On the Complexity of Calculating Approval-Based Winners in Candidates-Embedded Metrics

Yongjie Yang
Chair of Economic Theory, Saarland University, Saarbrücken, Germany
yyongjiecs@gmail.com

Abstract

We study approval-based multiwinner voting where candidates are in a metric space and committees are valuated in terms of their distances to the given votes. In particular, we consider three different distance functions, and for each of them we study both the utilitarian rules and the egalitarian rules, resulting in six variants of winners determination problems. We focus on the (parameterized) complexity of these problems for both the general metric and several special metrics. For hardness results, we also discuss their approximability.

1 Introduction

Approval-based multiwinner voting (ABMV) has absorbed considerable attention recently due to its wide applications in many areas [Aziz *et al.*, 2018; Brill *et al.*, 2020; Gawron and Faliszewski, 2019; Pierczynski and Skowron, 2019; Yang, 2019]. Canonical ABMV assumes that candidates are indistinguishable. However, in many applications candidates are correlated, driving researchers to consider more natural ABMV models where relations among candidates are taken into account. For instance, Izsak *et al*. [2018] considered ABMV with intraclass and interclass synergies, where a committee with some particular structure may produce extra utility according to some functions. Yang and Wang [2018] studied graph-based correlation among candidates, where candidates are represented by vertices in a graph and only committees inducing particular structures are eligible to be selected.

In this paper, we enrich this line of research by considering a model which captures the scenarios where candidates possess distance relationships defined within a metric space. The distance between two candidates is open to different interpretations, depending on concrete applications (e.g., it can be the real distance, similarity degree, or communication cost between candidates). This model is relevant for several real-world applications. For instance, imagine that several friends (or a travel agency, a performance group) are collectively making a decision on which k cities to travel together. The distance in the metric could be the physical distance, the traveling time, or the traveling cost among cities. Everyone has a set of cities which she would like to visit (determined due to, e.g., whether she has already visited a city before, the average expenditure of staying in a city, whether a city has some intriguing sights she wants to enjoy, etc.). Another example is as follows. Suppose that you want to borrow several books from the library of your university for your summer holidays. There are so many books that you can hardly determine which to borrow by yourself. Because you want to enhance your friendship with your friends or mates, or you want to better integrate into a reading club, you decide to ask some other people for recommendations. Books are classified into different types. Then, by defining the distances properly (e.g., books of the same type have distance one and books of different types have different distances larger than one) a platform deployed with our model can be useful.

Under our model, we consider rules which select winning k-committees minimizing either the sum of the distances (*utilitarian rules*) or the maximum distance (*egalitarian rules*) between the committee and the votes. To this end, we resort to three set functions to define the distances. For these rules, we study the complexity of the winners determination problems w.r.t. both the general metric and shortest-path metrics under paths, stars, and complete graphs. These special metrics are relevant to several applications, e.g., in the scenario where stations of a new railway line need to be collectively determined by a group of experts, or where some online resources are demanded to be deployed on several computers connected in a star-like network based on different criteria. For NP-hard problems, we also explore their approximation lower bounds and parameterized complexity. We offer a comprehensive landscape of the complexity of these problems. For our concrete results, we refer to Tables 1–2. We remark that as path-metrics are a special case of 1-dimensional Euclidean metrics, our hardness and inapproximability results obtained for path-metrics also hold for the case where candidates are embedded in the 1-dimensional Euclidean space.

2 Preliminaries

We assume the reader is familiar with the basics in (parameterized) complexity and graph theory [Downey, 2012; West, 2000]. For an undirected graph G and two vertices c and c' in G, $d_G(c, c')$ is the length of a shortest path between c and c' in G. A metric f over a set C is a *shortest-path metric* if there is an undirected graph G with vertex

metrics	utilitarian			egalitarian		
	minimum	maximum	sum	minimum	maximum	sum
general	NPC$^{0,\divideontimes}$ (Cor. 2)	NPC$^{\divideontimes}$ (Thm. 5)	**P** (Thm. 1)	NPC$^{0,\divideontimes}$ (Cor. 9)	**P** (Thm. 8)	NPC^{1,*} (Cor. 12, Thms. 13,15)
path	NPC$^{0,\divideontimes}$ (Cor. 2)	**NPC**$^{\divideontimes}$ (Thm. 5)	P (Thm. 1)	NPC$^{0,\divideontimes}$ (Cor. 9)	P (Thm. 8)	**NPC**^{1} (Thm. 13)
path*	**P** (Thm. 4)	**P** (Thm. 7)	P (Thm. 1)	**P** (Thm. 11)	P (Thm. 8)	$\boldsymbol{O^*(4^k)}$ (Thm. 20)
star/discrete	NPC$^{0,\divideontimes}$ (Cor. 2)	**NPC**$^{\divideontimes}$ (Thm. 5)	P (Thm. 1)	NPC$^{0,\divideontimes}$ (Cor. 9)	P (Thm. 8)	**NPC**^{1,*} (Cor. 12, Thms. 13,15)

utilitarian minimum	utilitarian maximum	egalitarian minimum	egalitarian sum
no FPT $\gamma(p)$-app (Cor. 3)	P $\boldsymbol{k}$**-app** (Thm. 6)	no FPT $\gamma(p)$-app (Cor. 10)	no P c-app for any $c < 2$ (Cor. 14)

Table 1: Our new important results are in boldface. Other results are either trivial or consequences of the boldfaced results. A problem labeled with $\divideontimes$ (resp. $\star$) means that it is W[2]-hard w.r.t. k (resp. $k+s$), labeled with 0 means that the hardness holds even if $s = 0$, and labeled with 1 means that the problem is NP-complete if $s = 1$ but becomes polynomial-time solvable if $s = 0$. Path* is a path-metric with the additional restriction that candidates approved in every vote are consecutive, p is the size of the input, and $\gamma(p)$ can be any computable function in p.

	utilitarian minimum	utilitarian maximum	egalitarian minimum	egalitarian sum
m	FPT (Cor. 16)	FPT (Cor. 16)	FPT (Cor. 16)	FPT (Cor. 16)
n	**FPT** (Thm. 17)	open	**FPT** (Thm. 18)	open
$n + \alpha(f)$	FPT (Thm. 17)	**FPT** (Thm. 19)	FPT (Thm. 18)	**FPT** (Thm. 19)

Table 2: A summary of our FPT-results. Our new important results are in boldface. Other results are either trivial or consequences of the boldfaced results. Here, $\alpha(f)$ is the number of different distances between candidates in C under f.

set C such that for all $c, c' \in C$, it holds that $f(c, c') = d_G(c, c')$. We call G a *graph-witness* of f. For simplicity, a path/star/complete-metric denotes a shortest-path metric with a path/star/complete graph witness. A complete-metric is often called a *discrete metric* in the literature.

In our model, there is a set C of *candidates*, a multiset V of *votes* over C cast by voters, and a *metric* $f : C \times C \to \mathbb{Q}_{\geq 0}$. Each vote over C is defined as a nonempty subset of C consisting of all candidates approved by the corresponding voter. A subset of exactly k candidates is called a *k-committee*. A *k-winners selection rule* (k-WSR) maps each (C, V, f) to a class of k-committees, the *winning committees*.

We study rules which select k-committees which are as close as possible to the votes, where the closeness is measured by certain *set functions* $g_f : 2^C \times 2^C \to \mathbb{Q}_{\geq 0}$. These rules can be categorized into two classes, namely the *utilitarian rules* and the *egalitarian rules*. Particularly, under a *g_f-utilitarian (resp. g_f-egalitarian) k-WSR*, *optimal* k-committees are those $w \subseteq C$ minimizing $\sum_{v \in V} g_f(v, w)$ (resp. $\max_{v \in V} g_f(v, w)$) among all k-committees of C. We define $g_f^{\circ}(v, c) = \min_{c' \in v} f(c', c)$ as the distance between vote v and candidate c. We study the following set functions.

- $g_f^{\text{min}}(v, w) = \min_{c \in w} g_f^{\circ}(v, c)$.
- $g_f^{\text{max}}(v, w) = \max_{c \in w} g_f^{\circ}(v, c)$.
- $g_f^{\text{sum}}(v, w) = \sum_{c \in w} g_f^{\circ}(v, c)$.

The above three functions respectively reflect scenarios where for a voter only her "best" candidate, her "worst" candidate, and all candidates in the winning committee matter.

Let X $\in \{\text{min}, \text{max}, \text{sum}\}$. We call $\sum_{v \in V} g_f^{\text{X}}(v, w)$ (resp. $\max_{v \in V} g_f^{\text{X}}(v, w)$) the utilitarian/egalitarian-X distance between w and V w.r.t. f. See Figure 1 for an illustration.

	a	b	c	d
a	0	4	5	3
b	4	0	8	2
c	5	8	0	6
d	3	2	6	0

	a	b	c	d
v_1	✓	✓	✓	
v_2		✓		
v_3			✓	✓
v_4	✓			✓

Figure 1: $C = \{a, b, c, d\}$ is a set of candidates. $V = \{v_1, v_2, v_3, v_4\}$ is a set of votes (see the right-handed matrix where a check mark means that the corresponding voter approves the corresponding candidate). A metric over C is represented by the left-handed matrix.

The utilitarian/egalitarian-X distances (U-X/E-X) between $\{a, b\}$ and V, and between $\{b, c\}$ and V in the example from Figure 1 are as follows:

between $\{a, b\}$ and V	U-min:	$0+0+2+0=2$	E-min:	2
	U-max:	$0+4+3+2=9$	E-max:	4
	U-sum:	$0+4+5+2=11$	E-sum:	5
between $\{b, c\}$ and V	U-min:	$0+0+0+2=2$	E-min:	2
	U-max:	$0+8+2+5=15$	E-max:	8
	U-sum:	$0+8+2+7=17$	E-sum:	8

In this paper, we study the following problems.

UTILITARIAN/EGALITARIAN-X WINNERS DETERMINATION (UWD-X/EWD-X)
Input: A set C of candidates, a multiset V of votes over C, a metric f over C, a nonnegative integer $k \leq |C|$, and a nonnegative rational number s.
Question: Is there a k-committee $w \subseteq C$ so that the utilitarian/egalitarian-X distance between w and V w.r.t. f is at most s?

For each X $\in \{\text{min}, \text{max}, \text{sum}\}$, OPT-UWD-X (resp.

OPT-EWD-X) denotes the optimization version of UWD-X (resp. EWD-X), where in the input we drop s, and the task is to find an optimal k-committee w.r.t. the utilitarian-X (resp. egalitarian-X) distance. Our hardness results proved in the paper are based on the following NP-complete problem [Gonzalez, 1985].

> RESTRICTED EXACT COVER BY 3 SETS (RX3C)
> **Input:** A universe H of cardinality 3κ for some positive integer κ, and a collection $\mathcal{B}$ of subsets of H such that every $B \in \mathcal{B}$ has cardinality 3 and every element of H is included in exactly three elements of $\mathcal{B}$.
> **Question:** Does H admit an exact set cover from $\mathcal{B}$, i.e., a subcollection $\mathcal{B}' \subseteq \mathcal{B}$ such that every element in H is in exactly one element of $\mathcal{B}'$?

Note that in every instance of RX3C it holds that $|\mathcal{B}| = |H|$. In addition, every exact set cover of H has cardinality κ.

3 Related Works

In this section, we discuss papers mostly related to our study.

Canonical ABMV. Clearly, the canonical ABMV can be incorporated into our model by ignoring the metric f. Many winners selection rules in this setting have been proposed, including the prevalent approval voting (AV), satisfaction approval voting (SAV) [Brams and Kilgour, 2014], Chamberlin-Courant approval voting (CCAV) [Chamberlin and Courant, 1983], proportional approval voting (PAV) [Thiele, 1895], etc. Generally, under these rules, each committee w receives a score given by a function of w and V, and k-committees with the maximum score are selected.[1] For instance, for AV and CCAV, the scores of a committee w are respectively $\sum_{v \in V} |v \cap w|$ and $|v \in V : v \cap w \neq \emptyset|$. We refer to [Lackner and Skowron, 2020] for a comprehensive survey.

For f being the discrete metrics, our k-WSRs are related to CCAV and AV in the following sense: CCAV and $g_f^{\min}$-utilitarian k-WSR (resp. AV and g_f^{sum}-utilitarian k-WSR) return exactly the same winning committees.

κ-Median, κ-Supplier, and κ-Center. In the metric κ-MEDIAN problem we are given a metric space (X, d), two subsets $P, Q \subseteq X$, and an integer κ, and the goal is to find a subset $P' \subseteq P$ of cardinality κ that minimizes $\sum_{q \in Q} \min_{p \in P'} d(p, q)$. If every voter approves exactly one candidate and every candidate is approved by exactly one voter, OPT-UWD-MIN becomes the special case of metric κ-MEDIAN where $P = Q = X$. It is well-known that metric κ-MEDIAN is polynomial-time approximatable to constant factors, with so far the best factor being $2.675 + \epsilon$ [Byrka *et al.*, 2017]. Moreover, there exists a tight $1 + 2/e + \epsilon$ approximation algorithm running in FPT-time w.r.t. κ [Cohen-Addad *et al.*, 2019]. When restricted to Euclidean space of fixed dimensions, κ-MEDIAN even admits a polynomial-time approximation scheme (PTAS) [Arora *et al.*, 1998]. When restricted to Euclidean space with unbounded dimensions, κ-MEDIAN becomes APX-hard [Guruswami and Indyk, 2003], and currently the best approximation factor achievable in polynomial time is 2.406 [Cohen-Addad *et al.*, 2022].

The metric κ-SUPPLIER problem has the same input as metric κ-MEDIAN, but the goal is to find a subset $P' \subseteq P$ of cardinality κ that minimizes $\max_{q \in Q} \min_{p \in P'} d(p, q)$. The metric κ-CENTER problem is a special case of metric κ-SUPPLIER where $P = Q = X$. It is easy to see that OPT-EWD-MIN generalizes metric κ-CENTER. It is known that metric κ-SUPPLIER admits a polynomial-time 3-approximation algorithm [Hochbaum and Shmoys, 1986]. In addition, metric κ-CENTER admits a polynomial-time 2-approximation algorithm, and unless P = NP this factor is the best possible even when restricted to Euclidean space of fixed dimensions [Gonzalez, 1985; Hochbaum and Shmoys, 1986; Hsu and Nemhauser, 1979]. It is even W[2]-hard to compute a $(2 - \epsilon)$-approximate solution for any $\epsilon > 0$ w.r.t. the parameter κ [Feldmann, 2019].

In contrast to the above approximability results, a reduction established by Betzler *et al.* [2013] implies that both OPT-UWD-MIN and OPT-EWD-MIN are W[2]-hard to approximate to any factor w.r.t. k, and this holds even when candidates are embedded in the 1-dimensional Euclidean space.

Spatial voting. Spatial voting is concerned with the setting where both candidates and voters are represented by points in a metric space (see, e.g., [Merrill, 1993; Duggan, 2007]). However, in our model, only candidates are embedded in a metric space but voters are not necessarily to be so. Moreover, spatial voting specifies the preferences of voters based on their distances to candidates, under the assumption that each voter prefers a closer candidate to a farther candidate. Then, based on the induced preferences, a canonical k-WSR (such as AV, PAV, CCAV, etc.) is used to select the winners. However, in our setting, we allow voters to have any possible dichotomous preferences. For comparison, consider a scenario where candidates are embedded in the 1-dimensional Euclidean space. Suppose that there are three candidates a, b, and c located at the points 0, 1, and 2, respectively. Then, in our model one can approve exactly a and c, but this cannot be the case in spatial voting. In addition, in our model winners are calculated based on distances between committees and the votes other than using canonical k-WSRs. A natural variant of spatial voting where the positions of candidates in the metric are publicly known, but those of voters are only private information has been studied in the literature (see, e.g., [Chen *et al.*, 2020; Anshelevich *et al.*, 2018; Gkatzelis *et al.*, 2020]). In this variant, voters provide ordinary preferences respecting their positions, and the goal is to compute a candidate that is as close as possible to voters' private positions.

4 Utilitarian Winners Determination

In this section, we study the complexity of computing winners of utilitarian rules w.r.t. both the general metric and some special metrics. We first observe that OPT-UWD-SUM has a straightforward polynomial-time algorithm: letting $\mathsf{sc}(c) = \sum_{v \in V} g_f^{\text{sum}}(v, \{c\})$ for each candidate $c \in C$, we sort elements of C in a descending order w.r.t. $\mathsf{sc}(c)$, and output the k-committee consisting of the first k candidates.

[1] There are also rules aimed to select committees that minimize some objectives (see, e.g., [Brams *et al.*, 2007]).

Theorem 1. OPT-UWD-SUM *can be solved in* $O(m \cdot \log m + m \cdot n))$ *time, where* m *and* n *denote the number of candidates and the number of votes, respectively.*

Now we study UWD-MIN. To show the W[2]-hardness of winners determination for CCAV w.r.t. the number of winners k, Betzler *et al.* [2013, Theorem 2] established a polynomial-time reduction which takes as input an instance of the W[2]-hard problem HITTING SET, and outputs a set of candidates C and a multiset of votes V so that the HITTING SET instance is a Yes-instance if and only if C contains a k-committee which intersects every vote in V. Observe that, regardless of the metric f, the utilitarian-min distance between a committee and V is 0 if and only if the committee intersects every vote in V. The work of Betzler *et al.* [2013] directly gives us the following two corollaries.

Corollary 2. *For all metrics,* UWD-MIN *is* NP-complete *and* W[2]-hard *w.r.t.* k. *This holds even if* $s = 0$.

Corollary 3. *For all metrics,* OPT-UWD-MIN *is* NP-hard *and* W[2]-hard *(w.r.t.* k*) to approximate to any factor.*

The above intractability results lead us to consider special cases of UWD-MIN. In this work, we study a special case where candidates are organized in a path and voters approve only consecutive candidates. This case is related to the domain of *voters interval* (VI) that has been extensively studied for canonical ABMV (see, e.g., [Betzler *et al.*, 2013; Peters, 2018; Elkind and Lackner, 2015; Liu and Guo, 2016]). Reiterate that a multiset of votes is VI if there is a linear order of the candidates so that every vote approves only consecutive candidates. Clearly, VI specifies only the order of candidates but not the distances among them. To fit our model, we regard the linear order in VI as a path.

For a path $P = (c_1, c_2, \ldots, c_m)$ over C, and a vote $v \subseteq C$, let $c(P, v, \mathrm{L})$ (resp. $c(P, v, \mathrm{R})$) be the candidate in v with the minimum (resp. maximum) subindex.

Theorem 4. UWD-MIN *restricted to path-metrics can be solved in* $O(m^6 \cdot n)$ *time when the candidates in every vote are consecutive in the path-witness of the given metric, where* m *and* n *are the numbers of candidates and votes, respectively.*

Proof. Let (C, V, f, k, s) be an instance of UWD-MIN, where f is a path-metric over C. Let $m = |C|$ and $n = |V|$. Let $P = (c_1, c_2, \ldots, c_m)$ be a path-witness of f so that each $v \in V$ induces a path in P. W.l.o.g., we assume that $m \geq k \geq 3$. For two integers x and y, let

$$C^{[x,y]} = \begin{cases} \{c_z \in C : x \leq z \leq y\} & 1 \leq x \leq y \leq m \\ \emptyset & \text{otherwise} \end{cases}$$

$$V^{[x,y]} = \begin{cases} \{v \in V : v \subseteq C^{[x,y]}\} & 1 \leq x \leq y \leq m \\ \emptyset & \text{otherwise} \end{cases}$$

We derive a dynamic programming algorithm by maintaining a 3-dimensional table $T(i, j, k')$ such that $1 \leq i \leq j \leq m$ and $1 \leq k' \leq k$. An entry $T(i, j, k')$ is presumed to store the utilitarian-min distance between $V^{[i,j]}$ and a k'-committee w' such that

(1) $c_i, c_j \in w'$ and $w' \subseteq C^{[i,j]}$; and

(2) $\sum_{v \in V^{[i,j]}} g_f^{\min}(v, w') \leq \sum_{v \in V^{[i,j]}} g_f^{\min}(v, w'')$ for all k'-committees w'' satisfying the first condition.

Let us call a k'-committee w' satisfying the above two conditions an optimal committee for $T(i, j, k')$. We fill out the table as follows.

- For all $i, j \in [m]$ and k' such that $i \leq j$ and $k' > j - i + 1$, let $T(i, j, k') = s + 1$.
- For all $i \in [m]$, let $T(i, i, 1) = 0$.
- For all $i, j \in [m]$ such that $i < j$, let $T(i, j, 2) = \sum_{v \in V^{[i,j]}} g_f^{\min}(v, \{c_i, c_j\})$ and let $T(i, j, 1) = s + 1$.

Now all entries with $k' \leq 2$ or $k' > j - i + 1$ are computed. We update other entries with $k' = 3, 4, \ldots, k$ as follows. Let $T(i, j, k')$ be an entry such that $i \in [j-2]$, $3 \leq j \leq m$, and $3 \leq k' \leq k$. We enumerate all pairs (i', j') of integers such that $i \leq i' \leq j' \leq j$ and $\{i', j'\} \neq \{i, j\}$. Each such a pair corresponds to a guess that there is an optimal committee for $T(i, j, k')$ which contains $c_{i'}$ and $c_{j'}$ but none of those in $C^{[i+1,i'-1]} \cup C^{[j'+1,j-1]}$. Let w' denote a k'-committee which contains all of c_i, c_j, $c_{i'}$, and $c_{j'}$, but none of $C^{[1,i-1]} \cup C^{[i+1,i'-1]} \cup C^{[j'+1,j-1]} \cup C^{[j+1,m]}$. For each enumerated (i', j'), we do the following. Let $b = |\{i, j\} \setminus \{i', j'\}|$. As $\{i, j\} \neq \{i', j'\}$, it holds that $b \in [2]$. Note that the utilitarian-min distance between w' and a vote approving at least one of c_i, c_j, $c_{i'}$, and $c_{j'}$ is 0. The utilitarian-min distance between w' and the votes in $V^{[i',j']}$ is indicated in $T(i', j', k' - b)$. Therefore, we need only to focus on the votes in $V^{[i+1,i'-1]} \cup V^{[j'+1,j-1]}$.

For two integers $x, y \in [m]$ such that $x \leq y$, let $\mathsf{sc}^{[x,y]} = \sum_{v \in V^{[x,y]}} g_f^{\min}(v, w')$ denote the utilitarian-min distance between w' and $V^{[x,y]}$. It is easy to see that the utilitarian-min distance between w' and $V^{[i+1,i'-1]}$ is $\mathsf{sc}^{[i+1,i'-1]} =$

$$\sum_{v \in V^{[i+1,i'-1]}} \min\{d_P(c(P, v, \mathrm{L}), c_i), d_P(c(P, v, \mathrm{R}), c_{i'})\},$$

and between w' and $V^{[j'+1,j-1]}$ is $\mathsf{sc}^{[j'+1,j-1]} =$

$$\sum_{v \in V^{[j'+1,j-1]}} \min\{d_P(c(P, v, \mathrm{L}), c_{j'}), d_P(c(P, v, \mathrm{R}), c_j)\}.$$

If w' is an optimal committee for $T(i, j, k')$, the utilitarian-min distance between w' and $V^{[i,j]}$ is then

$$\mathsf{sc}^{[i,j]} = \mathsf{sc}^{[i+1,i'-1]} + \mathsf{sc}^{[j'+1,j-1]} + T(i', j', k' - b).$$

We calculate $\mathsf{sc}^{[i,j]}$ for all enumerated pairs (i', j') and set $T(i, j, k')$ to be the minimum one among them.

After all entries are computed, we conclude that the given instance is a Yes-instance if and only if there exist i and j such that

$$T(i, j, k) + \sum_{v \in V^{[1,i-1]}} d_P(c(P, v, \mathrm{R}), c_i) + \sum_{v \in V^{[j+1,m]}} d_P(c(P, v, \mathrm{L}), c_j) \leq s.$$

Now we analyze the running time of the algorithm. To update an entry $T(i, j, k')$, we need to consider $O(m^2)$ pairs

(i', j') as discussed above. As calculating $\mathsf{sc}^{[i,j]}$ for each pair can be done in $O(m \cdot n)$-time, it takes $O(m^3 \cdot n)$ time to update the entry. Then, as we have $O(m^3)$ entries, the whole algorithm runs in $O(m^6 \cdot n)$ time. □

Now, we study UWD-MAX.

Theorem 5. UWD-MAX *is* NP-complete*, and is* W[2]-hard *w.r.t. k. This holds even when restricted to path-metrics, star-metrics, and discrete metrics.*

Unlike the minimum function, OPT-UWD-MAX can be approximated within factor k. In fact, every UWD-SUM optimal k-committee provides such an approximation solution.

Theorem 6. OPT-UWD-MAX *admits a polynomial-time k-approximation algorithm, where k denotes the size of the winning committee.*

However, if we further impose a particular restriction on the domain of path-metrics, the problem becomes tractable.

Theorem 7. UWD-MAX *restricted to path-metrics can be solved in $O(m^3 \cdot n)$ time if candidates approved by every vote are consecutive in the path-witness, where m denotes the number of candidates and n denotes the number of votes.*

Proof. Let (C, V, f, k, s) be an instance of UWD-MAX, where f is a path-metric over C with a path-witness P so that candidates in each vote are consecutive in P. Let $m = |C|$ and $n = |V|$. W.l.o.g., let us assume that $k \geq 2$ (otherwise, the instance can be solved easily). For a k-committee $w \subseteq C$, let $\widehat{C}(w)$ be the set of the two candidates $c, c' \in w$ such that all candidates in $w \setminus \{c, c'\}$ are between c and c' in the path-witness P. An observation is that the utilitarian-max distance between w and V can be determined by $\widehat{C}(w)$.

Observation. *For all $v \in V$ and all $w \subseteq C$ such that $|w| \geq 2$, it holds that $g_f^{\max}(v, w) = g_f^{\max}(v, \widehat{C}(w))$.*

By this observation, we enumerate all pairs $\{c, c'\} \subseteq C$ such that there are at least $k - 2$ candidates between them in the path-witness P. For each enumerated pair $\{c, c'\}$, we check if $\sum_{v \in V} g_f^{\max}(v, \{c, c'\}) \leq s$ holds, which can be done in $O(n \cdot m)$ time. If the inequality holds for at least one enumerated pair, we conclude that the given instance is a Yes-instance (due to the above analysis, every k-committee $w \subseteq C$ consisting of c, c' and any arbitrary $k - 2$ candidates between c and c' in P fulfills that $\sum_{v \in V} g_f^{\max}(v, w) \leq s$). Otherwise, we conclude that the instance is a No-instance.

As there are at most m^2 pairs to enumerate and checking if the above inequality holds for each of them can be done in $O(m \cdot n)$ time, the algorithm runs in $O(m^3 \cdot n)$ time. □

5 Egalitarian Winners Determination

In this section, we focus on egalitarian rules.

Theorem 8. EWD-MAX *can be solved in $O(m^2 \cdot n)$ time, where m and n are the number of candidates and the number of votes, respectively.*

Now we study egalitarian rules with the minimum function. Observe that a committee w is of utilitarian-min distance 0 from a multiset of votes V if and only if w is of egalitarian-min distance 0 from V. Then, we obtain the following two corollaries as a consequence of Corollary 2.

Corollary 9. *For all metrics,* EWD-MIN *is* NP-complete*, and is* W[2]-hard *w.r.t. k. This holds even if $s = 0$.*

Corollary 10. *For all metrics,* OPT-EWD-MIN *is* NP-hard *and* W[2]-hard *(w.r.t. k) to approximate to any factor.*

However, when restricted to a special case, the problem becomes polynomial-time solvable.

Theorem 11. EWD-MIN *restricted to path-metrics can be solved in $O(m^4 \cdot n)$ time if all votes approve only consecutive candidates in the path-witness, where m denotes the number of candidates and n denotes the number of votes.*

We are left with EWD-SUM. We show that, unlike the above two problems, a threshold value of s distinguishes the complexity of the problem. For $s = 0$, EWD-SUM and UWD-SUM (also EWD-MAX and UWD-MAX) are equivalent. Then, from Theorem 1, we have the following corollary.

Corollary 12. EWD-SUM *is polynomial-time solvable for the case where $s = 0$.*

However, we show that the complexity of the problem lifts as s increases to 1.

Theorem 13. EWD-SUM *is* NP-complete *even when restricted to path-metrics, star-metrics, and discrete metrics. Moreover, this holds even when $s = 1$.*

Proof for path-metrics. We construct a reduction from RX3C to EWD-SUM. Let $(H, \mathcal{B})$ be an RX3C instance where $|H| = |\mathcal{B}| = 3\kappa$. We assume that $\kappa \geq 3$. We create an EWD-SUM instance (C, V, f, k, s) as follows. For each $B \in \mathcal{B}$, we create one candidate $c(B)$. Let $C(\mathcal{B}) = \{c(B) : B \in \mathcal{B}\}$ be the set of these 3κ candidates. Additionally, we create a set D of $3\kappa + 1$ candidates. Let $C = C(\mathcal{B}) \cup D$. The path-witness of f is a path such that every candidate in $C(\mathcal{B})$ has exactly two neighbors from D. In other words, in the path, candidates in D and candidates in $C(\mathcal{B})$ occur alternatively beginning with someone from D. We create 6κ votes in V as follows. First, for each $h \in H$, we create one vote $v(h)$ which approves all candidates except those corresponding to the three 3-subsets including h, i.e., $v(h) = C \setminus \{c(B) : h \in B \in \mathcal{B}\}$. Then, for each $B \in \mathcal{B}$, we create one vote $v(B)$ which approves all candidates in $C(\mathcal{B})$ except $c(B)$, and approves arbitrarily one of the two neighbors of $c(B)$ in the path-witness and disapproves all the other candidates in D. We complete the reduction by setting $k = \kappa$ and $s = 1$. We refer to Figure 2 for an illustration of the reduction. In the following, we show that there is an exact set cover of H if and only if there is a k-committee of egalitarian-sum distance at most 1 from V.

($\Rightarrow$) Let $\mathcal{B}' \subseteq \mathcal{B}$ be an exact set cover of H. Consider the k-committee $w = \{c(B) : B \in \mathcal{B}'\}$ corresponding to $\mathcal{B}'$. Due to the construction, for each vote $v(h)$, where $h \in H$, there is exactly one candidate from w which is not approved in $v(h)$. In particular, this candidate is the one $c(B) \in w$ such that $h \in B \in \mathcal{B}'$. Note that $v(h)$ approves all candidates in D. As $c(B)$ has at least one neighbor from D in the path-witness, it follows that the distance between w and $v(h)$ is exactly 1. Now let us consider a vote $v(B)$ where $B \in \mathcal{B}$. Recall that $v(B)$ approves all candidates in $C(\mathcal{B})$ except only $c(B)$,

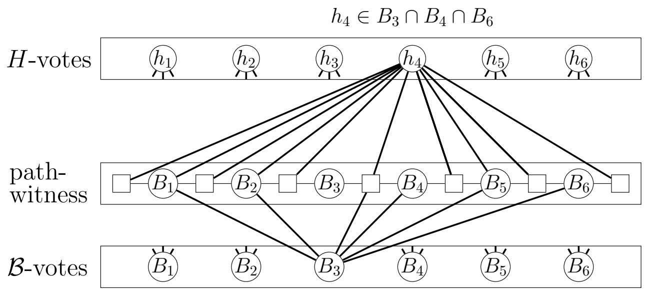

Figure 2: An illustration of the reduction in the proof of Theorem 13. An edge between a vote v and a candidate c means that $c \in v$. Square candidates (□) are those in D.

and approves one of the two neighbors of $c(B)$ in the path-witness. We distinguish between two cases. If $c(B) \notin w$, all candidates in w are approved by $v(B)$, and hence the distance between w and $v(B)$ is 0. Otherwise, $c(B)$ is the only one in w not approved by $v(B)$. As one of the neighbors of $c(B)$ is approved by $v(B)$, the distance between w and $v(B)$ is 1.

($\Leftarrow$) Assume that $w \subseteq C$ is a k-committee of egalitarian-sum distance at most 1 from V. We claim first that w does not include anyone from D.

Claim. $w \cap D = \emptyset$.

Proof of the claim. We first show that $|w \cap D| \leq 1$. For the sake of contradiction, assume that w contains at least two candidates from D. Then, as $|w \cap D| \leq |w| = \kappa$, $|D| = 3\kappa + 1$, and $\kappa \geq 3$, there must be at least one candidate $c(B)$, $B \in \mathcal{B}$, such that none of its two neighbors is in w. Then, according to the construction, the vote $v(B)$ approves none of $D \cap w$, implying that the distance between $v(B)$ and w is at least $|w \cap D| \geq 2$, a contradiction. In the following, we show that $|w \cap D| = 1$ is also impossible. For contradiction, assume that w contains exactly one candidate $d \in D$. Let $N(d)$ be the set of neighbors of d in the path-witness. As $|N(d)| \leq 2$ (if d is an end-vertex of the path-witness, $N(d)$ is a singleton; otherwise, $N(d)$ consists of exactly two candidates from $C(\mathcal{B})$) and $k = \kappa \geq 3$, the committee w must contain at least one candidate from $C(\mathcal{B}) \setminus N(d)$. Let $c(B)$ be such a candidate. Then, according to the construction, the vote $v(B)$ approves neither $c(B)$ nor d, implying that the distance between w and $v(B)$ is at least 2, a contradiction too.

Due to the above claim, we know that $w \subseteq C(\mathcal{B})$. Let $\mathcal{B}' = \{B \in \mathcal{B} : c(B) \in w\}$. Let $v(h)$ be a vote where $h \in H$. As $g_f^{\text{sum}}(v(h), w) \leq 1$, w contains at most one candidate not approved in $v(h)$. This means that there is at most one $B \in \mathcal{B}'$ such that $h \in B$. As this holds for all $h \in H$ and $|\mathcal{B}'| = \kappa$, it follows that $\mathcal{B}'$ is an exact set cover of H. □

In the above proof, if there is an exact set cover, there is a k-committee of egalitarian-sum distance at most 1 from V, otherwise the optimal k-committee is of egalitarian-sum distance at least 2 from V (this is also the case for star-metrics and discrete metrics). This leads to the following result.

Corollary 14. OPT-EWD-SUM *is* NP-hard *to* c-*approximate for any* $c < 2$ *even when restricted to path-metrics, star-metrics, and discrete metrics.*

Finally, we study EWD-SUM w.r.t. $k + s$.

Theorem 15. EWD-SUM *restricted to star-metrics and discrete metrics is* W[2]-hard *w.r.t.* $k + s$.

6 Fixed-Parameter Algorithms

In this section, we study FPT-algorithms for the winners determination problems. The first natural parameter is the number of candidates m. The following result is easy to see.

Corollary 16. UWD-X *and* EWD-X *where* $X \in \{\min, \max, \text{sum}\}$ *can be solved in* $O^*(2^m)$ *time.*

The second natural parameter is the number of votes n. For this parameter, we obtain the following results.

Theorem 17. UWD-MIN *can be solved in* $O^*(n^n)$ *time.*

Theorem 18. EWD-MIN *can be solved in* $O^*(2^{2^n})$ *time.*

Next, we study a combined parameter. For a metric f over C, let $\alpha(f) = |\{f(c, c') : c, c' \in C\}|$ be the number of different distances between candidates of C. Clearly, for a graph metric f whose graph-witness has diameter r, it holds that $\alpha(f) \leq r + 1$. For instance, for f being a star-metric or a discrete metric, $\alpha(f)$ is equal to 3 and 2, respectively.

Theorem 19. UWD-MAX *and* EWD-SUM *are* FPT *w.r.t. the combined parameter* $\alpha(f) + n$.

Finally, we study EWD-SUM restricted to path-metrics. If voters approve only consecutive candidates, we have shown that all the problems except EWD-SUM are polynomial-time solvable. Though we are unable to prove the same result for EWD-SUM, we show that this restriction makes EWD-SUM tractable from the perspective of parameterized complexity when parameterized by the number of winners.

Theorem 20. EWD-SUM *can be solved in* $O^*(4^k)$ *time when restricted to path-metrics and when the approved candidates of every vote are consecutive in the path-witness.*

7 Concluding Remarks

For future research, one could consider other set functions. A notable example is the Hausdorff function which has been successfully used in a broad range of areas (see, e.g., [Jesorsky *et al.*, 2001]). Additionally, it is interesting to consider a generalization of our model by adopting the notion of ordered weighted average (OWA) operation to derive the distances between votes and committees. Finally, we point out that in some scenarios it is nontrivial to precisely measure the distances between candidates. In this case, it is important to investigate the robustness of our model against the inaccuracy of distances among candidates.

References

[Anshelevich *et al.*, 2018] Elliot Anshelevich, Onkar Bhardwaj, Edith Elkind, John Postl, and Piotr Skowron. Approximating optimal social choice under metric preferences. *Artif. Intell.*, 264:27–51, 2018.

[Arora *et al.*, 1998] Sanjeev Arora, Prabhakar Raghavan, and Satish Rao. Approximation schemes for Euclidean k-medians and related problems. In *STOC*, pages 106–113, 1998.

[Aziz *et al.*, 2018] Haris Aziz, Edith Elkind, Shenwei Huang, Martin Lackner, Luis Sánchez-Fernández, and Piotr Skowron. On the complexity of extended and proportional justified representation. In *AAAI*, pages 902–909, 2018.

[Betzler *et al.*, 2013] Nadja Betzler, Arkadii Slinko, and Johannes Uhlmann. On the computation of fully proportional representation. *J. Artif. Intell. Res.*, 47:475–519, 2013.

[Brams and Kilgour, 2014] Steven J. Brams and D. Marc Kilgour. Satisfaction approval voting. In Rudolf Fara, Dennis Leech, and Maurice Salles, editors, *Voting Power and Procedures*, Studies in Choice and Welfare, pages 323–346. Springer, 2014.

[Brams *et al.*, 2007] Steven J. Brams, D. Marc Kilgour, and M. Remzi Sanver. A minimax procedure for electing committees. *Public Choice*, 132(3-4):401–420, 2007.

[Brill *et al.*, 2020] Markus Brill, Paul Gölz, Dominik Peters, Ulrike Schmidt-Kraepelin, and Kai Wilker. Approval-based apportionment. In *AAAI*, pages 1854–1861, 2020.

[Byrka *et al.*, 2017] Jarosław Byrka, Thomas Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for k-median and positive correlation in budgeted optimization. *ACM Trans. Algorithms*, 13(2):23:1–23:31, 2017.

[Chamberlin and Courant, 1983] John. R. Chamberlin and Paul N. Courant. Representative deliberations and representative decisions: Proportional representation and the Borda rule. *Am. Polit. Sci. Rev.*, 77(3):718–733, 1983.

[Chen *et al.*, 2020] Xujin Chen, Minming Li, and Chenhao Wang. Favorite-candidate voting for eliminating the least popular candidate in a metric space. In *AAAI*, pages 1894–1901, 2020.

[Cohen-Addad *et al.*, 2022] Vincent Cohen-Addad, Hossein Esfandiari, Vahab Mirrokni, and Shyam Narayanan. Improved approximations for Euclidean k-means and k-median, via nested quasi-independent sets. In *STOC*, 2022.

[Cohen-Addad *et al.*, 2019] Vincent Cohen-Addad, Anupam Gupta, Amit Kumar, Euiwoong Lee, and Jason Li. Tight FPT approximations for k-median and k-means. In *ICALP*, pages 42:1–42:14, 2019.

[Downey, 2012] Rod Downey. A parameterized complexity tutorial. In *LATA*, pages 38–56, 2012.

[Duggan, 2007] John Duggan. Equilibrium existence for zero-sum games and spatial models of elections. *Game. Econ. Behav.*, 60(1):52–74, 2007.

[Elkind and Lackner, 2015] Edith Elkind and Martin Lackner. Structure in dichotomous preferences. In *IJCAI*, pages 2019–2025, 2015.

[Feldmann, 2019] Andreas Emil Feldmann. Fixed-parameter approximations for k-center problems in low highway dimension graphs. *Algorithmica*, 81(3):1031–1052, 2019.

[Gawron and Faliszewski, 2019] Grzegorz Gawron and Piotr Faliszewski. Robustness of approval-based multiwinner voting rules. In *ADT*, pages 17–31, 2019.

[Gkatzelis *et al.*, 2020] Vasilis Gkatzelis, Daniel Halpern, and Nisarg Shah. Resolving the optimal metric distortion conjecture. In *FOCS*, pages 1427–1438, 2020.

[Gonzalez, 1985] Teofilo Francisco Gonzalez. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.*, 38:293–306, 1985.

[Guruswami and Indyk, 2003] Venkatesan Guruswami and Piotr Indyk. Embeddings and non-approximability of geometric problems. In *SODA*, pages 537–538, 2003.

[Hochbaum and Shmoys, 1986] Dorit S. Hochbaum and David Bernard Shmoys. A unified approach to approximation algorithms for bottleneck problems. *J. ACM*, 33(3):533–550, 1986.

[Hsu and Nemhauser, 1979] Wen-Lian Hsu and George Lann Nemhauser. Easy and hard bottleneck location problems. *Discret. Appl. Math.*, 1(3):209–215, 1979.

[Izsak *et al.*, 2018] Rani Izsak, Nimrod Talmon, and Gerhard Woeginger. Committee selection with intraclass and interclass synergies. In *AAAI*, pages 1071–1078, 2018.

[Jesorsky *et al.*, 2001] Oliver Jesorsky, Klaus J. Kirchberg, and Robert W. Frischholz. Robust face detection using the Hausdorff distance. In *AVBPA*, pages 90–95, 2001.

[Lackner and Skowron, 2020] Martin Lackner and Piotr Skowron. Multi-winner voting with approval preferences. https://arxiv.org/pdf/2007.01795.pdf, 2020. Accessed: 2022-05-15.

[Liu and Guo, 2016] Hong Liu and Jiong Guo. Parameterized complexity of winner determination in minimax committee elections. In *AAMAS*, pages 341–349, 2016.

[Merrill, 1993] Samuel Merrill III. Voting behavior under the directional spatial model of electoral competition. *Public Choice*, 77:739–756, 1993.

[Peters, 2018] Dominik Peters. Single-peakedness and total unimodularity: New polynomial-time algorithms for multi-winner elections. In *AAAI*, pages 1169–1176, 2018.

[Pierczynski and Skowron, 2019] Grzegorz Pierczyński and Piotr Skowron. Approval-based elections and distortion of voting rules. In *IJCAI*, pages 543–549, 2019.

[Thiele, 1895] Thorvald Nicolai Thiele. Om flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger*, pages 415–441, 1895.

[West, 2000] Douglas Brent West. *Introduction to Graph Theory*. Prentice-Hall, 2000.

[Yang and Wang, 2018] Yongjie Yang and Jianxin Wang. Multiwinner voting with restricted admissible sets: Complexity and strategyproofness. In *IJCAI*, pages 576–582, 2018.

[Yang, 2019] Yongjie Yang. Complexity of manipulating and controlling approval-based multiwinner voting. In *IJCAI*, pages 637–643, 2019.