to V such that for every $i \leq n(G)$, the closed neighborhood of $\sigma(i)$ in the subgraph G_i induced by $\{\sigma(i), \ldots, \sigma(n(G))\}$ is a clique ($\sigma(i)$ is usually called *simplicial vertex*). It is well known that a graph is *chordal* iff it has a peo and a graph is a *unit interval graph* iff it has an ordering σ such that σ and σ_{inv} are peo where $\sigma_{inv}(i) = \sigma(n(G)+1-i)$, [5, 19]. Moreover, if G is connected, then for all $i \leq n(G)-1$, $[\sigma(i), \sigma(i+1)] \in E$. From now on, we assume $\sigma(i) = v_i$ for $i \leq n(G)$, G_i is the subgraph induced by $\{v_i, \ldots, v_{n(G)}\}$ and $d^i = d_{G_i}[v_i]$.

A generalization of the decision version of Coloring where some vertices have a given color and the goal is to extend it to a proper coloring, has been studied in the literature under the name PreColoring Extension [3].

PreColoring Extension	
Input:	A simple graph $G = (V, E)$, a subset $W \subset V$, a coloring c' of W.
Parameter:	An integer k.
Question:	Deciding whether there is a proper k-coloring c of G extending c', i.e. $c(v) = c'(v)$ for $v \in W$.

Obviously, when $W = \emptyset$ we obtain the decision version of Coloring. PreColoring Extension has been proved **NP**-complete in interval graphs even if each precolored set has size at most 2 [3] and in unit interval graphs [20]. We are now ready to prove Theorem 5.

Proof (of Theorem 5). We propose a reduction from PreColoring Extension in unit interval graphs proved **NP**-complete in [20]. Let $G = (V, E)$ be a connected unit interval graph, a subset $W \subset V$, a coloring c' of W and an integer k'. We assume $V = \{v_1, \ldots, v_{n(G)}\}$ where $\sigma(i) = v_i$ as indicated previously, G_i is the subgraph induced by $\{v_i, \ldots, v_{n(G)}\}$ and $d^i = d_{G_i}[v_i]$ for $i \leq n(G)$. We build an instance of pre-CPD as follows:

- $n = n(G)$ agents $L = \{1, \ldots, n\}$ and n items $X = \{x_1, \ldots, x_n\}$;
- two positive integers $k = k'$ and $q = n$;
- $X' = \{x_i : v_i \in W\}$, a mapping $g(x_i) = c'(v_i)$ for $x_i \in X'$;
- a utility function for each agent defined by $u_i(x_\ell) = 1$ iff $v_\ell \in N_{G_i}[v_i]$ and $u_i(x_\ell) = 0$ otherwise.

Note that the agent preferences are dichotomous and single peaked. Actually, let $\ell_i = \max\{\ell : v_\ell \in N_{G_i}[v_i]\}$ for $i \leq n$; we have that $u_i(x_j) = 1$ iff $v_j \in I_i = \{v_i, \ldots, v_{\ell_i}\}$ because on the one hand, $[v_i, v_{i+1}] \in E$ (G is assumed to be connected) and on the other hand $N_{G_i}[v_i]$ is a clique of G_i.

We claim that the answer of PreColoring Extension is `yes` iff there exists a solution S made of k disjoint subsets $S_1, \ldots, S_k$ of X with utilitarian social welfare $U(S) = \sum_{\ell \in L} u_\ell(S) \geq n(G) + m(G)$.

If c is a proper k-coloring of G extending c', then by setting $S_i = \{x_\ell \in X : c(v_\ell) = i\}$ for $i \leq k$, we obtain a solution S with $U(S) = \sum_{\ell \in L} u_\ell(S) = n(G) + m(G)$ because $k = k' \geq max_{i \leq n(G)} d^i$ (recall that $N_{G_i}[v_i]$ is a clique of size d^i and it is a yes-instance of PreColoring Extension). Hence $U(S) = \sum_{\ell \in L} u_\ell(S) = \sum_{i \leq n} d^i = n(G) + m(G)$ and $g^{-1}(\ell) = \{x \in X' : g(x) = \ell\} \subseteq S_\ell$.

Conversely, assume that there is a collection S of k disjoint subsets $S_1, \dots, S_k$ of X with $U(S) \geq n(G) + m(G)$ and $g^{-1}(i) \subseteq S_i$. We also have $U(S) = \sum_{\ell \in L} u_\ell(S_\ell) \leq \sum_{i \leq n(G)} d^i = n(G) + m(G)$ because there are n agents and agent ℓ approves at most d^ℓ items. Thus, we deduce $U(S) = n(G)+m(G)$ and then $u_\ell(S_\ell) = d^\ell$ or equivalently each agent approves all these items. In conclusion, by setting $c(v_\ell) = i$ iff $x_\ell \in S_i$, c is a k-coloring extending c'. □

A.14 Missing Details from the Proof of Theorem 6

To compute these values required by the dynamic programming algorithm in the proof of Theorem 6, we start with the basis, where $U(0, (i_1, t_1), \dots, (i_k, t_k)) = 0$, for all pairs $(i_1, t_1), \dots, (i_k, t_k)$.

Then, we compute the values $U(1, (i_1, t_1), \dots, (i_k, t_k))$ for all pairs $(i_1, t_1), \dots, (i_k, t_k)$ such that $t_1 + \cdots + t_k \leq 1$. Specifically, we have that

$$U(1, (i_1, 0), \dots, (i_p, 1), \dots, (i_k, 0)) = \sum_{\nu=1}^{i_p} u_\nu(x_1)\,,$$

and that

$$U(1, (i_1, 0), \dots, (i_p, 0), \dots, (i_k, 0)) = 0\,.$$

Next, we proceed to compute the values $U(2, (i_1, t_1), \dots, (i_k, t_k))$ for all pairs $(i_1, t_1), \dots, (i_k, t_k)$ such that $t_1 + \cdots + t_k \leq 2$. Specifically, we have that

$$U(2, (i_1, 0), \dots, (i_p, 2), \dots, (i_k, 0)) = \max_{1 \leq \ell \leq i_p} \left\{ U(1, (i_1, 0), \dots, (\ell, 1), \dots, (i_k, 0)) + \sum_{\nu=\ell+1}^{i_p} u_\nu(x_2) \right\},$$

which is exactly as in the proof of [27, Theorem 6]. Moreover,

$$U(2, \dots, (i_{p_1}, 1), \dots, (i_{p_2}, 1), \dots) = \max \begin{Bmatrix} U(1, \dots, (i_{p_1}, 1), \dots, (i_{p_2}, 0), \dots) + \sum_{\nu=1}^{i_{p_2}} u_\nu(x_2) \\ U(1, \dots, (i_{p_1}, 0), \dots, (i_{p_2}, 1), \dots) + \sum_{\nu=1}^{i_{p_1}} u_\nu(x_2) \end{Bmatrix}$$

$$U(2, (i_1, 0), \dots, (i_p, 1), \dots, (i_k, 0)) = \max \left\{ U(1, (i_1, 0), \dots, (i_p, 1), \dots, (i_k, 0)), \sum_{\nu=1}^{i_p} u_\nu(x_2) \right\}$$

and $U(2, (i_1, 0), \dots, (i_k, 0)) = 0$.

For each $j = 3, \dots, m$, we compute similarly the values $U(j, (i_1, t_1), \dots, (i_k, t_k))$, for all pairs $(i_1, t_1), \dots, (i_k, t_k)$ such that $t_1 + \cdots + t_k \leq j$. □