Proportional Approval Voting, Harmonic k-median, and Negative Association

Jarosław Byrka
University of Wrocław, Wrocław, Poland
jby@cs.uni.wroc.pl

Piotr Skowron
University of Warsaw, Warsaw, Poland
p.skowron@mimuw.edu.pl

Krzysztof Sornat
University of Wrocław, Wrocław, Poland
krzysztof.sornat@cs.uni.wroc.pl

Abstract

We study a generic framework that provides a unified view on two important classes of problems: (i) extensions of the k-median problem where clients are interested in having multiple facilities in their vicinity (e.g., due to the fact that, with some small probability, the closest facility might be malfunctioning and so might not be available for using), and (ii) finding winners according to some appealing multiwinner election rules, i.e., election system aimed for choosing representatives bodies, such as parliaments, based on preferences of a population of voters over individual candidates. Each problem in our framework is associated with a vector of weights: we show that the approximability of the problem depends on structural properties of these vectors. We specifically focus on the harmonic sequence of weights, since it results in particularly appealing properties of the considered problem. In particular, the objective function interpreted in a multiwinner election setup reflects to the well-known Proportional Approval Voting (PAV) rule.

Our main result is that, due to the specific (harmonic) structure of weights, the problem allows constant factor approximation. This is surprising since the problem can be interpreted as a variant of the k-median problem where we do not assume that the connection costs satisfy the triangle inequality. To the best of our knowledge this is the first constant factor approximation algorithm for a variant of k-median that does not require this assumption. The algorithm we propose is based on dependent rounding [Srinivasan, FOCS'01] applied to the solution of a natural LP-relaxation of the problem. The rounding process is well known to produce distributions over integral solutions satisfying *Negative Correlation* (NC), which is usually sufficient for the analysis of approximation guarantees offered by rounding procedures. In our analysis, however, we need to use the fact that the carefully implemented rounding process satisfies a stronger property, called *Negative Association* (NA), which allows us to apply standard concentration bounds for *conditional* random variables.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis

Keywords and phrases approximation algorithms, computational social choice, k-median, dependent rounding, negative association

Digital Object Identifier 10.4230/LIPIcs.ICALP.2018.26

Related Version A full version of the paper is available at `https://arxiv.org/abs/1704.02183`.

Funding Jarosław Byrka was supported by the National Science Centre, Poland, grant number 2015/18/E/ST6/00456. Piotr Skowron was supported by a Humboldt Research Fellowship

45th International Colloquium on Automata, Languages, and Programming (ICALP 2018).
Editors: Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella;
Article No. 26; pp. 26:1–26:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

for Postdoctoral Researchers. Krzysztof Sornat was supported by the National Science Centre, Poland, grant number 2015/17/N/ST6/03684.

Acknowledgements We thank Ola Svensson and Aravind Srinivasan for their helpful and insightful comments.

1 Introduction

This paper considers a general unified framework for two classes of problems: (i) extensions of the k-median problem where clients care about having multiple facilities in their vicinity, and (ii) finding winning committees according to a number of well-known, but hard-to-compute multiwinner election systems[1]. Let us first formalize our framework; we will discuss motivation and explain the relation to k-median and to multiwinner elections later on.

For a natural number $t \in \mathbb{N}$, by $[t]$ we denote the set $\{1, \dots, t\}$. Let $\mathcal{F} = \{F_1, \dots, F_m\}$ be the set of m facilities and let $\mathcal{D} = \{D_1, \dots, D_n\}$ be the set of n clients (demands). The goal is to pick a set of k facilities that altogether are most satisfying for the clients. Different clients can have different preferences over individual facilities – by $c_{i,j}$ we denote the *cost* that client D_j suffers when using facility F_i (this can be, e.g., the communication cost of client D_j to facility F_i, or a value quantifying the level of personal dissatisfaction of D_j from F_i). Following Yager [34], we use ordered weighted average (OWA) operators to define the cost of a client for a bundle of k facilities C. Formally, let $w = \big(w_1, \dots, w_k\big)$ be a non-increasing vector of k weights. We define the w-cost of a client D_j for a size-k set of facilities C as $w(C, j) = \sum_{i=1}^{k} w_i c_i^{\rightarrow}(C, j)$, where $c^{\rightarrow}(C, j) = (c_1^{\rightarrow}(C, j), \dots, c_k^{\rightarrow}(C, j)) = \mathrm{sort}_{\mathrm{ASC}}\big(\{c_{i,j} \colon F_i \in C\}\big)$ is a non-decreasing permutation of the costs of client D_j for the facilities from C. Informally speaking, the highest weight is applied to the lowest cost, the second highest weight to the second lowest cost, etc. In this paper we study the following computational problem.

▶ **Definition 1** (OWA k-MEDIAN). In OWA k-MEDIAN we are given a set $\mathcal{D} = \{D_1, \dots, D_n\}$ of clients, a set $\mathcal{F} = \{F_1, \dots, F_m\}$ of facilities, a collection of clients' costs $\big(c_{i,j}\big)_{i\in[m], j\in[n]}$, a positive integer k ($k \le m$), and a vector of k non-increasing weights $w = \big(w_1, \dots, w_k\big)$. The task is to compute a subset C of $\mathcal{F}$ that minimizes the value

$$w(C) = \sum_{j=1}^{n} w(C, j) = \sum_{j=1}^{n} \sum_{i=1}^{k} w_i c_i^{\rightarrow}(C, j).$$

Note that OWA k-MEDIAN with weights $(1, 0, 0, \dots, 0)$ is the k-MEDIAN problem. Sometimes the costs represent distances between clients and facilities. Formally, this means that there exists a metric space $\mathcal{M}$ with a distance function $d \colon \mathcal{M} \times \mathcal{M} \to \mathbb{R}_{\ge 0}$, where each client and each facility can be associated with a point in $\mathcal{M}$ so that for each $F_i \in \mathcal{F}$ and each $D_j \in \mathcal{D}$ we have $d(i, j) = c_{i,j}$. When this is the case, we say that the *costs satisfy the triangle inequality*, and use the terms "costs" and "distance" interchangeably. Then, we use the prefix METRIC for the names of our problems. E.g., by METRIC OWA k-MEDIAN we denote the variant of OWA k-MEDIAN where the costs satisfy the triangle inequality.

We are specifically interested in the following two sequences of weights:

[1] We note that multiwinner election rules have many applications beyond the political domain – such applications include finding a set of results a search engine should display [12], recommending a set of products a company should offer to its customers [25, 26], allocating shared resources among agents [29, 28], solving variants of segmentation problems [23], or even improving genetic algorithms [15].

(1) **harmonic:** $w_{\mathrm{har}} = \big(1, 1/2, 1/3, \ldots, 1/k\big)$. By Harmonic k-median we denote the OWA k-median problem with the harmonic vector of weights.
(2) ***p*-geometric:** $w_{\mathrm{geom}} = \big(1, p, p^2, \ldots, p^{k-1}\big)$, for some $p < 1$.

The two aforementioned sequences of weights, w_{har} and w_{geom}, have their natural interpretations, which we discuss later on (for instance, see Examples 3 and 4).

1.1 Motivation

In this subsection we discuss the applicability of the studied model in two settings.

Multiwinner Elections

Different variants of the OWA k-median problem are very closely related to the preference aggregation methods and multiwinner election rules studied in the computational social choice, in particular, and in AI, in general – we summarize this relation in Table 1 and in Figure 1. In particular, one can observe that each "median" problem is associated with a corresponding "winner" problem. Specifically, the k-median problem is known in computational social choice as the Chamberlin–Courant rule. Let us now explain the differences between the winner ("election") and the median ("facility location") problems:

1. The election problems are usually formulated as maximization problems, where instead of (negative) costs we have (positive) utilities. The two variants, the minimization (with costs) and the maximization (with utilities) have the same optimal solutions. Yet, there is a substantial difference in their approximability.
   Approximating the minimization variant is usually much harder. For instance, consider the Chamberlin–Courant (CC) rule which is defined by using the sequence of weights $(1, 0, 0, \ldots, 0)$. In the maximization variant standard arguments can be used to prove that a greedy procedure yields the approximation ratio of $(1 - 1/e)$. This stands in a sharp contrast to the case when the same rule is expressed as the minimization one; in such a case we cannot hope for virtually any approximation [30] (we extend this result in Theorem 21 in [6]). Approximating the minimization variant is also more desired. E.g., a $1/2$-approximation algorithm for (maximization) CC can effectively ignore half of the population of clients, whereas it was argued [30] that a 2-approximation algorithm for the minimization (if existed) would be more powerful. In this paper we study the harder minimization variant, and give the first constant-factor approximation algorithm for the minimization OWA-Winner with the harmonic weights.
2. In facility location problems it is usually assumed that the costs satisfy the triangle inequality. This relates to the previous point: since the problem cannot be well approximated in the general setting, one needs to make additional assumptions. One of our main results is showing that there is a k-median problem (OWA k-median with harmonic weights) that admits a constant-factor approximation without assuming that the costs satisfy the triangle inequality; this is the first known result of this kind.

The special case of Harmonic k-median where each cost belongs to the binary set $\{0, 1\}$ is equivalent to finding winners according to Proportional Approval Voting. The harmonic sequence $w_{\mathrm{har}} = (1, 1/2, 1/3, \ldots, 1/k)$ is in a way exceptional: indeed, PAV can be viewed as an extension of the well known D'Hondt method of apportionment (used for electing parliaments in many contemporary democracies) to the case where the voters can vote for individual candidates rather than for political parties [4]. Further, PAV satisfies several other appealing properties, such as extended justified representation [3]. This is one of the reasons why we are specifically interested in the harmonic weights. For more discussion on PAV and other approval-based rules, we refer the reader to the survey of Kilgour [22].

Table 1 The relation between the k-MEDIAN problems and the corresponding problems studied in AI, in particular in the computational social choice community.

k-median problem	election rule	comment
OWA k-MEDIAN	OWA-Winner [29]	Finding winners according to OWA-Winner rules is the maximization variant of OWA k-MEDIAN (utilities instead of costs).
	Thiele methods [33]	Thiele methods are OWA-Winner rules for 0/1 costs.
HARMONIC k-MEDIAN	PAV [33]	In PAV we assume the 0/1 costs. So far, only the maximization variant was considered in the literature.
k-MEDIAN	Chamberlin–Courant [9]	In CC, usually some specific form of utilities is assumed – different utilities have been considered, but always in the maximization variant (utilities instead of costs).

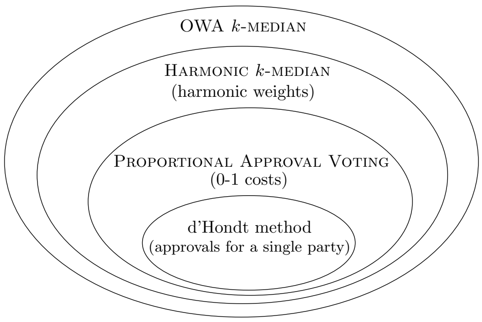

Figure 1 The relation between the considered models. OWA k-MEDIAN is the most general model. PROPORTIONAL APPROVAL VOTING and HARMONIC k-MEDIAN due to the use of harmonic weights can be viewed as natural extensions of the well known and commonly used D'Hondt method of apportionment [4].

OWA k-median as an Extension of k-median

Intuitively, our general formulation extends k-MEDIAN to scenarios where the clients not only use their most preferred facilities, but when there exists a more complex relation of "using the facilities" by the clients. Similar intuition is captured by the FAULT TOLERANT version of the k-MEDIAN problem introduced by Swamy and Shmoys [32] and recently studied by Hajiaghayi et al. [17]. There, the idea is that the facilities can be malfunctioning, and to increase the resilience to their failures each client needs to be connected to several of them.

▶ **Definition 2** (FAULT TOLERANT k-MEDIAN)**.** In FAULT TOLERANT k-MEDIAN problem we are given the same input as in k-MEDIAN, and additionally, for each client D_j we are given a natural number $r_j \geq 1$, called *the connectivity requirement.* The cost of a client D_j is the sum of its costs for the r_j closest open facilities. Similarly as in k-MEDIAN, we aim at choosing at most k facilities so that the sum of the costs is minimized.

When the values $(r_j)_{j \in [n]}$ are all the same, i.e., if $r_j = r$ for all j, then FAULT TOLERANT k-MEDIAN is called r-FAULT TOLERANT k-MEDIAN and it can be expressed as OWA k-MEDIAN for the weight vector w with r ones followed by $k - r$ zeros. Yet, in the typical setting of k-MEDIAN problems one additionally assumes that the costs between clients and facilities behave like distances, i.e., that they satisfy the triangle inequality. Indeed, the $(2.675+\epsilon)$-approximation algorithm for k-MEDIAN [5], the 93-approximation algorithm for FAULT TOLERANT k-MEDIAN [17], the 2-approximation algorithm for k-CENTER [18], and the 6.357-approximation algorithm for k-MEANS [1], they all use triangle inequalities. Moreover it can be shown by straightforward reductions from the SET COVER problem that there are no constant factor approximation algorithms for all these settings with general (non-metric) connection costs unless $\mathsf{P} = \mathsf{NP}$.

Using harmonic or geometric OWA weights is also well-justified in case of facility location problems, as illustrated by the following examples.

▶ **Example 3** (Harmonic weights: proportionality)**.** Assume there are $\ell \leq k$ cities, and for $i \in [\ell]$ let N_i denote the set of clients who live in the i-th city. For the sake of simplicity, let us assume that $k \cdot |N_i|$ is divisible by n. Further, assume that the cost of traveling between any two points within a single city is negligible (equal to zero), and that the cost of traveling between different cities is equal to one. Our goal is to decide in which cities the k facilities should be opened; naturally, we set the cost of a client for a facility opened in the same city to zero, and – in another city – to one. Let us consider OWA k-MEDIAN with the harmonic sequence of weights w_{har}. Let n_i denote the number of facilities opened in the i-th city in the optimal solution. We will show that for each i we have $n_i = \frac{k|N_i|}{n}$, i.e., that the number of facilities opened in each city is proportional to its population. Towards a contradiction assume there are two cities, i and j, with $n_i \geq \frac{k|N_i|}{n} + 1$ and $n_j \leq \frac{k|N_j|}{n} - 1$. By closing one facility in the i-th city and opening one in the j-th city, we decrease the total cost by at least:

$$|N_j| \cdot w_{n_j+1} - |N_i| \cdot w_{n_i} = \frac{|N_i|}{n_j+1} - \frac{|N_i|}{n_i} > \frac{|N_j|n}{k|N_j|} - \frac{|N_i|n}{k|N_i|} = 0.$$

Since, we decreased the cost of the clients, this could not be an optimal solution. As a result we see that indeed for each i we have $n_i = \frac{k|N_i|}{n}$.

▶ **Example 4** (Geometric weights: probabilities of failures)**.** Assume that we want to select k facilities and that each client will be using his or her favorite facility only. Yet, when a client wants to use a facility, it can be malfunctioning with some probability p; in such a case the client goes to her second most preferred facility; if the second facility is not working properly, the client goes to the third one, etc. Thus, a client uses her most preferred facility with probability $1-p$, her second most preferred facility with probability $p(1-p)$, the third one with probability $p^2(1-p)$, etc. As a result, the expected cost of a client D_j for the bundle of k facilities C is equal to $w(C, j)$ for the weight vector $w = \big(1-p, (1-p)p, \ldots, (1-p)p^{k-1}\big)$. Finding a set of facilities, that minimize the expected cost of all clients is equivalent to solving OWA k-MEDIAN for the p-geometric sequence of weights (in fact, the sequence that we use is a p-geometric sequence multiplied by $(1-p)$, yet multiplication of the weight vector by a constant does not influence the structure of the optimal solutions).

1.2 Our Results and Techniques

Our main result is showing, that there exists a 2.3589-approximation algorithm for HARMONIC k-MEDIAN for general connection costs (not assuming triangle inequalities). This is in contrast to the innaproximability of most clustering settings with general connection costs.

Our algorithm is based on dependent rounding of a solution to a natural linear program (LP) relaxation of the problem. We use the *dependent rounding* (DR) studied by Srinivasan et al. [31, 16], which transforms in a randomized way a fractional vector into an integral one. The sum-preservation property of DR ensures that exactly k facilities are opened.

DR satisfies, what is well known as *negative correlation (NC)* – intuitively, this implies that the sums of subsets of random variables describing the outcome are more centered around their expected values than if the fractional variables were rounded independently. More precisely, negative correlation allows one to use standard concentration bounds such as the Chernoff-Hoeffding bound. Yet, interestingly, we find out that NC is not sufficient for our analysis in which we need a conditional variant of the concentration bound. The property that is sufficient for conditional bounds is *negative association* (NA) [20]. In fact its special case that we call *binary negative association* (BNA), is sufficient for our analysis. It captures the capability of reasoning about conditional probabilities. Thus, our work demonstrates how to apply the (B)NA property in the analysis of approximation algorithms based on DR. To the best of our knowledge, Harmonic k-median is the first natural computational problem, where it is essential to use BNA in the analysis of the algorithm.

We additionally show that the 93-approximation algorithm of Hajiaghayi et al. [17] can be extended to OWA k-median (our technique is summarized in Section 3) – this time we additionally need to assume that the costs satisfy the triangle inequality. Indeed, without this assumption the problem is hard to approximate for a large class of weight vectors; for instance, for p-geometric sequences with $p < 1/e$ or for sequences where there exists $\lambda \in (0, 1)$ such that clients care only about the λ-fraction of opened facilities. Due to space constraints the formulation and the discussion on these hardness results are redelegated to the full version of the paper [6, Appendix E].

For the paper to be self-contained in [6, Appendix A] we discuss in detail the process of dependent rounding (including a few illustrative examples); in particular, we provide an alternative proof that DR satisfies binary negative association. Our proof is more direct and shorter than the proofs known in the literature [24].

2 Harmonic k-median and Proportional Approval Voting: a 2.3589-approximation Algorithm

In this section we demonstrate how to use the Binary Negative Association (BNA) property of Dependent Rounding (DR) to derive our main result – a randomized constant-factor approximation algorithm for Harmonic k-median. In [6, Appendix A] we provide a detailed discussion on DR and BNA, including a proof that DR satisfies BNA, and several examples.

▶ **Theorem 5.** *There exists a polynomial time randomized algorithm for* Harmonic k-median *that gives* 2.3589*-approximation in expectation.*

▶ **Corollary 6.** *There exists a polynomial time randomized algorithm for the minimization* Proportional Approval Voting *that gives* 2.3589*-approximation in expectation.*

In the remainder of this section we will prove the statement of Theorem 5. Consider the following linear program (1–5) that is a relaxation of a natural ILP for Harmonic k-median.

$$\min \sum_{j=1}^{n}\sum_{\ell=1}^{k}\sum_{i=1}^{m} w_\ell \cdot x_{ij}^\ell \cdot c_{ij} \qquad (1)$$

$$\sum_{i=1}^{m} y_i = k \qquad (2)$$

$$\sum_{\ell=1}^{k} x_{ij}^\ell \leq y_i \qquad \forall i \in [m],\ j \in [n] \qquad (3)$$

$$\sum_{i=1}^{m} x_{ij}^\ell \geq 1 \qquad \forall j \in [n],\ \ell \in [k] \qquad (4)$$

$$y_i, x_{ij}^\ell \in [0, 1] \qquad \forall i \in [m],\ j \in [n],\ \ell \in [k] \qquad (5)$$

The intuitive meaning of the variables and constraints of the above LP is as follows. Variable y_i denotes how much facility F_i is opened. Integral values 1 and 0 correspond to, respectively, opening and not opening the i-th facility. Constraint (2) encodes opening exactly k facilities. Each client $D_j \in \mathcal{D}$ has to be assigned to each among k opened facilities with different weights. For that we copy each client k times: the ℓ-th copy of a client D_j is assigned to the ℓ-th closest to D_j open facility. Variable x_{ij}^{ℓ} denotes how much the ℓ-th copy of D_j is assigned to facility F_i. In an integral solution we have $x_{ij}^{\ell} \in \{0, 1\}$, which means that the ℓ-th copy of a client can be either assigned or not to the respective facility. The objective function (1) encodes the cost of assigning all copies of all clients to the opened facilities, applying proper weights. Constraint (3) prevents an assignment of a copy of a client to a not-opened part of a facility. In an integer solution it also forces assigning different copies of a client to different facilities. Observe that, due to non-increasing weights w_ℓ, the objective (1) is smaller if an ℓ'-th copy of a client is assigned to a closer facility than an ℓ''-th copy, whenever $\ell' < \ell''$. Constraint (4) ensures that each copy of a client is served by some facility.

Just like in most facility location settings it is crucial to select the facilities to open, and the later assignment of clients to facilities can be done optimally by a simple greedy procedure. We propose to select the set of facilities in a randomized way by applying the DR procedure to the y vector from an optimal fractional solution to linear program (1–5). This turns out to be a surprisingly effective methodology for Harmonic k-median.

2.1 Analysis of the Algorithm

Let $\mathrm{OPT}^{\mathrm{LP}}$ be the value of an optimal solution (x^*, y^*) to the linear program (1–5). Let OPT be the value of an optimal solution $(x^{\mathrm{OPT}}, y^{\mathrm{OPT}})$ for Harmonic k-median. Easily we can see that $(x^{\mathrm{OPT}}, y^{\mathrm{OPT}})$ is a feasible solution to the linear program (1–5), so $\mathrm{OPT}^{\mathrm{LP}} \le \mathrm{OPT}$. Let $Y = (Y_1, \ldots, Y_m)$ be the random solution obtained by applying the DR procedure described in [6, Appendix A] to the vector y^*. Recall that DR preserves the sum of entries (see [6, Appendix A]), hence we have exactly k facilities opened. It is straightforward to assign clients to the open facilities, so the variables $X = (X_{ij}^{\ell})_{j\in[n], i\in[m], \ell\in[k]}$ are easily determined.

We will show that $\mathbb{E}[\mathrm{cost}(Y)] \le 2.3589 \cdot \mathrm{OPT}^{\mathrm{LP}}$. In fact, we will show that $\mathbb{E}[\mathrm{cost}_j(Y)] \le 2.3589 \cdot \mathrm{OPT}_j^{\mathrm{LP}}$, where the subindex j extracts the cost of assigning client D_j to the facilities in the solution returned by the algorithm. In our analysis we focus on a single client $D_j \in \mathcal{D}$. Next, we reorder the facilities $\{F_1, F_2, \ldots, F_m\}$ in the non-decreasing order of their connection costs to D_j (i.e., in the non-decreasing order of c_{ij}). Thus, from now on, facility F_i is the i-th closest facility to client D_j; ties are resolved in an arbitrary but fixed way.

The ordering of the facilities is depicted in Figure 2, which also includes information about the fractional opening of facilities in y^*, i.e., facility F_i is represented by an interval of length y_i^*. The total length of all intervals equals k. Next, we subdivide each interval into a set of (small) ϵ-size pieces (called ϵ-subintervals); ϵ is selected so that $1/\epsilon$, and y_i^*/ϵ for each i, are integers. Note that the values y_i^*, which originate from the solution returned by an LP solver, are rational numbers. The subdivision of $[0, k]$ into ϵ-subintervals is shown in Figure 2 on the "$(Z_r)_{r\in\{1,2,\ldots,k/\epsilon\}}$" level.

The idea behind introducing the ϵ-subintervals is the following. Although computationally the algorithm applies DR to the y^* variables, for the sake of the analysis we may think that the DR process is actually rounding z variables corresponding to ϵ-subinterval under the additional assumption that rounding within individual facilities is done before rounding between facilities. Formally, we replace the vector $Y = (Y_1, Y_2, \ldots, Y_m)$ by an equivalent vector of random variables $Z = (Z_1, Z_2, \ldots, Z_{k/\epsilon})$. Random variable Z_r represents the r-th ϵ-subinterval. We will use the following notation to describe the bundles of ϵ-subintervals

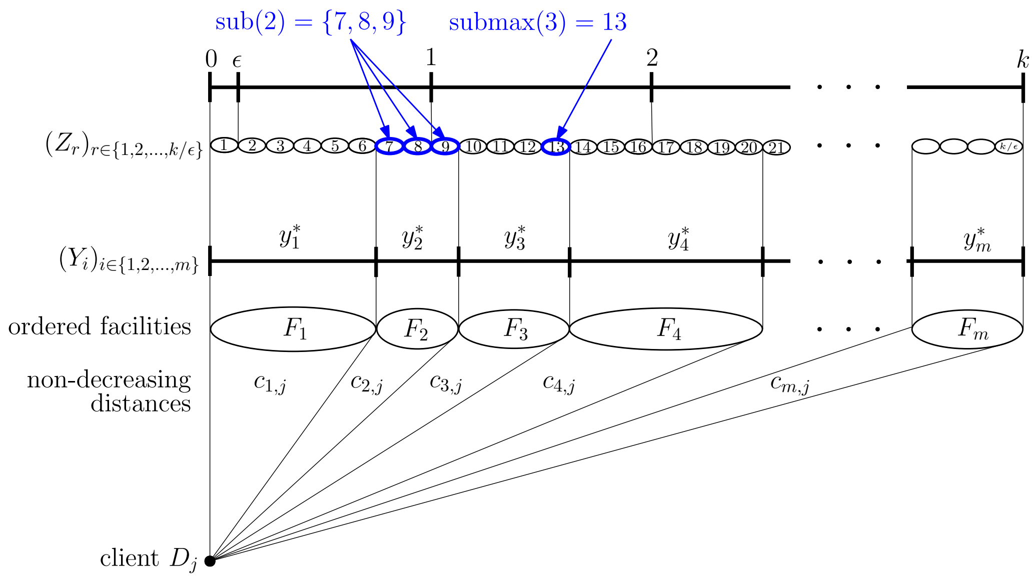

Figure 2 Ordering of the facilities by $c_{i,j}$ for the chosen client D_j. Definitions of the variables Y_i, Z_r and of the indices $\mathrm{sub}(i)$ and $\mathrm{submax}(i)$.

that correspond to particular facilities:

$$\mathrm{submax}(0) = 0 \qquad \text{and} \qquad \mathrm{submax}(i) = \mathrm{submax}(i-1) + \frac{y_i^*}{\epsilon}, \tag{6}$$

$$\mathrm{sub}(i) = \{\mathrm{submax}(i-1)+1, \ldots, \mathrm{submax}(i)\}. \tag{7}$$

Intuitively, $\mathrm{sub}(i)$ is the set of indexes r such that Z_r represents an interval belonging to the i-th facility. Examples for both definitions are shown in Figure 2 in the upper level. Formally, the random variables Z_r are defined so that:

$$Y_i = \sum_{r\in\mathrm{sub}(i)} Z_r \qquad \text{and} \qquad Y_i = 1 \implies \exists!\ r \in \mathrm{sub}(i) \quad Z_r = 1. \tag{8}$$

For each $r \in \{1, 2, \ldots, {}^{k}\!/\!_{\epsilon}\}$ we can write that:

$$\Pr[Z_r = 1] = \Pr[Z_r = 1 \big| Y_{\mathrm{sub}^{-1}(r)} = 1] \cdot \Pr[Y_{\mathrm{sub}^{-1}(r)} = 1] = \frac{\epsilon}{y^*_{\mathrm{sub}^{-1}(r)}} \cdot y^*_{\mathrm{sub}^{-1}(r)} = \epsilon \tag{9}$$

and $\Pr[Z_r = 0] = 1 - \epsilon$, hence $\mathbb{E}[Z_r] = \epsilon$. Also we have:

$$\Pr\left[Y_i = 1\right] = \Pr\left[\sum_{r\in\mathrm{sub}(i)} Z_r = 1\right] = \Pr\left[\bigvee_{r\in\mathrm{sub}(i)} Z_r = 1\right] = \sum_{r\in\mathrm{sub}(i)} \Pr\left[Z_r = 1\right]. \tag{10}$$

When $Y_i = 1$ its representative is chosen randomly among $(Z_r)_{r\in\mathrm{sub}(i)}$ independently of the choices of representatives of other facilities. Therefore

$$\forall_{i\in[m]} \quad \forall_{r\in\mathrm{sub}(i)} \quad \mathbb{E}\left[f(Y) \mid Y_i = 1\right] = \mathbb{E}\left[f(Y) \mid Y_i = 1 \ \wedge\ Z_r = 1\right], \tag{11}$$

for any function f on vector $Y = (Y_1, Y_2, \ldots, Y_m)$. Now we are ready to analyze the expected cost for any client $D_j \in \mathcal{D}$. Here we use the special assumption on the harmonic weights.

$$
\begin{aligned}
\mathbb{E}[\mathrm{cost}_j(Y)] \le\ & \sum_{i=1}^{m} \left(\mathbb{E}\left[\frac{c_{ij}}{1+\sum_{i'=1}^{i-1} Y_{i'}} \middle| Y_i = 1 \right] \cdot \Pr\left[Y_i = 1\right] \right) \\
\overset{(10)}{=}\ & \sum_{i=1}^{m} \left(c_{ij} \cdot \mathbb{E}\left[\frac{1}{1+\sum_{i'=1}^{i-1} Y_{i'}} \middle| Y_i = 1 \right] \cdot \sum_{r \in \mathrm{sub}(i)} \Pr\left[Z_r = 1\right] \right) \\
=\ & \sum_{i=1}^{m} \left(c_{ij} \cdot \sum_{r \in \mathrm{sub}(i)} \mathbb{E}\left[\frac{1}{1+\sum_{i'=1}^{i-1} Y_{i'}} \middle| Y_i = 1 \right] \cdot \Pr\left[Z_r = 1\right] \right) \\
\overset{(11)}{=}\ & \sum_{i=1}^{m} \left(c_{ij} \cdot \sum_{r \in \mathrm{sub}(i)} \mathbb{E}\left[\frac{1}{1+\sum_{i'=1}^{i-1} Y_{i'}} \middle| Y_i = 1 \wedge Z_r = 1 \right] \cdot \Pr\left[Z_r = 1\right] \right) \\
\overset{(8),(9)}{=}\ & \sum_{i=1}^{m} \left(\epsilon \cdot c_{ij} \cdot \sum_{r \in \mathrm{sub}(i)} \mathbb{E}\left[\frac{1}{1+\sum_{r'=1}^{\mathrm{submax}(i-1)} Z_{r'}} \middle| Z_r = 1 \right] \right) \\
\overset{(8)}{=}\ & \sum_{i=1}^{m} \left(\epsilon \cdot c_{ij} \cdot \sum_{r \in \mathrm{sub}(i)} \mathbb{E}\left[\frac{1}{1+\sum_{r'=1}^{r-1} Z_{r'}} \middle| Z_r = 1 \right] \right)
\end{aligned}
\tag{12}
$$

W.l.o.g., assume that $\mathrm{OPT}_j^{\mathrm{LP}} > 0$. Hence the approximation ratio for any client D_j is

$$
\frac{\mathbb{E}[\mathrm{cost}_j(Y)]}{\mathrm{OPT}_j^{\mathrm{LP}}} \overset{(7),(12)}{\le} \frac{\sum_{r=1}^{k/\epsilon} \epsilon \cdot c_{\mathrm{sub}^{-1}(r),j} \cdot \mathbb{E}\left[\frac{1}{1+\sum_{r'=1}^{r-1} Z_{r'}} \middle| Z_r = 1 \right]}{\sum_{r=1}^{k/\epsilon} \epsilon \cdot c_{\mathrm{sub}^{-1}(r),j} \cdot \frac{1}{\lceil r\epsilon \rceil}} =
$$

note that $\mathrm{sub}^{-1}(r)$ is an index of a facility that contains Z_r. Now we convert the sum over facilities into a sum over unit intervals. A unit interval is represented as a sum of $1/\epsilon$ many ϵ-subintervals:

$$
= \frac{\sum_{\ell=1}^{k} \sum_{r=(\ell-1)/\epsilon+1}^{\ell/\epsilon} c_{\mathrm{sub}^{-1}(r),j} \cdot \mathbb{E}\left[\frac{1}{1+\sum_{r'=1}^{r-1} Z_{r'}} \middle| Z_r = 1 \right]}{\sum_{\ell=1}^{k} \sum_{r=(\ell-1)/\epsilon+1}^{\ell/\epsilon} c_{\mathrm{sub}^{-1}(r),j} \cdot \frac{1}{\ell}} \le
$$

W.l.o.g., we can assume that first interval has non-zero costs: $\sum_{r=1}^{1/\epsilon} c_{\mathrm{sub}^{-1}(r),j} > 0$, otherwise the LP pays 0 and our algorithm pays 0 in expectation on intervals from non-empty prefix of $(1, 2, \ldots, k)$. With this assumption we can take maximum over intervals:

$$
\overset{\text{Lemma 17 in [6]}}{\le} \max_{\ell \in [k]} \left(\frac{\sum_{r=(\ell-1)/\epsilon+1}^{\ell/\epsilon} c_{\mathrm{sub}^{-1}(r),j} \cdot \mathbb{E}\left[\frac{1}{1+\sum_{r'=1}^{r-1} Z_{r'}} \middle| Z_r = 1 \right]}{\sum_{r=(\ell-1)/\epsilon+1}^{\ell/\epsilon} c_{\mathrm{sub}^{-1}(r),j} \cdot \frac{1}{\ell}} \right) \le
$$

Costs $c_{\mathrm{sub}^{-1}(r),j}$ can be general and they could be hard to analyze. Therefore we would like to remove costs from the analysis. We will use Lemma 18 from [6] for which the technique of splitting variables Y_i into Z_r was needed. We are using the fact that the variables Z_r have the same expected values; otherwise the coefficient in front of the expected value would be $c_{ij} \cdot y_i^*$, i.e., not monotonic. Thus

$$\stackrel{\text{Lemma 18 in [6]}}{\leq} \max_{\ell \in [k]} \left(\epsilon \cdot \ell \cdot \sum_{r={(\ell-1)}/{\epsilon}+1}^{\ell/\epsilon} \mathbb{E}\left[\frac{1}{1+\sum_{r'=1}^{r-1} Z_{r'}} \middle| Z_r = 1 \right] \right). \tag{13}$$

Consider the expected value in the above expression for a fixed $r \in \{(\ell-1)/\epsilon + 1, \ldots, \ell/\epsilon\}$:

$$\begin{aligned} E_r &= \mathbb{E}\left[\frac{1}{1+\sum_{r'=1}^{r-1} Z_{r'}} \middle| Z_r = 1 \right] = \sum_{t=1}^{k} \frac{1}{t} \Pr\left[\sum_{r'=1}^{r-1} Z_{r'} = t-1 \middle| Z_r = 1 \right] = \\ &= \sum_{t=1}^{\ell} \frac{1}{t} \Pr\left[\sum_{r'=1}^{r-1} Z_{r'} = t-1 \middle| Z_r = 1 \right] + \sum_{t=\ell+1}^{k} \frac{1}{t} \Pr\left[\sum_{r'=1}^{r-1} Z_{r'} = t-1 \middle| Z_r = 1 \right]. \end{aligned} \tag{14}$$

For $t \in \{1, 2, \ldots, \ell\}$ we consider the conditional probability in the above expression, denote it by $p_r(t-1)$, and analyze the corresponding cumulative distribution function $H_r(t-1)$:

$$p_r(t-1) = \Pr\left[\sum_{r'=1}^{r-1} Z_{r'} = t-1 \middle| Z_r = 1 \right], \tag{15}$$

$$H_r(t-1) = \Pr\left[\sum_{r'=1}^{r-1} Z_{r'} \leq t-1 \middle| Z_r = 1 \right] = \sum_{t'=0}^{t-1} p_r(t'), \tag{16}$$

We continue the analysis of E_r:

$$\begin{aligned} E_r &\stackrel{(14),(15)}{=} \sum_{t=1}^{\ell} \frac{1}{t} p_r(t-1) + \sum_{t=\ell+1}^{k} \frac{1}{t} p_r(t-1) \\ &\stackrel{(16)}{=} H_r(0) + \sum_{t=2}^{\ell} \frac{1}{t} \left(H_r(t-1) - H_r(t-2) \right) + \sum_{t=\ell+1}^{k} \frac{1}{t} p_r(t-1) \\ &= H_r(0) + \sum_{t=2}^{\ell} \frac{1}{t} H_r(t-1) - \sum_{t=2}^{\ell} \frac{1}{t} H_r(t-2) + \sum_{t=\ell+1}^{k} \frac{1}{t} p_r(t-1) \\ &= \sum_{t=1}^{\ell} \frac{1}{t} H_r(t-1) - \sum_{t=1}^{\ell-1} \frac{1}{t+1} H_r(t-1) + \sum_{t=\ell+1}^{k} \frac{1}{t} p_r(t-1) \\ &= \sum_{t=1}^{\ell-1} \frac{1}{t} H_r(t-1) - \sum_{t=1}^{\ell-1} \frac{1}{t+1} H_r(t-1) + \frac{1}{\ell} H_r(\ell-1) + \sum_{t=l+1}^{k} \frac{1}{t} p_r(t-1) \\ &\leq \sum_{t=1}^{\ell-1} \left(\frac{1}{t} - \frac{1}{t+1} \right) H_r(t-1) + \frac{1}{\ell} \left(H_r(\ell-1) + \sum_{t=\ell+1}^{k} p_r(t-1) \right) \\ &= \sum_{t=1}^{\ell-1} \frac{1}{t(t+1)} H_r(t-1) + \frac{1}{\ell} \left(H_r(\ell-1) + \sum_{t=\ell+1}^{k} p_r(t-1) \right) \\ &\leq \sum_{t=1}^{\ell-1} \frac{1}{t(t+1)} H_r(t-1) + \frac{1}{\ell}. \end{aligned} \tag{17}$$

▶ **Lemma 7.** *For any $\ell \in [k]$, $t \in [\ell-1]$ and $r \in \{(\ell-1)/\epsilon + 1, (\ell-1)/\epsilon + 2, \ldots, \ell/\epsilon\}$ we have*

$$H_r(t-1) \leq e^{-r \cdot \epsilon} \cdot \left(\frac{e \cdot r \cdot \epsilon}{t} \right)^t.$$

The proof of Lemma 7 combines the use of the BNA property of variables $\{Z_1, Z_2, \ldots, Z_{k/\epsilon}\}$ with applications of Chernoff-Hoeffding bounds. Due to the space constraints, the proof is moved to the full version of the paper [6, Appendix C]. In the end, we get the following bound on the approximation ratio.

▶ **Lemma 8.** *For any $j \in [n]$ we have*

$$\frac{\mathbb{E}[\text{cost}_j(Y)]}{\text{OPT}_j^{\text{LP}}} \leq 2.3589.$$

A proof uses inequalities (13), (17) as well as Lemma 7 with an upper bound derived by an integral of the function $f_t(x) = e^{-x}$. We made numerical calculation for $\ell \in \{1, 2, \ldots, 88\}$ and for other case we used Stirling formula and Taylor series for e^{ℓ} to derive analytical upper bound. Full proof, including a plot of numericaly obtained values, is presented in [6, Appendix C].

3 OWA k-median with Costs Satisfying the Triangle Inequality

In this section we construct an algorithm for OWA k-MEDIAN with costs satisfying the triangle inequality. Thus, the problem we address in this section is more general than HARMONIC k-MEDIAN (i.e., the problem we have considered in the previous section) in a sense that we allow for arbitrary non-increasing sequences of weights. On the other hand, it is less general in a sense that we require the costs to form a specific structure (a metric).

In our approach we first adapt the algorithm of Hajiaghayi et al. [17] for FAULT TOLERANT k-MEDIAN so that it applies to the following, slightly more general setting: for each client D_j we introduce its multiplicity $m_j \in \mathbb{N}$ – intuitively, this corresponds to cloning D_j and co-locating all such clones in the same location as D_j. However, this will require a modification of the original algorithm for FAULT TOLERANT k-MEDIAN, since we want to allow the multiplicities $\{m_j\}_{D_j \in \mathcal{D}}$ to be exponential with respect to the size of the instance (otherwise, we could simply copy each client a sufficient number of times, and use the original algorithm of Hajiaghayi et al.).

Next, we provide a reduction from OWA k-MEDIAN to such a generalization of FAULT TOLERANT k-MEDIAN. The resulting FAULT TOLERANT k-MEDIAN WITH CLIENTS MULTIPLICITIES problem can be cast as the following integer program:

$$\begin{aligned} \min \quad & \sum_{j=1}^{n} \sum_{i=1}^{m} m_j \cdot x_{ij} \cdot c_{ij} & \sum_{i=1}^{m} x_{ij} &= r_j & \forall j &\in [n] \\ & & x_{ij} &\leq y_i & \forall i \in [m],\ j &\in [n] \\ & \sum_{i=1}^{m} y_i = k & y_i, x_{ij} &\in \{0,1\} & \forall i &\in [m] \\ & & m_j &\in \mathbb{N} & \forall j &\in [n] \end{aligned}$$

▶ **Theorem 9.** *There is a polynomial-time* 93*-approximation algorithm for* METRIC FAULT TOLERANT k-MEDIAN WITH CLIENTS MULTIPLICITIES.

Proof can be found in [6, Appendix D]. Consider reduction from OWA k-MEDIAN to FAULT TOLERANT k-MEDIAN WITH CLIENTS MULTIPLICITIES depicted on Figure 3.

▶ **Lemma 10.** *Let I be an instance of* OWA k-MEDIAN*, and let I' be an instance of* FAULT TOLERANT k-MEDIAN WITH CLIENTS MULTIPLICITIES *constructed from I through reduction from Figure 3. An α-approximate solution to I' is also an α-approximate solution to I.*

Proof can be found in [6, Appendix D].

▶ Reduction. Let us take an instance I of OWA k-MEDIAN $\left(\mathcal{D}, \mathcal{F}, k, w, \{c_{ij}\}_{F_i \in \mathcal{F}, D_j \in \mathcal{D}}\right)$ where $w_i = \frac{p_i}{q_i}, i \in [k]$ are rational numbers in the canonical form. We construct an instance I' of FAULT TOLERANT k-MEDIAN WITH CLIENTS MULTIPLICITIES with the same set of facilities and the same number of facilities to open, k. Each client $D_j \in \mathcal{D}$ is replaced with clients $D_{j,1}, D_{j,2}, \ldots, D_{j,k}$ with requirements $1, 2, \ldots, k$, respectively. For $Q = \prod_{r=1}^{k} q_r$, the multiples of the clients are defined as follows:

- $m_{j,\ell} = (w_\ell - w_{\ell+1}) \cdot Q$, for each $\ell \in [k-1]$, and
- $m_{j,k} = w_k \cdot Q$.

Figure 3 Reduction from OWA k-MEDIAN to FAULT TOLERANT k-MEDIAN WITH CLIENTS MULTIPLICITIES.

▶ **Corollary 11.** *There exists a 93-approximation algorithm for* METRIC OWA k-MEDIAN *that runs in polynomial time.*

4 Concluding Remarks and Open Questions

We have introduced a new family of k-median problems, called OWA k-MEDIAN, and we have shown that our problem with the harmonic sequence of weights allows for a constant factor approximation even for general (non-metric) costs. This algorithm applies to Proportional Approval Voting. In the analysis of our approximation algorithm for HARMONIC k-MEDIAN, we used the fact that the dependent rounding procedure satisfies Binary Negative Association.

We showed that any METRIC OWA k-MEDIAN can be approximated within a factor of 93 via a reduction to FAULT TOLERANT k-MEDIAN WITH CLIENTS MULTIPLICITIES. We also obtained that OWA k-MEDIAN with p-geometric weights with $p < 1/e$ cannot be approximated without the assumption of the costs being metric. The status of the non-metric problem with p-geometric weights with $p > 1/e$ remains an intriguing open problem.

Using approximation and randomized algorithms for finding winners of elections requires some comment. First, the multiwinner election rules such as PAV have many applications in the voting theory, recommendation systems and in resource allocation. Using (randomized) approximation algorithms in such scenarios is clearly justified. However, even for other high-stake domains, such as political elections, the use of approximation algorithms is a promising direction. One approach is to view an approximation algorithm as a new, full-fledged voting rule (for more discussion on this, see the works of Caragiannis et al. [7, 8], Skowron et al. [30], and Elkind et al. [13]). In fact, the use of randomized algorithms in this context has been advocated in the literature as well – e.g., one can arrange an election where each participant is allowed to suggest a winning committee, and the best out of the suggested committees is selected; in such case the approximation guaranty of the algorithm corresponds to the quality of the outcome of elections (for a more detailed discussion see [30]) [2]. Nonetheless, we think that it would be beneficial to learn whether our algorithm can be efficiently derandomized.

[2] Indeed, approximation algorithms for many election rules have been extensively studied in the literature. In the world of single-winner rules, there are already very good approximation algorithms known for the Kemeny's rule [2, 10, 21] and for the Dodgson's rule [27, 19, 7, 14, 8]. A hardness of approximation has been proven for the Young's rule [7]. For the multiwinner case we know good (randomized) approximation algorithms for Minimax Approval Voting [11], Chamberlin–Courant rule [30], Monroe rule [30], or maximization variant of PAV [29].

References

1 S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward. Better Guarantees for k-Means and Euclidean k-Median by Primal-Dual Algorithms. In *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science*, pages 61–72, 2017. doi:10.1109/FOCS.2017.15.

2 N. Ailon, M. Charikar, and A. Newman. Aggregating Inconsistent Information: Ranking and Clustering. *Journal of the ACM*, 55(5):23:1–23:27, 2008. doi:10.1145/1411509.1411513.

3 H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh. Justified Representation in Approval-Based Committee Voting. *Social Choice and Welfare*, 48(2):461–485, 2017. doi:10.1007/s00355-016-1019-3.

4 M. Brill, J-F. Laslier, and P. Skowron. Multiwinner Approval Rules as Apportionment Methods. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 414–420, 2017. URL: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14782.

5 J. Byrka, T. Pensyl, B. Rybicki, A. Srinivasan, and K. Trinh. An Improved Approximation for k-Median and Positive Correlation in Budgeted Optimization. *ACM Transactions on Algorithms*, 13(2):23:1–23:31, 2017. doi:10.1145/2981561.

6 J. Byrka, P. Skowron, and K. Sornat. Proportional Approval Voting, Harmonic k-Median, and Negative Association. *CoRR*, abs/1704.02183v3, 2017.

7 I. Caragiannis, J. A. Covey, M. Feldman, C. M. Homan, C. Kaklamanis, N. Karanikolas, A. D. Procaccia, and J. S. Rosenschein. On the Approximability of Dodgson and Young Elections. *Artificial Intelligence*, 187:31–51, 2012. doi:10.1016/j.artint.2012.04.004.

8 I. Caragiannis, C. Kaklamanis, N. Karanikolas, and A. D. Procaccia. Socially Desirable Approximations for Dodgson's Voting Rule. *ACM Transactions on Algorithms*, 10(2):6:1–6:28, 2014. doi:10.1145/2556950.

9 B. Chamberlin and P. Courant. Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule. *American Political Science Review*, 77(3):718–733, 1983.

10 D. Coppersmith, L. Fleischer, and A. Rudra. Ordering by Weighted Number of Wins Gives a Good Ranking for Weighted Tournaments. *ACM Transactions on Algorithms*, 6(3):55:1–55:13, 2010. doi:10.1145/1798596.1798608.

11 M. Cygan, Ł. Kowalik, A. Socała, and K. Sornat. Approximation and Parameterized Complexity of Minimax Approval Voting. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 459–465, 2017 (full version will appear in Journal of Artificial Intelligence Research).

12 C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank Aggregation Methods for the Web. In *Proceedings of the 10th International World Wide Web Conference*, pages 613–622, 2001.

13 E. Elkind, P. Faliszewski, P. Skowron, and A.i Slinko. Properties of Multiwinner Voting Rules. *Social Choice and Welfare*, 48(3):599–632, 2017. doi:10.1007/s00355-017-1026-z.

14 P. Faliszewski, E. Hemaspaandra, and L. A. Hemaspaandra. Multimode Control Attacks on Elections. *Journal of Artificial Intelligence Research*, 40:305–351, 2011. doi:10.1613/jair.3136.

15 P. Faliszewski, J. Sawicki, R. Schaefer, and M. Smolka. Multiwinner Voting in Genetic Algorithms for Solving Ill-Posed Global Optimization Problems. In *Proceedings of the 19th International Conference on the Applications of Evolutionary Computation*, pages 409–424, 2016. doi:10.1007/978-3-319-31204-0_27.

16 R. Gandhi, S. Khuller, S. Parthasarathy, and A. Srinivasan. Dependent Rounding and Its Applications to Approximation Algorithms. *Journal of the ACM*, 53(3):324–360, 2006.

17 M. T. Hajiaghayi, W. Hu, J. Li, S. Li, and B. Saha. A Constant Factor Approximation Algorithm for Fault-Tolerant k-Median. *ACM Transactions on Algorithms*, 12(3):36:1–36:19, 2016. doi:10.1145/2854153.

18 D.S. Hochbaum and D.B. Shmoys. A Best Possible Heuristic for the k-Center Problem. *Mathematics of Operations Research*, 10(2):180–184, 1985.

19 C. M. Homan and L. A. Hemaspaandra. Guarantees for the Success Frequency of an Algorithm for Finding Dodgson-Election Winners. *Journal of Heuristics*, 15(4):403–423, 2009. doi:10.1007/s10732-007-9068-5.

20 K. Joag-Dev and F. Proschan. Negative Association of Random Variables with Applications. *The Annals of Statistics*, 11(1):286–295, 1983.

21 C. Kenyon-Mathieu and W. Schudy. How to Rank with Few Errors. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages 95–103, 2007. doi:10.1145/1250790.1250806.

22 D. Kilgour. Approval Balloting for Multi-Winner Elections. In J. Laslier and R. Sanver, editors, *Handbook on Approval Voting*, pages 105–124. Springer, 2010.

23 J. M. Kleinberg, C. H. Papadimitriou, and P. Raghavan. Segmentation Problems. *Journal of the ACM*, 51(2):263–280, 2004. doi:10.1145/972639.972644.

24 J. B. Kramer, J. Cutler, and A. J. Radcliffe. Negative Dependence and Srinivasan's Sampling Process. *Combinatorics, Probability and Computing*, 20(3):347–361, 2011.

25 T. Lu and C. Boutilier. Budgeted Social Choice: From Consensus to Personalized Decision Making. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 280–286, 2011. doi:10.5591/978-1-57735-516-8/IJCAI11-057.

26 T. Lu and C. Boutilier. Value-Directed Compression of Large-Scale Assignment Problems. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 1182–1190, 2015. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9557.

27 J. C. McCabe-Dansted, G. Pritchard, and A. M. Slinko. Approximability of Dodgson's rule. *Social Choice and Welfare*, 31(2):311–330, 2008. doi:10.1007/s00355-007-0282-8.

28 B. Monroe. Fully Proportional Representation. *American Political Science Review*, 89(4):925–940, 1995.

29 P. Skowron, P. Faliszewski, and J. Lang. Finding a Collective Set of Items: From Proportional Multirepresentation to Group Recommendation. *Artificial Intelligence*, 241:191–216, 2016. doi:10.1016/j.artint.2016.09.003.

30 P. Skowron, P. Faliszewski, and A. M. Slinko. Achieving Fully Proportional Representation: Approximability Results. *Artificial Intelligence*, 222:67–103, 2015. doi:10.1016/j.artint.2015.01.003.

31 A. Srinivasan. Distributions on Level-Sets with Applications to Approximation Algorithms. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 588–597, 2001.

32 C. Swamy and D.B. Shmoys. Fault-Tolerant Facility Location. *ACM Transactions on Algorithms*, 4(4):51:1–51:27, 2008.

33 T. N. Thiele. Om flerfoldsvalg. In *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger*, pages 415–441. København: A.F. Høst., 1895.

34 R. R. Yager. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking. *IEEE Trans. Systems, Man, and Cybernetics*, 18(1):183–190, 1988. doi:10.1109/21.87068.