Parameterized Dichotomy of Choosing Committees Based on Approval Votes in the Presence of Outliers

Palash Dey
Tata Institute of Fundamental Research, Mumbai
palash.dey@tifr.res.in

Neeldhara Misra
Indian Institute of Technology Gandhinagar
mail@neeldhara.com

Y. Narahari
Indian Institute of Science Bangalore
hari@csa.iisc.ernet.in

ABSTRACT

Approval ballots provide an opportunity for agents to make a comment about every candidate, without incurring the overhead of determining a full ranking on the set of candidates; they are very natural for many practical settings. We study the computational complexity of the committee selection problem for several approval-based voting rules in the presence of outliers. Our first result shows that outliers render the committee selection problem intractable for approval, net approval, and minisum approval voting rules. We next study the parameterized complexity of this problem with five natural parameters, namely the target score, the size of the committee (and its dual parameter namely the number of candidates outside the committee); and the number of outliers (and its dual parameter namely the number of non-outliers). For approval, net approval, and minisum approval voting rules, we provide a dichotomous result, which resolves the parameterized complexity of this problem for all subsets of the above five natural parameters considered (by showing either FPT or W[1]-hardness for all subsets of parameters).

CCS Concepts

•Theory of computation → Analysis of Algorithms and Problem Complexity; •Distributed Artificial Intelligence → Multi-agent Systems;

Keywords

voting; committee selection; outliers; social choice; parameterized complexity

1. INTRODUCTION

Aggregating preferences of a set of agents is a fundamental problem in artificial intelligence and social choice theory [11]. Typically, agents (or voters) express their preferences over alternatives (or candidates). There are many different models for expressing preferences, ranging from the simple plurality voting (each voter provides his or her favorite choice) to the most comprehensive method where each voter provides a complete ranking over the set of all candidates. *Approval ballots* represent an intermediate model, where a voter approves of or disapprove each candidate – thus a vote may be captured in the form of a subset of approved candidates or as a binary string indexed by the candidate set. Approval votes provide a well-studied model in their own right (see, for instance, [20, 9, 23, 25, 5]). Approval ballots provide the agents an opportunity to make a comment about every candidate, without incurring the overhead of determining a full ranking on the candidate set.

Just as there are multiple ways of expressing or modelling preferences, there exist several notions of aggregating preferences as well. Often, the goal is to just find a winning alternative. In an extreme scenario we may wish to derive a complete ranking of the alternatives that suitably reflects the preferences of the agents. Another possibility is the desire to find the "top k" alternatives, or to find an unordered subset of k alternatives. Our focus in this paper will be on finding the "best" subset of k candidates, and we will refer to a subset of k candidates as a *committee*.

A natural way of aggregating approval votes is approval voting – every candidate gets one point for every approval, and the candidates can be ranked according to score. In a single-winner scenario, this method of aggregation works rather well; in fact, it coincides with Condorcet winners (for an appropriate interpretation of the notion in the dichotomous setting, see [25, Chapter 11]) and moreover it is robust to strategic behavior of the voters [20, 9, 3, 25]. However, when it comes to choosing a committee, issues begin to emerge, especially to do with fairness, susceptibility to manipulation, and proportional representation of agents. Several other voting rules have been proposed to address some of these concerns.

We now informally describe a general way of thinking about a certain class of voting rules, in the context of approval ballots. Recall that a committee is a subset of candidates, and note that a vote can also be interpreted as a subset of candidates. Therefore, one might consider various natural notions of distance between these two sets. A certain fixed notion of distance leads to a measure of suitability of a committee with respect to an election – for instance, by considering the sum of distances to all voters, or the maximum distance incurred from any voter. For a committee $\mathcal{X}$ and a vote $\mathcal{S}$, the notions of distance that are well-studied in the literature include the size of the symmetric difference (leading to the *minisum* or *minimax* rules) [10, 24, 14, 21]; the size of the intersection (leading to the notion of *approval score*); or the difference between $|\mathcal{X} \cap \mathcal{S}|$ and $|\mathcal{X} \setminus \mathcal{S}|$ (leading to the notion of *net approval* scores) [28, 30, 2].

Appears in: *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

Motivation

The standard approach to the selection of a winning committee is to look for one that optimizes these scoring functions over the entire election – that is, when the sum or maximum of scores is taken over all the votes. However, in many scenarios where voting has been applied successfully, for example, social networks [6, 29], computational biology [22], spam detection [15], restaurant rating, etc., it is common to have some outliers or errors in the input set of votes. In such situations, it is plausible that there is a committee that represents a satisfactory consensus when the scores are taken over all the voters barring a few rather than the entire set of voters. For example, consider the minisum voting rule, which minimizes the sum of the Hamming distances of the committee from the votes in the electorate (the sum is taken over the votes). For a committee of size k, the best possible minisum score is zero. Clearly, the lower the minisum score, the greater the overall satisfaction of the electorate. Now, imagine in an electorate where most voters agree on a committee $\mathcal{C}^*$ of size k, and a very small number of voters do not approve anyone from this committee. Ignoring the voters who do not approve anyone from $\mathcal{C}^*$ as outliers leads us to a situation where there exists a committee with the best possible minisum score. The discussion certainly provokes the following lines of thought: how much can the best minisum committee change if a few voters are removed? More specifically, is there a subset of at least n^* votes that admits a committee whose minisum score is at most k^*? The problem of finding the best minisum committee is a special case of this problem (when $n = n^*$). This generalization allows us to explore trade-offs and structure in the election: if there is a committee that prescribes a satisfactory consensus over a large fraction of the election, then it is likely to be perceived a more suitable choice by the community compared to a committee that optimizes the same score function over the entire election. We acknowledge that it may be unfair to "leave out" some of the votes in many real life applications of voting, proportional representation and political election for example. However, in many other real life applications of voting as mentioned above, the presence of outliers are quite common and we believe that considering outliers in such applications is important. Indeed, the notion of outliers is very popular in the closest string problem [8] whose main motivation comes from biology and in social network scenario.

Related Work

The notion of finding a satisfactory consensus over a large subset has been explored in multiple contexts. For instance, the well-studied the Young voting rule [31] can be interpreted as finding the minimum number of "outliers" whose removal makes a certain candidate a Condorcet winner. A closely related notion is *control*, where one is interested in the possibility of influencing the outcome of an election by either restricting the set of candidates or voters, or introducing "spoiler" candidates or voters. This is a well-studied phenomenon, see, for instance, [4] for an overview of the computational complexity of various types of control problems. A specific type of control is *control by deleting voters*, which arises in the context of single-winner elections: we wish to know if it is possible to make a particular candidate c win by removing at most k votes from the instance. For a fixed candidate c, this problem has the flavor of identifying outliers, with respect to making c a winner.

Outliers are also quite common in the literature of *Closest String* problems. The setting of closest string involves a collection of n strings, and the goal is to find a single string that minimizes either the maximum distance, or the sum of distances, from all the input strings. The most commonly studied notion of distance is the Hamming distance. Another question that is often asked is the following: given a budget k, is there a string that is "close to" at least k strings? This question has been studied for both the minimax and minisum notions of closeness [7, 26].

While the problem of finding outliers in multiwinner elections shares similarities with the aforementioned problems, nevertheless, there are also significant differences. For instance, the problem of control by deleting voters involves a favorite candidate and the goal is to manipulate the outcome. To begin with, the control problems studied so far are mostly in the context of single-winner elections. More importantly, the question we are raising is in a different spirit. Here, we don't identify favorites – our goal is to find out if the election structurally admits a satisfactory consensus barring a small number of votes.

Further, once we interpret votes as binary strings, we are asking a question that is rather similar to the Closest String family of problems. Here, the main distinction is that our search is only over strings that have a fixed number of ones. This similarity has been noted and explored in some previous works on voting (see, for instance, Byrka and Sornat [13] and Boucher and Ma [8]). More specifically, and again in the spirit of observing similarities with the closest string family of problems, the minimax approval rule has been studied from the perspective of outliers [27]. To the best of our knowledge, our work is the first comprehensive study of computational complexity of approval ballots in the presence of outliers.

Our Framework

An advantage of using scoring rules for approval ballots described above is that the winning committees are polynomially computable for most of them (with the notable exception of the minimax voting rule). However, once we pose the question of whether a target score t is achievable after the removal of at most k outliers, the complexity landscape changes quite dramatically. We show that this question is a computationally hard problem – in particular, we establish NP-hardness. Having shown hardness in the classical setting, we explore the complexity issue further, primarily from the perspective of fixed-parameter algorithms, using the framework of parameterized complexity. We refer the reader to the book [19] for a comprehensive overview of parameterized complexity.

Our Contributions

Our main result is a set of parameterized *dichotomy* theorems for the problem of identifying a small subset of outliers. We focus on three specific scoring-oriented voting rules for aggregating approval ballots. These are Minisum, Approval, and Net Approval (we refer the reader to the next section for a detailed description of these voting rules). Next, our task is to identify reasonable parameters. The fundamental structural parameters for a voting problem are the number of voters or candidates. These parameters provide natural starting points, and have been popular among several works that address the parameterized complexity of computational problems in voting.

Also, for any optimization problem, a natural choice of parameter is the quantity that is being optimized, often called the *standard parameter*. In our setting, the standard parameter would be the target score. In the context of outliers for multiwinner elections, the other natural parameters would be the size of the committee that is sought, and the number of outliers. We explore the parameterized complexity of finding outliers from the lens of all possible combinations of these parameters, and obtain *classification* results for the three voting rules under investigation. We defer a precise statement of these results to the next section. We believe resolving parameterized complexity for *all subsets of the parameters considered* for the Minisum, Approval, and Net Approval voting rules are novel as well as non-trivial. We remark that proving such parameterized dichotomy results for important problems has recently gained a lot of interest in theoretical computer science; see, for example, the work of Bringmann et al. [12], Aravind et al. [1], etc. and our work is the first of such kind in computational social choice.

Organization

The rest of this paper is organized as follows. In Section 2, we describe our main contributions, summarizing it in Section 2.1, and along the way we setup the notation and the definitions that will be used in subsequent sections. As we will see, our results are parameterized dichotomy theorems that classify the parameterized complexity of the problem of finding outliers. Subsequently, in Section 3, we begin with presenting our algorithm results in Section 3.1; then move on to our parameterized hardness results in Section 3.2 and finally combine them to prove the main theorems in Section 3.3. We conclude in Section 4. In the interest of space, we skip few proofs which can be found in the full version [17].

2. DEFINITIONS AND OVERVIEW OF KEY RESULTS

We consider three standard approval scoring rules, namely the *minisum, approval, and net approval scores*. The last two scores, as originally defined in the literature, are designed to simply give us the total amount of approval that a committee incurs from all the voters. The scores themselves are therefore non-decreasing functions of n, and as such, the question of outliers is not interesting if we use the scores directly (in particular, it is impossible to improve these scores by removing votes). Therefore, we consider the dual scoring system that complements the original – namely, we score a committee based on the amount of *disapproval* that it incurs from all the votes, and seek to minimize the *total disapproval*. Typically, for any notion of approval, there are either one or two natural complementary notions of disapproval that present themselves (discussed in greater detail below). This formulation is consistent with the idea that we want our scores to capture "distance" rather than closeness. We note that in terms of scores, these rules are equivalent to the original, but choosing to ask the minimization question allows us to formulate the problem of finding the best committee in the presence of outliers.

Remark 1. *We chose to use the notion of disapproval instead of approval because of its consistency with the other distance-based rules (like minisum and minimax). All the variations are equivalent as scoring functions, and we note that our choice is only a matter of exposition.*

<table>
<tr><th colspan="2">Measures of approval</th><th colspan="2">Measures of disapproval</th></tr>
<tr><td>Minisum</td><td colspan="2">$|\mathcal{X}\Delta\mathcal{S}|$</td><td>Minisum</td></tr>
<tr><td rowspan="2">Approval</td><td rowspan="2">$|\mathcal{X}\cap\mathcal{S}|$</td><td>$|\mathcal{X}\setminus\mathcal{S}|$</td><td rowspan="2">Disapproval</td></tr>
<tr><td>$|\mathcal{S}\setminus\mathcal{X}|$</td></tr>
<tr><td>Net Approval</td><td>$|\mathcal{X}\cap\mathcal{S}|-|\mathcal{X}\setminus\mathcal{S}|$</td><td>$|\mathcal{X}\setminus\mathcal{S}|-|\mathcal{X}\cap\mathcal{S}|$</td><td>Net Disapproval</td></tr>
</table>

Table 1: The approval-based rules that are considered in this paper. Here, $\mathcal{X}$ is a committee and $\mathcal{S}$ is a vote, and the table illustrates the possible scores that $\mathcal{X}$ can incur from $\mathcal{S}$. The total score of $\mathcal{X}$ will be the sum of these scores taken over $\mathcal{S}$. Δ denotes the symmetric difference of sets.

Number of votes	n	Set of candidates	$\mathcal{C}$
Number of candidates	m	Committee chosen	$\mathcal{C}^*$
Number of outliers	$\overline{n}$	Non-committee	$\overline{\mathcal{C}}(=\mathcal{C}\setminus\mathcal{C}^*)$
Number of non-outliers	$n^*(=n-\overline{n})$	Set of votes	$\mathcal{V}$
Size of committee	m^*	Set of non-outliers	$\mathcal{V}^*$
Size of non-committee	$\overline{m}(=m-m^*)$	Set of outliers	$\overline{\mathcal{V}}(=\mathcal{V}\setminus\mathcal{V}^*)$
Score of the committee	t	Hamming distance	$h(\cdot)$

Table 2: Notation.

In Table 1, we summarize the notions of distances between a committee and a vote. We also refer the reader to Table 2 for an overview of the notation we use in this paper. Each of these notions naturally gives rise to a score-based voting rule. Formally, for any distance function $s : 2^{\mathcal{C}} \times 2^{\mathcal{C}} \to \mathbb{N}$ between two subsets of candidates, we overload notation and define the corresponding score function $s : 2^{\mathcal{C}} \times 2^{\mathcal{V}} \to \mathbb{N}$ as follows (the view of $s(\cdot)$ will be clear from the context):

$$s(\mathcal{X},\mathcal{W}) := \sum_{\mathcal{S}\in\mathcal{W}} s(\mathcal{X},\mathcal{S}).$$

For the *winner determination* problem, the goal is to find a committee $\mathcal{X}$ of size m^* that minimizes $s(\mathcal{X},\mathcal{V})$. For all the scoring rules in Table 1, a winning committee of size m^* can be found in polynomial time for any $m^* \leqslant m$. We are now ready to define the problem of winner determination for a scoring rule s in the presence of outliers:

> s-Outliers
> **Input:** A set of votes $\mathcal{V} = \{\mathcal{S}_1,\ldots,\mathcal{S}_n\}$ over a set of candidates $\mathcal{C} = \{c_1,\ldots,c_m\}$, a committee size m^*, a requirement n^*, and a target score t.
> **Question:** Does there exist a committee $\mathcal{C}^* \subset \mathcal{C}$ and a set of non-outliers $\mathcal{V}^* \subset \mathcal{V}$ such that $|\mathcal{V}^*| \geqslant n^*$, $|\mathcal{C}^*| = m^*$, and $s(\mathcal{C}^*,\mathcal{V}^*) \leqslant t$?

Remark 2. *We will focus on only one variant of disapproval for the approval scoring rule, namely the one given by the top row in Table 1. The other variation is symmetric and our results hold in the exact same fashion.*

At this point, we briefly introduce terminology that will be pertinent to the description of our contributions. A *parameterized problem instance* comprises of an instance x in the usual sense, and a parameter k. A problem with parameter k is called *fixed parameter tractable* (FPT) if it is solvable in time $f(k) \cdot p(|x|)$, where f is an arbitrary function of k and p is a polynomial in the input size $|x|$. There is also a hierarchy of complexity classes above FPT, and showing that a parameterized problem is hard for one of these classes is considered evidence that the problem is unlikely to be fixed-parameter tractable. Indeed, assuming the Exponential Time Hypothesis, a problem hard for W[1] does not belong to FPT [18]. We refer the reader to [16, 19] for a detailed introduction to parameterized algorithms and complexity.

2.1 Summary of Results

We show that the s-Outliers problem is NP-complete for all the scoring rules considered here, even in the special cases when every vote approves exactly two candidates or every candidate is approved by exactly two votes (these results follow from Lemma 3 and Lemma 6, respectively). To initiate the parameterized study of the s-Outliers problem, we have to identify suitable parameters. Recall that for an optimization problem, a natural choice of parameter is the quantity that we are optimizing, or the *standard parameter*. In our setting, the standard parameter would be the target score t. However, as we will see, the problem turns out to be $W[1]$-hard when parameterized by the target score. Therefore, we consider other natural parameters in the hope that they may lead to tractability.

We now turn our attention to the next most natural parameters, that is, the number of voters or candidates. In contrast to the standard parameter, the s-Outliers problem is FPT for all the scoring rules that we consider, with respect to either the number of candidates or voters. We briefly describe the overall approach here. Note that we can, in "FPT time", enumerate all choices of outliers or all possible committees, for the parameters n and m, respectively. For any fixed choice of outliers (respectively, committee), it turns out that a natural greedy strategy leads us to the optimal solution for any of the scoring rules studied here. The greedy algorithm is easy to implement in polynomial time, and we describe this argument formally in Proposition 1.

Given this initial landscape, we approach the problem with a finer set of parameters, which are motivated by the definition of the problem: the size of the committee (m^*), the number of candidates *not* in the committee ($\overline{m}$), the number of non-outliers (n^*), the number of outliers ($\overline{n}$). Along with the target score, we therefore have five parameters to consider. Note that the parameters m and n implicitly accounted for as combinations of these parameters, since $m = m^* + \overline{m}$ and $n = n^* + \overline{n}$.

Our main results are the following three dichotomy theorems, that completely classify the parameterized complexity of the problem for the parameters considered above. In particular, for any subset of these parameters, we establish if the s-Outliers problem is FPT or W[1]-hard when $s \in$ {Minisum, Net Disapproval, Disapproval}. In all cases, parameterizing by m or n makes the problem tractable, but beyond that, the classifications diverge and are slightly different for each of the three problems. The precise classifications are given by the following theorems.

Theorem 1. *Let $\mathcal{P} = \{m^*, \overline{m}, n^*, \overline{n}, t\}$, and $\mathcal{Q} \subseteq \mathcal{P}$. The* Minisum-outliers *problem parameterized by $\mathcal{Q}$ is* FPT *if $\mathcal{Q}$ contains either $\{m^*, \overline{m}\}$, $\{n^*, \overline{n}\}$, or $\{\overline{n}, t\}$, and is* W[1]*-hard otherwise.*

Theorem 2. *Let $\mathcal{P} = \{m^*, \overline{m}, n^*, \overline{n}, t\}$, and $\mathcal{Q} \subseteq \mathcal{P}$. The* Net-Disapproval-outliers *problem parameterized by $\mathcal{Q}$ is* FPT *if $\mathcal{Q}$ contains either $\{m^*, \overline{m}\}$ or $\{n^*, \overline{n}\}$, and is* W[1]*-hard otherwise.*

Theorem 3. *Let $\mathcal{P} = \{m^*, \overline{m}, n^*, \overline{n}, t\}$, and $\mathcal{Q} \subseteq \mathcal{P}$. The* Disapproval-Outliers *problem parameterized by $\mathcal{Q}$ is* FPT *if $\mathcal{Q}$ contains either $\{m^*, \overline{m}\}$, $\{n^*, \overline{n}\}$, or $\{\overline{m}, \overline{n}, t\}$ and is* W[1]*-hard otherwise.*

3. RESULTS

In this section, we present our parameterized dichotomy results. We first describe our FPT algorithms for s-Outliers when parameterized by either m, n, or $\overline{n} + t$. In the next subsection, we describe our W[1]-hardness results for various carefully chosen combinations of parameters. In the last subsection, we tie these results together into a proof of Theorems 1 to 3. The reason that a small number of classifications account for all possible cases is the following. Note that if a problem is FPT (respectively, W[1]-hard) with respect to some subset $\mathcal{Q} \subseteq \mathcal{P}$ of parameters, then it is also FPT (W[1]-hard) with respect to any superset (subset) of $\mathcal{Q}$. Therefore, typically it suffices to demonstrate algorithms and hardness results on carefully constructed subsets of parameters, that will in turn account for all possibilities.

3.1 FPT Algorithms

We begin by discussing the FPT algorithms. While describing running times, we will use the $O^*()$ notation to suppress polynomial factors. Our first result relies on the fact that the s-Outliers problem is polynomial time solvable for all the voting rules considered here if we know either the committee or the non-outliers of the solution. Hence by guessing either the committee or the the set of non-outliers as applicable, we immediately obtain the following result.

Proposition 1. *The* Minisum-Outliers *problem admits algorithms running in time $O^*(2^n)$ and $O^*(2^m)$.*

Now we show that the Minisum-outliers problem, parameterized by $(t + \overline{n})$, is in FPT.

Lemma 1. *There is a $O^*(2^{t+\overline{n}})$ time algorithm for the* Minisum-outliers *problem.*

Proof. Notice that if $t \geqslant n - \overline{n}$ (that is $n \leqslant t + \overline{n}$), then we can simply use the algorithm in Proposition 1, which runs in $O^*(2^n) = O^*(2^{t+\overline{n}})$ time. Otherwise, we have $t < n - \overline{n}$. In this case, consider any committee $\mathcal{X}$ that achieves the target score t with respect to some set of at most $\overline{n}$ outliers. If $t < n - \overline{n}$, then there exists a vote, say $\mathcal{S}_i$, such that $s(\mathcal{X}, \mathcal{S}_i) = 0$. Indeed, otherwise every vote contributes at least one to the Minisum score and the total score of $\mathcal{X}$ will be at least $n - \overline{n}$, contradicting the assumption that $\mathcal{X}$ had a Minisum score of at most t.

Since $s(\mathcal{X}, \mathcal{S}_i) = 0$, observe that the vote $\mathcal{S}_i$ determines the committee. We can proceed by guessing $\mathcal{S}_i$, declaring the candidates approved by $\mathcal{S}_i$ as the committee, and then finding the corresponding outliers (recall that the last step is easy when the committee is fixed). Therefore, the algorithm iterates over all votes $\mathcal{S}$ that approve exactly m^* candidates, and fixes $\mathcal{X}$ to be $\mathcal{S}$. It then identifies the non-outliers $\mathcal{V}^*$ with respect to the committee $\mathcal{X}$, and returns YES if there is at least one choice of $\mathcal{S}$ for which the number of outliers is at most $\overline{n}$, and NO otherwise. □

We begin with the details of FPT algorithm for DISAPPROVAL-OUTLIERS when parameterized by $t, \overline{m}, \overline{n}$. To begin with, we have the following straight forward observations that naturally lead to preprocessing rules,:

OBSERVATION 1. *Let $(\mathcal{V} = \{\mathcal{S}_i : i \in [n]\}, \mathcal{C}, n^*, m^*, t)$ be a* YES *instance of* DISAPPROVAL-OUTLIERS.

- *If $\mathcal{S}_j = \mathcal{C}$ for some $j \in [n]$, then there exists a set $\mathcal{V}^* \subseteq \mathcal{V}$ of non-outliers of size n^* containing $\mathcal{S}_j$ and a committee $\mathcal{C}^* \subseteq \mathcal{C}$ of size m^* such that the disapproval score of $\mathcal{C}^*$ with votes $\mathcal{V}^*$ is at most t.*
- *If $|\mathcal{S}_j| > \bar{m} + t$ for some $j \in [n]$, then there does not exist a set $\mathcal{V}^* \subseteq \mathcal{V}$ of non-outliers of size n^* containing $\mathcal{S}_j$ and a committee $\mathcal{C}^* \subseteq \mathcal{C}$ of size m^* such that the disapproval score of $\mathcal{C}^*$ with votes $\mathcal{V}^*$ is at most t.*

Next, we need to define an annotated variant of the problem.

DEFINITION 1. FORCED-DISAPPROVAL-OUTLIERS
Given a set of votes $\mathcal{V} = \{\mathcal{S}_i : i \in [n]\}$, a set of candidates $\mathcal{C}$, a subset $\mathcal{C}' \subseteq \mathcal{C}$ of candidates that should be in the committee, a subset $\mathcal{V}' \subseteq \mathcal{V}$ of votes that should be in the set of non-outliers the number n^* of non-outliers, the size m^* of the committee, and target score t, do there exists a set of non-outliers $\mathcal{V}^* \subseteq \mathcal{V}$ and a committee $\mathcal{C}^* \subseteq \mathcal{C}$ such that $|\mathcal{V}^*| \geqslant n^*, |\mathcal{C}^*| = m^*, \mathcal{C}' \subseteq \mathcal{C}^*, \mathcal{V}' \subseteq \mathcal{V}*$ and the disapproval score of $\mathcal{C}^*$ with votes $\mathcal{V}^*$ is at most t.

It is easy to establish the following.

OBSERVATION 2. DISAPPROVAL-OUTLIERS *many-to-one reduces to* FORCED-DISAPPROVAL-OUTLIERS.

We are now ready to describe the branching algorithm.

LEMMA 2. *There is a $O^*((t + \bar{m} + \bar{n})^{t+\bar{m}+\bar{n}})$ time algorithm for the* FORCED-DISAPPROVAL-OUTLIERS *problem.*

PROOF. We delete all votes $\mathcal{S} \in \mathcal{V}$ for which $\mathcal{S} = \mathcal{C}$. The correctness of this step follows from Observation 1. We pick any arbitrary non-outlier vote $\mathcal{T} \in \mathcal{V}$ and branch according to the following guesses:

1. If $\bar{n} > 0$, then we guess that $\mathcal{T}$ is an outlier - in this case, we decrement $\bar{n}$ by one and remove $\mathcal{T}$ from the instance.
2. If $\mathcal{T}$ is not an outlier, then let $\bar{\mathcal{T}} = \mathcal{C} \setminus \mathcal{T}$ be the set of candidates which are not approved in the vote $\mathcal{T}$. We branch on every possible intersection $\mathcal{D}$ of $\bar{\mathcal{T}}$ with the "final" set of candidates $\bar{\mathcal{C}} \subseteq \mathcal{C}$ not in the committee. In the branch corresponding to $\mathcal{D}$, we remove all the candidates in $\mathcal{D}$ from the instance, mark all the candidates $\bar{\mathcal{T}} \setminus \mathcal{D}$ to be in the committee if they are not already been marked, reduce the target score t by $|\bar{\mathcal{T}} \setminus \mathcal{D}|$, and mark $\mathcal{T}$ to be a non-outlier.

At every node of the branching, we again delete all votes $\mathcal{S} \in \mathcal{V}$ for which $\mathcal{S} = \mathcal{C}$ and branch again. We have the following base cases:

- If $t < 0$, then we return NO.
- If either $\bar{n} = O(1)$ or $\bar{m} = O(1)$, then we solve the problem in $O(m^{O(1)})$ time by either trying all possible outliers or all possible candidates not in the committee.

Notice that, in every branch, the value of the parameter $(t + \bar{m} + \bar{n})$ decreases by at least one. Also the branching factor is at most $(t + \bar{m} + \bar{n})$ and at every node we perform $O(poly(m, n))$ time operations. Hence, the running time of the algorithm is $O^*((t + \bar{m} + \bar{n})^{t+\bar{m}+\bar{n}})$. □

3.2 W[1]-Hardness Results

In this section, we establish our W[1]-hardness results. To begin with, we focus on parameters combined with the target score t. Notice that we have tractability when we combine either $(\overline{n}+n^*)$ or $(\overline{m}+m^*)$ along with t (follows from Proposition 1), and also for t and $\overline{n}$, from Lemma 1. Therefore, the interesting combinations remaining are $(t+n^*+\overline{m})$ and $(t+n^*+m^*)$. We first consider the parameter $(t+n^*+m^*)$. To show hardness, we will need the d-CLIQUE problem, which is known to be W[1]-hard parameterized by clique size [19]. The problem is defined below.

d-CLIQUE	**Parameter:** k
Input: A d-regular graph $\mathcal{G} = (U, E)$ with vertex set U and edge set E and a positive integer k.	
Question: Is there a clique of size k in $\mathcal{G}$?	

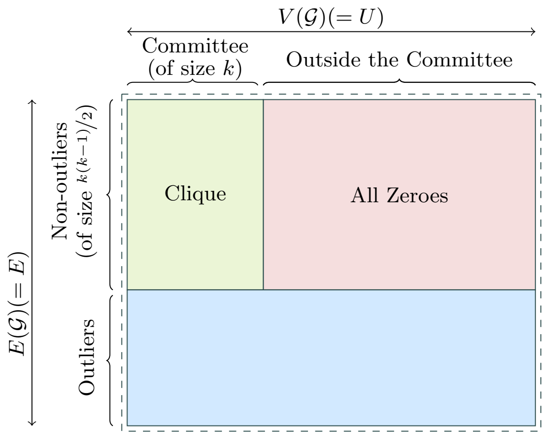
All the results in this section are based on reductions from the d-CLIQUE problem. In general, our reductions will typically have candidates and voters based on the vertices and edges, and the approval ballots will be designed to encode the structure of the graph. The instance will be engineered such that the vertices and edges involved in the clique will naturally correspond to outliers and committee members, for a carefully chosen value of the target score.

Our first result is the W[1]-hardness of the MINISUM-OUTLIERS problem when parameterized by $(t + m^* + n^*)$, even when every vote approves exactly two candidates. This also establishes NP-hardness of the natural decision version of the problem in the classical setting (as the proof involves a polynomial time reduction from a NP-hard problem).

LEMMA 3. *Let $s \in$ {Minisum, Net-Disapproval, Disapproval}. Then the s-*OUTLIERS *problem is* W[1]*-hard, when parameterized by $(t+m^*+n^*)$, even when every vote approves exactly two candidates.*

PROOF. First let us prove the result for the MINISUM-OUTLIERS problem. We exhibit a parameterized reduction from the CLIQUE problem to the MINISUM-OUTLIERS problem thereby proving the result. Let $(\mathcal{G} = (U, E), k)$ be an arbitrary instance of the d-CLIQUE problem. Let $U = \{u_1, \cdots, u_n\}$ and $E = \{e_1, \cdots, e_m\}$. For constructing the corresponding instance of the MINISUM-OUTLIERS problem, we introduce one candidate for each vertex, and one vote for each edge (see Figure 1). The i^{th} vote approves the candidates corresponding to the endpoints of the edge e_i. We ask for a k-sized committee, set a target score of $t = (k-2)^{k(k-1)/2}$, and allow for at most $m - {k(k-1)}/2$ outliers. Formally, the instance $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$ of the MINISUM-OUTLIERS is the following.

$$\mathcal{V} = \{\mathcal{S}_1, \cdots, \mathcal{S}_m\}, \mathcal{C} = \{c_1, \cdots, c_n\}, n^* = {k(k-1)}/2, m^* = k,$$



Incidence Matrix of $\mathcal{G}$

Figure 1: A schematic depicting the reduction in Lemma 3.

$$t = (k-2)^{k(k-1)}/2, \mathcal{S}_i = \{c_j : u_j \in e_i\}, \forall i \in [m]$$

We briefly describe the intuition for the equivalence of these two instances. Given a clique, suppose we choose a committee based on the vertices of the clique, and the non-outliers based on the edges of the clique. Then the sum of the distances incurred by this committee is $(k-2)^{k(k-1)}/2$, since every non-outlier approves of two members from the k-sized committee, and disapproves of everyone else. Notice also that since every voter approves exactly two candidates, this is the best that we can hope for if we are allowed at most $m - {}^{k(k-1)}/2$ outliers.

We now turn to a formal proof of the equivalence. In the forward direction, suppose $W \subset U$ forms a clique with $|W| = k$ in $\mathcal{G}$, and let Q denote the set of edges that have both endpoints in W. Consider the committee $\mathcal{C}^* = \{c_i : u_i \in W\}$ and the set of non-outliers $\mathcal{V}^* = \{\mathcal{S}_i : e_i \in Q\}$. This achieves the minisum score of $(k-2)^{k(k-1)}/2$.

In the reverse direction, suppose there exists a set of non-outliers $\mathcal{V}^* \subset \mathcal{V}$ with $|\mathcal{V}^*| \geqslant {}^{k(k-1)}/2$ and a committee $\mathcal{C}^* \subset \mathcal{C}$ with $|\mathcal{C}^*| = k$ such that $s(\mathcal{C}^*, \mathcal{V}^*) \leqslant (k-2)^{k(k-1)}/2$. We claim that the vertices corresponding to $\mathcal{C}^*$ form a clique of size k. To see this, we consider the edges given by $Q = \{e_j : \mathcal{S}_j \in \mathcal{V}^*\}$. We show that each edge e_j has both its endpoints in $W = \{u_i : c_i \in \mathcal{C}^*\}$. Indeed otherwise, suppose that there is an edge $e_j \in Q$ for which at least one of its endpoints are not in W. Then $s(\mathcal{S}_j, \mathcal{C}^*) > (k-2)$. On the other hand we have $s(\mathcal{S}, \mathcal{C}^*) \geqslant (k-2)$ for every $\mathcal{S} \in \mathcal{V}^*$. Hence we have $s(\mathcal{C}^*, \mathcal{V}^*) > (k-2)^{k(k-1)}/2$, a contradiction.

The proofs for the other rules are identical, except for the values of the target score. We define $t = (k-2)^{k(k-1)}/2$ for the disapproval and $t = (k-4)^{k(k-1)}/2$ for the net disapproval voting rules. It is easily checked that the details are analogous. □

We next consider the parameter $(t+n^*+\overline{m})$ and exhibit a parameterized reduction from the d-Clique problem. This reduction is very similar to the one before, except that we need to bound the number of non-committee members $(\overline{m})$ as a function of k, as opposed to the committee. The theme of the previous reduction was to give every voter (edge) two approvals that was forced to correspond to the chosen candidates, by the choice of the target score. Now we flip the structure of the approval ballots, and encode the edges with two *disapprovals* instead. In particular, the i^{th} vote approves all candidates *except* the ones corresponding to the endpoints of the edge e_i. We ask for an $(n-k)$-sized committee. As before, the target score is again such that all the disapprovals among non-outliers must correspond to members not chosen, which in turn will correspond to the clique. We now turn to the formal proof.

Lemma 4. *The* Minisum-outliers *problem is* W[1]*-hard, when parameterized by* $(t + n^* + \overline{m})$*. Also, for* $s \in$ *{disapproval, net disapproval}, the* s-Outliers *problem is* W[1]*-hard when parameterized by* $(n^*+\overline{m})$*, even when* $t = 0$.

Proof. First let us prove the result for the Minisum-outliers problem. We exhibit a parameterized reduction from the d-Clique problem to the Minisum-outliers problem thereby proving the result. Let $(\mathcal{G} = (U, E), k)$ be an arbitrary instance of the d-Clique problem. Let $U = \{u_1, \cdots, u_n\}$ and $E = \{e_1, \cdots, e_m\}$. For constructing the corresponding instance of the Minisum-outliers problem, we introduce one candidate for each vertex, and one vote for each edge. The i^{th} vote approves all candidates *except* the ones corresponding to the endpoints of the edge e_i. We ask for an $(n-k)$-sized committee. The other details are similar to our previous reduction, in that we have a target score of $t = (k-2)^{k(k-1)}/2$, and allow for at most $m - {}^{k(k-1)}/2$ outliers.

We define the corresponding instance $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$ of the Minisum-outliers problem as follows:

$$\mathcal{V} = \{\mathcal{S}_1, \cdots, \mathcal{S}_m\}, \mathcal{C} = \{c_1, \cdots, c_n\},$$

$$\mathcal{S}_i = \{c_j : u_j \notin e_i\} \; \forall i \in [n],$$

$$n^* = {}^{k(k-1)}/2, \overline{m} = k, t = (k-2)^{k(k-1)}/2$$

We claim that the two instances are equivalent. In the forward direction, suppose $W \subset U$ forms a clique with $|W| = k$, and let Q denote the set of edges that have both endpoints in W. Then the committee $\mathcal{C}^* = \{c_i : u_i \notin W\}$ along with the set of non-outliers $\mathcal{V}^* = \{\mathcal{S}_i : u_i \in Q\}$ achieves minisum score of $(k-2)^{k(k-1)}/2$.

In the reverse direction, suppose there exists a set of non-outliers $\mathcal{V}^* \subset \mathcal{V}$ and a committee $\mathcal{C}^* \subset \mathcal{C}$ such that $s(\mathcal{V}^*, \mathcal{C}^*) \leqslant (k-2)^{k(k-1)}/2$. We claim that the vertices in $W = \{u_i : c_i \in \overline{\mathcal{C}}\}$ form a clique with edge set $Q = \{e_i : \mathcal{S}_i \in \mathcal{V}^*\}$. If not then, there exist a vote $\mathcal{S}_j \in \mathcal{V}^*$ such that $s(\mathcal{S}_j, \mathcal{C}) \geqslant k$. On the other hand we have $s(\mathcal{S}_i, \mathcal{C}) \geqslant k-2$ for every $\mathcal{S}_i \in \mathcal{V}^*$. This implies that $s(\mathcal{V}^*, \mathcal{C}^*) > (k-2)^{k(k-1)}/2$, which is a contradiction. We define $t = 0$ for the disapproval and net disapproval voting rules. It is easily checked that the details are analogous. □

Now we turn to parameter combinations that do not involve t. Notice that the only interesting combinations here are $\overline{n}+m^*$ and $\overline{n}+\overline{m}$, since combinations involving n^* and either m^* or $\overline{m}$ are accounted for in the W[1]-hardness results of Lemmas 3 and 4, and the remaining two combinations are FPT due to Proposition 1.

In the next two reductions, the contrast from the proofs of Lemmas 3 and 4 is that we will have candidates corresponding to edges, and votes corresponding to vertices. Another significant divergence is the fact that the target scores we demand in the next two reductions are no longer pure functions of k (which we do not need also since the parameters considered do not involve the target score t). To begin with, we establish W[1]-hardness with respect to the parameter $(\overline{n} + m^*)$. Note that for this, we will need both the committee size and the number of outliers to be bounded by a function of k only. We lead up to a scenario where the outliers correspond to the clique, and the committee corresponds to the edges incident to the clique. To make this work, we will have a slightly counter-intuitive construction of the approval ballots – namely that the i^{th} vote (corresponding to the vertex u_i) approves of every candidate that corresponds to an edge *not* incident to u_i. We now turn to the formal proof.

LEMMA 5. *The* MINISUM-OUTLIERS *problem is* W[1]*-hard when parameterized by* $(\overline{n} + m^*)$*. Also, the* DISAPPROVAL-OUTLIERS *problem is* W[1]*-hard parameterized by* $(m^* + \overline{n})$*, even when the target score* $t = 0$*. In particular,* s-OUTLIERS *is* W[1]*-hard parameterized by* $(m^* + \overline{n} + t)$*.*

PROOF. To begin with, let us prove the result for the MINISUM-OUTLIERS problem. We exhibit a parameterized reduction from the d-CLIQUE problem to the MINISUM-OUTLIERS problem thereby proving the result. Let $(\mathcal{G} = (U, E), k)$ be an instance of the d-CLIQUE problem. Let $U = \{u_1, \cdots, u_n\}$ and $E = \{e_1, \cdots, e_m\}$ be respectively the set of vertices and the set of edges of $\mathcal{G}$. For constructing the corresponding instance of the MINISUM-OUTLIERS problem, we introduce one candidate for each edge, and one vote for each vertex. The i^{th} vote approves of every candidate that corresponds to an edge *not* incident to the vertex u_i. We ask for a ${}^{k(k-1)}/_2$-sized committee, set a target score of $t = (m - {}^{k(k-1)}/_2 - d)\,(n - k)$, and allow for at most k outliers. Formally, the instance $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$ of the MINISUM-OUTLIERS is the following.

$$\mathcal{V} = \{\mathcal{S}_1, \cdots, \mathcal{S}_n\}, \mathcal{C} = \{c_1, \cdots, c_m\},$$

$$\mathcal{S}_i = \{c_j : u_i \notin e_j\} \ \forall i \in [n],$$

$$\overline{n} = k, m^* = {}^{k(k-1)}/_2, t = (m - {}^{k(k-1)}/_2 - d)\,(n - k)$$

We briefly describe the intuition for the equivalence of these two instances. Given a clique in $\mathcal{G}$, suppose we choose a committee based on the edges of the clique, and the outliers based on the vertices of the clique. Then every non-outlier approves of every candidate from the clique (recall that the approvals are "flipped"). Also, among the members not chosen in the committee, there are exactly d disapprovals (since $\mathcal{G}$ is d-regular), causing a total distance of $(m - {}^{k(k-1)}/_2 - d)$ per non-outlier. Notice also that since every voter approves exactly $m - d$ candidates, this is the best distance that we can hope from a committee of size ${}^{k(k-1)}/_2$. If we are allowed at most k outliers, it is clear that our best bet overall is to remove vertices corresponding to a clique of size k.

We claim that the two instances are equivalent. In the forward direction, suppose $W \subset U$ forms a clique with $|W| = k$, and let Q denote the set of edges that have both endpoints in W. Then the committee $\mathcal{C}^* = \{c_i : e_i \in Q\}$ and the set of non-outliers $\mathcal{V}^* = \{\mathcal{S}_i : u_i \notin W\}$ achieves minisum score of $(m - {}^{k(k-1)}/_2 - d)\,(n - k)$.

In the reverse direction, suppose there exists a set of outliers $\overline{\mathcal{V}} \subset \mathcal{V}$ and a committee $\mathcal{C}^* \subset \mathcal{C}$ such that $|\overline{\mathcal{V}}| \leqslant k, |\mathcal{C}^*| = {}^{k(k-1)}/_2$, and $s(\mathcal{V}^*, \mathcal{C}^*) \leqslant (m - {}^{k(k-1)}/_2 - d)\,(n - k)$. We claim that the vertices in $W = \{u_i : \mathcal{S}_i \in \overline{\mathcal{V}}\}$ form a clique with the corresponding set of edges $Q = \{e_j \in E : c_j \in \mathcal{C}^*\}$. If not, then there exists a candidate $x \in \mathcal{C}^*$ that is not approved by at least one vote in $\mathcal{V}^*$. However, this implies that $s(\mathcal{V}^*, \mathcal{C}^*) > (m - {}^{k(k-1)}/_2 - d)\,(n - k)$, which is a contradiction.

We define $t = 0$ for the disapproval voting rule. It is easily checked that the details are analogous. $\square$

Finally, we show that the MINISUM-OUTLIERS problem is W[1]-hard when parameterized by $(\overline{n} + \overline{m})$. This is again achieved by "flipping" the definition of the approval ballot from the previous proof, with the same reasoning that allowed the transition from Lemma 3 to Lemma 4.

LEMMA 6. *Let* $s \in$ *{Minisum, Disapproval, Net-Disapproval}. Then* s-OUTLIERS *is* W[1]*-hard, when parameterized by* $(\overline{n} + \overline{m})$*, even when every candidate is approved in exactly two votes. Further, the* NET-DISAPPROVAL-OUTLIERS *problem, parameterized by* $(\overline{n} + \overline{m})$*, is* W[1]*-hard even when* $t = 0$ *and every candidate is approved in exactly two votes. In particular, the* NET-DISAPPROVAL-OUTLIERS *problem is* W[1]*-hard parameterized by* $(\overline{n} + \overline{m} + t)$*.*

PROOF. First let us prove the result for the MINISUM-OUTLIERS problem. We exhibit a parameterized reduction from the d-CLIQUE problem to the MINISUM-OUTLIERS problem thereby proving the result. Let $(\mathcal{G} = (U, E), k)$ be an instance of the d-CLIQUE problem. Let $U = \{u_1, \cdots, u_n\}$ and $E = \{e_1, \cdots, e_m\}$. We define the corresponding instance $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$ of the MINISUM-OUTLIERS problem as follows:

$$\mathcal{V} = \{\mathcal{S}_1, \cdots, \mathcal{S}_n\}, \mathcal{C} = \{c_1, \cdots, c_m\},$$

$$\mathcal{S}_i = \{c_j : u_i \in e_j\}, \forall i \in [n],$$

$$\overline{n} = k, \overline{m} = {}^{k(k-1)}/_2, t = (n - k)\,(m - {}^{k(k-1)}/_2 - d)$$

We claim that the two instances are equivalent. In the forward direction, suppose $W \subset U$ forms a clique with $|W| = k$, and let Q denotes the set of edges that have both endpoints in W. Consider the committee $\mathcal{C}^* = \{c_i : e_i \notin Q\}$ and the set of outliers $\overline{\mathcal{V}} = \{\mathcal{S}_i : u_i \in W\}$. Consider now a vote $\mathcal{S}_j \in \mathcal{V}^*$. Note that $|\mathcal{C}^* \cap \mathcal{S}_j| = d$, since each edge incident to u_j is an edge that does not belong to Q, by the definition of $\mathcal{C}^*$. Further, this also implies that $|\mathcal{C}^* \setminus \mathcal{S}_j| = (m - {}^{k(k-1)}/_2 - d)$. Therefore, $s(\mathcal{S}_j, \mathcal{C}^*) = (m - {}^{k(k-1)}/_2 - d)$. Hence we have $s(\mathcal{V}^*, \mathcal{C}^*) = (n - k)\,(m - {}^{k(k-1)}/_2 - d)$.

For the reverse direction, suppose there exists a set of non-outliers $\mathcal{V}^* \subset \mathcal{V}$ and a committee $\mathcal{C}^* \subset \mathcal{C}$ such that $s(\mathcal{V}^*, \mathcal{C}^*) \leqslant (n - k)\,(m - {}^{k(k-1)}/_2 - d)$. We claim that the vertices in $W = \{u_i : \mathcal{S}_i \in \overline{\mathcal{V}}\}$ form a clique. If not then there exists a vote $\mathcal{S} \in \mathcal{V}^*$, such that $s(\mathcal{S}, \mathcal{C}^*) > (m - {}^{k(k-1)}/_2 - d)$. However, for every vote $\mathcal{S}' \in \mathcal{V}^*$, we have $s(\mathcal{S}', \mathcal{C}^*) \geqslant (m - {}^{k(k-1)}/_2 - d)$. This makes $s(\mathcal{V}^*, \mathcal{C}^*) > (n - k)\,(m - {}^{k(k-1)}/_2 - d)$ which is a contradiction.

The proofs for the other rules are identical, except for the values of the target score. We define $t = (n - k)\,(m - {}^{k(k-1)}/_2 - d)$ for the disapproval voting rule. For the net disapproval voting rule, we add $(m - 2d)$ many dummy candidates who are approved by every vote. We keep $\overline{m} = {}^{k(k-1)}/_2$ and make $t = 0$. It is easily checked that the remaining details are analogous. $\square$

We show next that NET-DISAPPROVAL-OUTLIERS is W[1]-hard with respect to the combined parameter $(\overline{n} + t + m^*)$.

LEMMA 7. *The* NET-DISAPPROVAL-OUTLIERS *problem is* W[1]*-hard, when parameterized by* $(\overline{n} + m^*)$*, even when the target score is* 0*. In particular,* NET-DISAPPROVAL-OUTLIERS *problem is* W[1]*-hard, when parameterized by* $(t + \overline{n} + m^*)$*.*

PROOF. We reduce the CLIQUE problem to the NET-DISAPPROVAL-OUTLIER problem. Let $(\mathcal{G} = (U, E), k)$ be an arbitrary instance of the CLIQUE problem. Let $U = \{u_1, \cdots, u_n\}$. We define the corresponding instance $(\mathcal{V}, \mathcal{C}, n^*, m^*, t)$ of the NET-DISAPPROVAL-OUTLIER problem as follows:

$$\mathcal{C} = \{c_e : e \in E\} \cup D, \text{ where } |D| = {}^{k(k-1)}/_{2},$$

$$\mathcal{V} = \{\mathcal{S}_1, \cdots, \mathcal{S}_n\} \cup \{\mathcal{T}_1, \cdots, \mathcal{T}_{n+2k}\},$$

$$\mathcal{S}_i = \{c_e : u_i \notin e\}, \forall i \in [n], \ \mathcal{T}_j = D, \forall j \in [{}^{k(k-1)}/_{2}],$$

$$\overline{n} = k, \ m^* = 2^{k(k-1)}/_{2}, \ t = 0$$

We claim that the two instances equivalent. In the forward direction, suppose $\mathcal{G}$ has a clique on a subset of vertices $W \subset U$ of size k; let Q be the clique edges. Then we define the set of outlier votes to be $\overline{\mathcal{V}} = \{\mathcal{S}_i : u_i \in W\}$ and the committee to be the set of candidates $\mathcal{C}^* = D \cup \{c_e : e \in Q\}$. Now we have $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S} \cap \mathcal{C}^*| = (n-k){}^{k(k-1)}/_{2} + (n+2k)|D| = (n+k){}^{k(k-1)}/_{2}$ and $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S}^c \cap \mathcal{C}^*| = (n-k)|D| + (n+2k){}^{k(k-1)}/_{2} = (n+k){}^{k(k-1)}/_{2}$ thereby achieving the net approval score t of 0.

In the reverse direction, suppose we have a set of outlier votes $\overline{\mathcal{V}}$ and a committee $\mathcal{C}^*$ which achieves a net approval score t of 0. First observe that we can assume without loss of generality that D is a subset of $\mathcal{C}^*$ since irrespective of the outliers chosen, every candidate in D receives at least $(n+k)$ approvals and every candidate not in D receives at most n approvals. Now since the committee $\mathcal{C}^*$ contains D, for every $j \in [n+2k]$, the vote $\mathcal{T}_j$ contributes at most $|D| - {}^{k(k-1)}/_{2} = 0$ to the net approval score, whereas, for every $i \in [n]$, the vote $\mathcal{S}_i$ contributes at least ${}^{k(k-1)}/_{2} - |D| = 0$ to the net approval score. Hence, we may assume without loss of generality that $\mathcal{T}_j$ belongs to the set of non-outliers $\mathcal{V}^*$ for every $j \in [n+2k]$. Now we claim that the set of edges $Q = \{e : c_e \in \mathcal{C}^*\}$ must form a clique on the set of vertices $W = \{u : \mathcal{S}_u \in \overline{\mathcal{V}}\}$. If not then, $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S} \cap \mathcal{C}^*| < (n+k){}^{k(k-1)}/_{2}$ and $\sum_{\mathcal{S} \in \mathcal{V}^*} |\mathcal{S}^c \cap \mathcal{C}^*| > (n+k){}^{k(k-1)}/_{2}$ thereby making the net approval score t strictly more than 0. □

3.3 Proof of the Main Theorems

Proof of Theorem 1. We are now ready to describe the proof of Theorem 1. Since $m^* + \overline{m} = m$ and $n^* + \overline{n} = n$, the tractability results follow from Proposition 1 and Lemma 1. Now, we have the following cases.

1. $\mathcal{Q}$ excludes at least one of m^* and $\overline{m}$, and
2. $\mathcal{Q}$ excludes at least one of n^* and $\overline{n}$, and
3. $\mathcal{Q}$ excludes at least one of t and $\overline{n}$.

Among such choices of $\mathcal{Q}$, we have the following cases.

1. Suppose $t \in \mathcal{Q}$. Then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{t, n^*, \overline{m}\}$ or a subset of $\mathcal{Q}_2 = \{t, n^*, m^*\}$. The hardness for all of these cases follow from Lemma 4 and Lemma 3, respectively.
2. Suppose $\overline{n} \in \mathcal{Q}$. Then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{\overline{n}, \overline{m}\}$ or a subset of $\mathcal{Q}_2 = \{\overline{n}, m^*\}$. The hardness for all of these cases follow from Lemma 5 and Lemma 6, respectively.
3. If neither t nor $\overline{n}$ belongs to $\mathcal{Q}$, then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{n^*, m^*\}$, or $\mathcal{Q}_2 = \{n^*, \overline{m}\}$. Note that these cases are already subsumed by Case (1) above.

This completes the proof of the theorem. □

Proof of Theorem 2. We now turn to the proof of Theorem 2. Since $m^* + \overline{m} = m$ and $m^* + \overline{n} = n$, the tractability results follow from Proposition 1. Now, we only have to consider subsets of parameters $\mathcal{Q}$ such that $\mathcal{Q}$ does not contain both m^* and $\overline{m}$, and $\mathcal{Q}$ does not contain both n^* and $\overline{n}$. Among such choices of $\mathcal{Q}$, we have the following cases.

1. Suppose $n^* \in \mathcal{Q}$. Then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{n^*, \overline{m}, t\}$ or a subset of $\mathcal{Q}_2 = \{n^*, m^*, t\}$. The hardness for all of these cases follow from Lemma 4 and Lemma 3, respectively.
2. Suppose $\overline{n} \in \mathcal{Q}$. Then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{\overline{n}, \overline{m}, t\}$ or a subset of $\mathcal{Q}_2 = \{\overline{n}, m^*, t\}$. The hardness for all these cases follow from Lemma 6 and Lemma 7, respectively.
3. If neither n^* nor $\overline{n}$ belongs to $\mathcal{Q}$, then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{t, m^*\}$, or $\mathcal{Q}_2 = \{t, \overline{m}\}$. Note that these cases are already subsumed by the cases above.

This completes the proof of the theorem. □

Proof of Theorem 3. Finally, we turn to the case of the disapproval voting rules. Since $m^* + \overline{m} = m$ and $m^* + \overline{n} = n$, the tractability results follow from Proposition 1. Now, we only have to consider subsets of parameters $\mathcal{Q}$ such that $\mathcal{Q}$ does not contain both m^* and $\overline{m}$, and $\mathcal{Q}$ does not contain both n^* and $\overline{n}$. Among such choices of $\mathcal{Q}$, we have the following cases.

1. Suppose $n^* \in \mathcal{Q}$. Then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{n^*, \overline{m}, t\}$ or a subset of $\mathcal{Q}_2 = \{n^*, m^*, t\}$. The hardness for all of these cases follows from Lemma 4 and Lemma 3, respectively.
2. Suppose $\overline{n} \in \mathcal{Q}$. Then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{\overline{n}, m^*, t\}$ or $\mathcal{Q}_2 = \{\overline{n}, \overline{m}, t\}$. The hardness for all subsets of $\mathcal{Q}_1$ follows from Lemma 5. The status for $\mathcal{Q}_2$ is FPT from Theorem 2. The hardness for all strict subsets of $\mathcal{Q}_2$ follows from Lemma 6 and the cases that are already resolved.
3. If neither n^* nor $\overline{n}$ belongs to $\mathcal{Q}$, then $\mathcal{Q}$ is either a subset of $\mathcal{Q}_1 = \{t, m^*\}$, or $\mathcal{Q}_2 = \{t, \overline{m}\}$. Note that these cases are already subsumed by the cases above.

This completes the proof of the theorem. □

4. CONCLUDING REMARKS

We have showed that consideration of outliers makes the problem of choosing a good committee substantially harder, even in the framework of parameterized complexity. However, there are few fixed parameter tractable algorithms which we feel are quite practically appealing. Refining the FPT fragment of our dichotomy to kernelization is an exciting direction for further investigation. We are also interested in extending our dichotomous results to a dichotomy based on more parameters, such as the maximum number of candidates in a vote. Another important future work is to perform similar study for other committee selection rules like minimax, Chamberlin-Courant's, and Monroe's rules.

REFERENCES

[1] N. Aravind, R. Sandeep, and N. Sivadasan. Parameterized lower bounds and dichotomy results for the np-completeness of h-free edge modification problems. In *Latin American Symposium on Theoretical Informatics*, pages 82–95. Springer, 2016.

[2] H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh. Justified representation in approval-based committee voting. In *Proc. Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 784–790, 2015.

[3] H. Aziz, S. Gaspers, J. Gudmundsson, S. Mackenzie, N. Mattei, and T. Walsh. Computational aspects of multi-winner approval voting. In *Proc. International Conference on Autonomous Agents and Multiagent Systems*, pages 107–115. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[4] J. J. Bartholdi, C. A. Tovey, and M. A. Trick. How hard is it to control an election? *Math Comput Model.*, 16(8):27–40, 1992.

[5] D. Baumeister, G. Erdélyi, E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Computational aspects of approval voting. In *Handbook on Approval Voting*, pages 199–251. Springer, 2010.

[6] P. Boldi, F. Bonchi, C. Castillo, and S. Vigna. Voting in social networks. In *Proc. 18th ACM Conference on Information and Knowledge Mmanagement*, pages 777–786. ACM, 2009.

[7] C. Boucher, G. M. Landau, A. Levy, D. Pritchard, and O. Weimann. On approximating string selection problems with outliers. *Theor. Comput. Sci.*, 498:107–114, 2013.

[8] C. Boucher and B. Ma. Closest string with outliers. *BMC Bioinformatics*, 12(S-1):S55, 2011.

[9] S. J. Brams. Mathematics and democracy: Designing better voting and fair-division procedures. *Math Comput Model.*, 48(9-10):1666–1670, 2008.

[10] S. J. Brams, D. M. Kilgour, and M. R. Sanver. A minimax procedure for electing committees. *Public Choice*, 132(3-4):401–420, 2007.

[11] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia. Handbook of computational social choice, 2016.

[12] K. Bringmann, D. Hermelin, M. Mnich, and E. J. van Leeuwen. Parameterized complexity dichotomy for steiner multicut. *J. Comput. Syst. Sci*, 2016.

[13] J. Byrka and K. Sornat. PTAS for minimax approval voting. In *Web and Internet Economics*, volume 8877 of *Lecture Notes in Computer Science*, pages 203–217. Springer, 2014.

[14] I. Caragiannis, D. Kalaitzis, and E. Markakis. Approximation algorithms and mechanism design for minimax approval voting. In *Proc. Twenty-Fourth AAAI Conference on Artificial Intelligence, Atlanta, Georgia, USA, July 11-15, 2010*, 2010.

[15] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *J. Artif. Int. Res.*, 10(1):243–270, May 1999.

[16] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh. *Parameterized Algorithms*. Springer, 2015.

[17] P. Dey, N. Misra, and Y. Narahari. On choosing committees based on approval votes in the presence of outliers. *CoRR*, abs/1511.04190, 2015.

[18] R. G. Downey and M. R. Fellows. *Parameterized complexity*. Springer-Verlag, New York, 1999.

[19] R. G. Downey and M. R. Fellows. *Fundamentals of Parameterized Complexity*. Springer Publishing Company, Incorporated, 2013.

[20] P. C. Fishburn. Axioms for approval voting: Direct proof. *J. Econ. Theory*, 19(1):180–185, 1978.

[21] J. Gramm, R. Niedermeier, and P. Rossmanith. Fixed-parameter algorithms for closest string and related problems. *Algorithmica*, 37(1):25–42, 2003.

[22] B. G. Jackson, P. S. Schnable, and S. Aluru. Consensus genetic maps as median orders from inconsistent sources. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 5(2):161–171, 2008.

[23] D. M. Kilgour. Approval balloting for multi-winner elections. In *Handbook on approval voting*, pages 105–124. Springer, 2010.

[24] R. LeGrand, E. Markakis, and A. Mehta. Some results on approximating the minimax solution in approval voting. In *Proc. International Conference on Autonomous Agents and Multiagent Systems*, volume 7, pages 1185–1187, 2007.

[25] D. Lepelley. Jean-françois laslier and M. remzi sanver (eds.): Handbook on approval voting. *Soc. Choice Welf.*, 40(4):1231–1235, 2013.

[26] C. Lo, B. Kakaradov, D. Lokshtanov, and C. Boucher. SeeSite: characterizing relationships between splice junctions and splicing enhancers. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 11(4):648–656, July 2014.

[27] N. Misra, A. Nabeel, and H. Singh. On the parameterized complexity of minimax approval voting. In *Proc. 2015 International Conference on Autonomous Agents and Multiagent Systems, Istanbul, Turkey, May 4-8, 2015*, pages 97–105, 2015.

[28] A. D. Procaccia, J. S. Rosenschein, and A. Zohar. On the complexity of achieving proportional representation. *Soc. Choice Welf.*, 30(3):353–362, 2008.

[29] M. A. Rodriguez, D. J. Steinbock, J. H. Watkins, C. Gershenson, J. Bollen, V. Grey, and B. Degraf. Smartocracy: Social networks for collective decision making. In *In 40th Annual Hawaii International Conference on Systems Science (HICSS'07). Waikoloa*, pages 90–99. IEEE, 2007.

[30] P. K. Skowron and P. Faliszewski. Fully proportional representation with approval ballots: Approximating the maxcover problem with bounded frequencies in FPT time. In *Proc. Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 2124–2130, 2015.

[31] H. P. Young and A. Levenglick. A consistent extension of Condorcet's election principle. *SIAM J. Appl. Math*, 35(2):285–300, 1978.