Privacy in Elections: k-Anonymizing Preference Orders*

Nimrod Talmon**

Institut für Softwaretechnik und Theoretische Informatik, TU Berlin, Germany
nimrodtalmon77@gmail.com

Abstract. We study the (parameterized) complexity of a combinatorial problem, motivated by the desire to publish elections-related data, while preserving the privacy of the voters (humans or agents). In this problem, introduced and defined here, we are given an election, a voting rule, and a distance function over elections. The task is to find an election which is not too far away from the original election (with respect to the given distance function) while preserving the election winner (with respect to the given voting rule), and such that the resulting election is k-anonymized; an election is said to be k-anonymous if for each voter in it there are at least $k-1$ other voters with the same preference order. We consider the problem of k-anonymizing elections for the Plurality rule and for the Condorcet rule, for the Discrete distance and for the Swap distance. We show that the parameterized complexity landscape of our problem is diverse, with cases ranging from being polynomial-time solvable to Para-NP-hard.

1 Introduction

We consider privacy issues when publishing preferences-related (or, election-related) data. Assume being given data consisting of a set of records, where each record (corresponding to a human or an agent) contains preferences-related information as well as some private (side) information. The task is to publish this data (for example, to let researchers analyze it) while preserving the privacy of the entities in it. Two of the most well-studied approaches for achieving privacy when publishing information is *differential privacy* (see, for example, Dwork and Roth [12]) and k-*anonymity* (and l-*diversity*; see, for example, Sweeney [25] and Machanavajjhala et al. [20]). Here we follow the k-anonymity framework (see Clifton and Tassa [9] for a recent comparison between these two approaches).

We say that an election is k-*anonymous* if each preference order in it appears at least k times (where a preference order is an order over a set of predefined alternatives). Given an input election , the goal is to generate an election which is k-anonymous but still preserves some properties of the original election.

* To appear in the 20th International Symposium on Fundamentals of Computation Theory (FCT 2015).

** Supported by DFG Research Training Group MDS (GRK 1408).

It is natural to consider the distance between the original election and the resulting anonymized election (in a similar way as done when k-anonymizing graphs; there, one can define the distance as the symmetric difference of the edge set, for example). Therefore, we consider distances over elections. We study the Discrete distance (where each preference order can be transformed into any other preference order at unit cost) and the Swap distance (where each two consecutive alternatives can be swapped at unit cost), as these are the most basic and well-studied distances defined on elections (see, for example, [14]). The idea is that if the distance is small, then the anonymized election does not differ too much from the original election; this is, arguably, more apparent in the Swap distance.

Besides requiring that the original election and the resulting k-anonymized election will be close (with respect to the considered distance), we would like to preserve some specific properties of the original election (in this, we follow ideas presented by Bredereck et al. [5], who considered preserving graph properties, such as the connectivity, the relative distances, and the diameter). Here, we require that the winner of the election will be preserved. For this, we need to fix a voting rule. We study two voting rules, the Plurality rule and the Condorcet rule, as these are the most basic and well-studied voting rules, which are also good representative rules (specifically, the Plurality rule, albeit simple, can be seen as a representative scoring rule, while the Condorcet rule can be seen as a representative tournament-based rule).

In what follows, we study the parameterized complexity of k-anonymizing elections, under the Plurality rule and under the Condorcet rule, for the Discrete distance and for the Swap distance. We consider two election-related parameters, specifically, the number of voters and the number of alternatives, and one anonymity-related parameter, the anonymity level k. We show that the parameterized complexity landscape of our problem is diverse, with cases ranging from being polynomial-time solvable to Para-NP-hard.

In a way, this paper can be seen as bringing the well-studied field of k-anonymity to the well-studied field of voting systems and social choice, with the hope of better understanding complexity issues of preserving privacy when publishing election-related data. We view our definition of k-anonymous elections as being a natural adaptation of the concept of k-anonymity to preferences-related (or, election-related) data.

1.1 Related Work

There is a big body of literature on security of elections and on preserving privacy of voters participating in (digital) elections. Chaum [7], Nurmi et al. [24], and Cuvelier et al. [10], among others, considered cryptographic mechanisms to encrypt the votes, while Chen et al. [8], among others, considered differential privacy. Ashur and Dunkelman [1] showed how to breach the privacy of voters for the Israeli parliament when an adversary can look at the publicly-available nation-wide election statistics. This work is of some relevance to us as it considers privacy with respect to (publicly-available) published data.

Sweeney [25] introduced the concept of k-anonymity as a way to preserve privacy over published data, after demonstrating how to identify many individuals by mixing publicly-available medical data with publicly-available voter lists (interestingly, Sweeney [25] already somewhat focuses on election-related data, specifically on the party affiliation; informally speaking, party affiliation corresponds to the Plurality rule, since only the first choice counts, while for the Condorcet rule we would need the complete preference orders publicly available). Much work has been done on k-anonymizing tables (for example, Meyerson and Williams [22], Bredereck et al. [4], and Bredereck et al. [6]; some of these concentrate on parameterized complexity), on k-anonymizing graphs (for example, Liu and Terzi [19], Hartung et al. [17], and Bredereck et al. [5]; some of these concentrate on parameterized complexity). Here, we consider neither general tables nor graphs, but instead we consider elections. Indeed, elections can be described as tables, but here we require to preserve the winner and allow different, election specific, operations. Specifically, while the Discrete distance can be natural also for general tables, this is not the case for the Swap distance.

For general information about social choice, elections, voting systems, and voting rules, we point the reader to any textbook on social choice, for example the book by Brandt et al. [3].

2 Preliminaries

Considering distances over elections, we follow some notation from Elkind et al. [14]. We assume familiarity with standard notions regarding algorithms, computational complexity, and graph theory. For a non-negative integer z, we denote the set $\{1, \ldots, z\}$ by $[z]$.

2.1 Elections and Distances

An election E is a pair (C, V) where $C = \{c_1, \ldots, c_m\}$ is the set of alternatives and $V = (v_1, \ldots, v_n)$ is the collection of voters. Each voter v_i is represented by a total order $\succ_{v_i}$ over C (her preference order; we use voters and preference orders interchangeably). A voter which prefers c to all other alternatives is called a *c-voter*. For reading clarity, we refer to the voters as females, while the alternatives are males. A voting rule $\mathcal{R}$ is a function that, given an election $E = (C, V)$, returns a set $\mathcal{R}(E) \subseteq C$ of election winners (indeed, we use the non-unique winner model). We consider the following voting rules:

1. **the Plurality rule.** Each alternative receives one point for each voter which ranks him first, and the winners are the highest-scoring alternatives.
2. **the Condorcet rule.** An alternative c is a (weak) Condorcet winner if for each other alternative $c' \in C \setminus \{c\}$, it holds that $|\{v \in V : c \succ_v c'\} \geq |\{v \in V : c' \succ_v c\}|$, that is, if he beats (or ties with) all other alternatives in head-to-head contests. The Condorcet rule elects the Condorcet winners if some exist, returning an empty set otherwise.

Given a set V' of preference orders, we say that a function $d : V' \times V' \to \mathbb{N}$ is a *distance function* over preference orders if it is a metric over preference orders. Given a distance function over preference orders d and two elections (over the same set of alternatives), $E = (C, (v_1, \ldots, v_n))$ and $E' = (C, (v'_1, \ldots, v'_n))$, the above definition can be naturally extended by fixing an arbitrary order for the voters of E, that is, $[v_1, \ldots, v_n]$, considering all the possible permutations for the voters of E', that is, $[v'_{\pi(1)}, \ldots, v'_{\pi(n)}]$, and defining the d-distance between E and E' to be $d(E, E') = \min_{\pi \in S_n} \sum_{i \in [n]} d(v_i, v'_{\pi(i)})$, where S_n is the set containing all the possible permutations of $[n]$. We define the following distance functions:

1. **Discrete distance.** $d_{\text{discr}}(v_1, v_2) = 0$ if and only if $v_1 = v_2$, while otherwise $d_{\text{discr}}(v_1, v_2) = 1$. Indeed, for two elections, $E = (C, (v_1, \ldots, v_n))$ and $E' = (C, (v'_1, \ldots, v'_n))$, it holds that $d_{\text{discr}}(E, E') = |\{i : v_i \neq v'_i\}|$ (hence, the Hamming distance).
2. **Swap distance.** $d_{\text{swap}}(v_1, v_2) = |\{(c, c') \in C \times C : c \succ_{v_1} c' \ \wedge \ c' \succ_{v_2} c\}|$. Indeed, the swap distance $d_{\text{swap}}(v_1, v_2)$ (also called the Dodgson distance) is the minimum number of swaps of consecutive alternatives needed for transforming the preference order of v_1 to that of v_2.

Clearly, both the Discrete distance and the Swap distance are distance functions.

2.2 Anonymization

A group of voters with the same preference order is called a *block*. Using this notion, we have that an election is k-anonymous if and only if each block in it is of size at least k. We denote the number of voters in block B by $|B|$ and say that a block is *bad* if $0 < B < k$ (as it is not yet anonymized in this case). Since all voters in a block have the same preference order, it is valid to consider the preference order of the voters in the block. Specifically, a block of c-voters is called a *c-block*.

2.3 Parameterized Complexity

An instance (I, k) of a parameterized problem consists of the "classical" problem instance I and an integer k being the *parameter* [11, 15, 23]. A parameterized problem is called *fixed-parameter tractable* (FPT) if there is an algorithm solving it in $f(k) \cdot |I|^{O(1)}$ time, for an arbitrary computable function f only depending on the parameter k. In difference to that, algorithms running in $|I|^{f(k)}$ time prove membership in the class XP (clearly, FPT $\subseteq$ XP).

One can show that a parameterized problem L is (presumably) not fixed-parameter tractable by devising a *parameterized reduction* from a W[1]-hard problem (for example, the Clique problem, parameterized by the solution size) or a W[2]-hard problem (for example, the Set Cover problem, parameterized by the solution size) to L. A parameterized problem which is NP-hard even for instances for which the parameter is a constant is said to be Para-NP-hard.

	Discrete distance	Swap distance
Plurality rule	P (Theorem 1)	NP-h even when $n = k = 4$ (Theorem 3) FPT wrt. m (Theorem 5)
Condorcet rule	NP-h (Theorem 2) para. complexity wrt. k is open FPT wrt. n (Theorem 4) FPT wrt. m (Theorem 5)	NP-h even when $n = k = 4$ (Theorem 3) FPT wrt. m (Theorem 5)

Table 1: (parameterized) Complexity of EA.

2.4 Main Problem and Overview of Our Results

The main problem we consider in this paper is defined as follows.

> $\mathcal{R}$-d-Election Anonymization ($\mathcal{R}$-d-EA)
> **Input:** An election $E = (C, V)$ where $C = \{c_1, \dots, c_m\}$ is the set of alternatives and $V = (v_1, \dots, v_n)$ is the collection of voters, anonymity level k, and a budget s.
> **Question:** Is there a k-anonymous election E' such that $\mathcal{R}(E) = \mathcal{R}(E')$ and $d(E, E') \leq s$ (where $\mathcal{R}$ is a voting rule, d is a distance function over elections, and an election is said to be k-anonymous if for each voter in it there are at least $k - 1$ other voters with the same preference order)?

We study the (parameterized) complexity of $\mathcal{R}$-d-Election Anonymization, where we consider both the Plurality rule and the Condorcet rule as the voting rule $\mathcal{R}$, and where we consider both the Discrete distance and the Swap distance as the distance d (that is, we consider the following four variants: Plurality-Discrete-EA, Condorcet-Discrete-EA, Plurality-Swap-EA, and Condorcet-Swap-EA). Our results are summarized in Table 1. Due to the lack of space, some proof details are deferred to the full version.

3 Results

Intuitively, from all variants considered in this paper, Plurality-Discrete-EA should be the most tractable, as the Plurality rule is conceptually simpler than the Condorcet rule and the Discrete distance is conceptually simpler than the Swap distance. This intuition is correct: it turns out that Plurality-Discrete-EA is polynomial-time solvable, while all other variants are NP-hard. We begin by describing a polynomial-time algorithm, based on dynamic programming, for Plurality-Discrete-EA.

Theorem 1. *Plurality-Discrete-*EA *is polynomial-time solvable.*

Proof. We describe an algorithm based on applying dynamic programming twice, in a nested way. To understand the general idea, consider an alternative c and its corresponding c-voters. We have two cases to consider with respect to the solution election: either (1) some of the c-voters are transformed to be c'-voters (for some, possibly several, other alternatives $c' \neq c$), or (2) some c'-voters (for some, possibly several, other alternatives $c' \neq c$) are transformed to be c-voters. The crucial observation is that, with respect to anonymizing the c-voters, we do not need to remember the specific c'-voters discussed above, but only their number. Therefore, we define a first (outer) dynamic program, iterating over the alternatives, and computing the most efficient way of anonymizing the c-voters, while considering all possible values for these numbers of c'-voters, and while making sure that the initial winner of the election stays the winner. In order to compute how to anonymize the c-voters we define a second (inner) dynamic program, considering the c-blocks one at a time. For each c-block, the inner dynamic program decides whether to make the respective c-block empty (with zero voters) or full (with at least k voters), by considering the possible ways of transforming other c-voters or other c'-voters, similarly in spirit to the first (outer) dynamic programming. The full proof is deferred to the full version. $\square$

For the Condorcet rule, still considering the Discrete distance, we can show that EA is NP-hard, by a reduction from a restricted variant of the Exact Cover by 3-Sets problem.

Theorem 2. *Condorcet-Discrete-*EA *is* NP*-hard.*

Proof. We reduce from the following NP-hard problem [16], defined as follows.

> Restricted Exact Cover by 3-Sets
> **Input:** Collection $\mathcal{S} = \{S_1, \ldots, S_n\}$ of sets of size 3 over a universe $X = \{x_1, \ldots, x_n\}$ such that each element appears in exactly three sets.
> **Question:** Is there a subset $\mathcal{S}' \subseteq \mathcal{S}$ such that each element x_i occurs in exactly one member of $\mathcal{S}'$?

We assume, without loss of generality, that $\{\{x_1, x_2, x_3\}, \ldots, \{x_{n-2}, x_{n-1}, x_n\}\} = \{\{x_{3l-2}, x_{3l-1}, x_{3l}\} : l \in [n/3]\} \nsubseteq \mathcal{S}$, and that $n = 0 \pmod 3$. Given an instance for *Restricted Exact Cover by* 3*-Sets*, we create an instance for Condorcet-Discrete-EA, as follows.

We create two alternatives, p and d, and, also, for each element $x_i \in X$, we create an alternative x_i, such that the set of alternatives is $\{X, p, d\}$. We create a set of $n/3$ voters, called *jokers*, such that the ith joker (for $i \in [n/3]$) has preference order $x_{3i-2} \succ x_{3i-1} \succ x_{3i} \succ p \succ d \succ X \setminus \{x_{3i-2}, x_{3i-1}, x_{3i}\}$. For each set S_j, we create k voters, each with preference order $S_j \succ p \succ d \succ \overline{S_j}$. We refer to these voters as the *set voters*. We create another set of $(n-6)k + ((n/3) - 2)$ voters, each with preference order $d \succ X \succ p$, called the *init voters*. Finally, we set s to $n/3$ and k to $(n/3) + 2$. This finishes the construction.

Let us first compute the winner in the input election. Notice that the set voters and the jokers prefer p to d, while the init voters prefer d to p. As there are kn set voters, $n/3$ jokers, but only $(n-6)k+((n/3)-2)$ init voters, p defeats d. Consider an element x_i. There is exactly one joker which prefers x_i to p, and exactly three set voters which prefer x_i to p (as x_i appears in exactly three sets), and all the init voters prefer x_i to p. Other than these, all other voters prefer p to x_i. Therefore, there are $(n/3) - 1 + k(n-3)$ voters which prefer p to x_i and $1 + 3k + (n-6)k + ((n/3) - 2) = (n/3) - 1 + k(n-3)$ voters which prefer x_i to p, so p and x_i are tied (we use the weak Condorcet criterion, but the reduction can be changed to work for the strong Condorcet criterion as well). Finally, it is not hard to see that d defeats x_i. Summarizing the above computations, we see that p is the winner in the input election. Thus, in an anonymized solution election, p shall be the winner as well. Given an exact cover $\mathcal{S}'$, we move all jokers to the set voters corresponding to the exact cover, that is, we move all jokers such that, for each set $S \in \mathcal{S}'$, we will have one joker in the block of set voters corresponding to S. Notice that initially the jokers form an exact cover. Notice further that they still form an exact cover in the solution, as we moved them to blocks of set voters corresponding to the sets of $\mathcal{S}'$. Specifically, considering only the jokers, the relative score of p and the x_i's did not change between the input and the solution. Therefore p is still the winner. Moreover, it is not hard to see that the election is anonymized. For the other direction, notice that the jokers are not anonymized (that is, their blocks are bad; specifically, each joker forms its own block) while all other blocks are anonymized. It is not possible to anonymize the jokers by moving other voters (or other jokers) to their blocks, as the budget is too small for that. Therefore, in a solution, all jokers should move to other blocks. There are two possibilities for each joker, either to move to the init voters or to move to some set voters. It is not a good idea to move some voters to the init voters, as the init voters prefer X to p. More formally, if in a solution, some jokers move to the init voters, then we can instead move them to an arbitrarily-chosen set voter. Therefore, we can assume that, in a solution, all jokers move to set voters. If the jokers move to set voters in a way that does not correspond to an exact cover, than at least one x_i would win over p (as at least one x_i would be covered twice, that is, at least two jokers would move to set voters preferring x_i to p; therefore, the relative score between x_i and p would change in such a way that x_i would win over p in a head-to-head contest). Therefore, a solution must correspond to an exact cover, and we are done. $\square$

Moving further to the Swap distance, we show that EA is NP-hard for both voting rules considered, and even for elections with only four voters (and therefore, also for elections with anonymity level only four; indeed, any input with $k > n$ is a trivial no-instance). Technically, in the corresponding reduction, from Kemeny Distance, we set both n and k to four. We mention that the next theorem actually holds for all voting rules which are *unanimous* (a voting rule is unanimous if for all elections where all voters prefer the same alternative c, it selects the preferred alternative c; see, for example, [14]).

Theorem 3. *For $\mathcal{R} \in \{$Plurality,Condorcet$\}$, $\mathcal{R}$-Swap-*EA *is* NP*-hard even if the number n of voters is four and the anonymity level k is four.*

Proof. We reduce from the Kemeny Distance problem.

> Kemeny Distance
> **Input:** An election $E' = (C', V')$ and a positive integer h.
> **Question:** Is the Kemeny distance of $E' = (C', V')$ at most h (where the Kemeny distance is the minimum total number of swaps of neighboring alternatives needed to have all voters vote the same; such a similar vote is called a Kemeny vote)?

Kemeny Distance is NP-hard already for four voters [13, 2]. Given an input election for Kemeny Distance, we create an instance for EA, as follows. We initialize our election $E = (C, V)$ with the election given for Kemeny Distance and create another alternative c (that is, we set $C = C' \cup \{c\}$). For each voter in the election, we place c as the first choice of the voter, that is, for each voter $v' \in V'$, we create a voter $v \in V$ with the same preference order as v' while preferring c to all other alternatives. We set k to n and s to h. This finishes the construction.

It is clear that originally c is the winner of the election (both under the Plurality rule and under the Condorcet rule; indeed, for this theorem, we only require the rule to be *unanimous*; see, for example, [14]). The crucial observation is that there is no need to swap the new alternative c; this follows because all voters already agree on him, as they place him first in their preference orders. More formally, as k is set to n, it follows that if in a solution c is swapped in some voters, then he must be swapped in all voters. Therefore, we can simply "unswap" these swaps to get a cheaper solution.

Finally, since k is set to n, it follows that all voters should vote the same in the resulting k-anonymous solution election. Thus, the best way to anonymize the election is by finding a Kemeny vote, and transforming all voters to vote as this Kemeny vote. Therefore, the election can be k-anonymized by at most s swaps if and only if the Kemeny distance of the input election is at most h, and we are done. □

With respect to the parameter number n of voters, the situation for the Discrete distance is different than the situation for the Swap distance. Specifically, it turns out that Condorcet-Discrete-EA is FPT wrt. n.

Theorem 4. *Condorcet-Discrete-*EA *is* FPT *wrt. the number n of voters.*

Proof. The crucial observation here is that there is no need to create new blocks, besides, perhaps, one arbitrarily-chosen p-block. To see this, consider a solution that adds a new block B which is not a p-block (recall that a p-block is a block of p-voters). Change the solution by moving the voters in B to a new arbitrarily-chosen p-block (instead of the block B). While using the same budget, the election is still anonymized and p is still a Condorcet winner. It follows that no new blocks, besides, perhaps, one additional arbitrarily chosen new p-block, are needed. We mention that an additional p-block might be needed when there are no p-voters (and, therefore, no p-blocks) in the input election.

The above observation suggests the following simple algorithm. We begin by guessing whether we need a new p-block. Then, for each voter, we guess whether (1) it will stay the same, (2) it will move to some other original block (out of the possible $n-1$ other original blocks), or (3) it will move to the new p-block (if we guessed that such a new block p-block exists).

Correctness follows by the observation above and by the brute-force nature of the algorithm. Fixed-parameter tractability follows since we guess for each voter (out of the n original voters), where it will end up (out of n or $n+1$ possibilities), resulting in running time $O(m \cdot n^n)$. ◻

We move on to consider the number m of alternatives. Similarly to numerous other problems in computational social choice, all variants of EA considered in this paper are FPT with respect to this parameter. This follows by applying the celebrated result of Lenstra [18] after formulating the problem as an integer linear program where the number of variables is upper-bounded by a function dependent only on the number m of alternatives.

Theorem 5. *For $\mathcal{R} \in \{$Plurality, Condorcet$\}$ and $d \in \{$Discrete, Swap$\}$, $\mathcal{R}$-d-EA is FPT wrt. the number m of alternatives.*

Proof. The crucial observation is that the number of different preference orders is $m!$, therefore upper-bounded by a function depending only on the parameter, m. We enumerate the set of $m!$ different preference orders and create a variable $x_{i,j}$, for each $i, j \in [m!]$, with the intended meaning that $x_{i,j}$ will represent the number of voters with preference order i in the input and preference order j in the solution.

We add a budget constraint:

$$\sum_{i,j \in [m!]} x_{i,j} \cdot d(i,j) \leq s.$$

(Note that we can precompute all the distances, in polynomial time.)

For preference order i, we denote the number of voters with preference order i in the input by start_i and the number of voters with preference order i in the solution by end_i. For each i, it holds that:

$$\mathrm{end}_i = \mathrm{start}_i + \sum_{j \in [m!]} x_{j,i} - \sum_{j \in [m!]} x_{i,j}.$$

We guess a set $Z \subseteq [m!]$ of preference orders with the intent that these will be the preference orders that will be present in the solution. For each preference order $i \in Z$ we add a k-anonymity constraint, and, similarly, for each preference order $i \notin Z$ we require that its end_i will be 0, as follows:

$$\forall i \in Z : \mathrm{end}_i \geq k$$
$$\forall i \notin Z : \mathrm{end}_i = 0$$

Differently for the Plurality rule and the Condorcet rule, we add more constraints to make sure that the winner will not change, as follows.

For the Plurality rule, we guess the highest score z in the solution (that is, the winning score), and we check that the initial winner p gets exactly z points, by adding the following constraint:

$$\sum_{\{i\,:\,p \text{ is at the first position of } i\}} \mathrm{end}_i = z.$$

Similarly, for each non-winning alternative $c \neq p$, we add a non-winning constraint:

$$\sum_{\{i\,:\,c \text{ is at the first position of } i\}} \mathrm{end}_i \leq z.$$

For the Condorcet rule, for the initial Condorcet winner p (if it exists), we check that he indeed beats all other alternatives in the solution:

$$\forall c \neq p : \sum_{i \in [m!] \text{ and } p \succ_i c} \mathrm{end}_i \geq \sum_{i \in [m!] \text{ and } c \succ_i p} \mathrm{end}_i.$$

Since the number of variables and the number of constraints is upper-bounded by the parameter, fixed-parameter tractability follows by applying the result of Lenstra [18]. □

4 Conclusion

Motivated by privacy issues when publishing election-related data, we initiated the study of k-anonymizing preference orders, investigating its computational complexity for the Plurality rule and the Condorcet rule, while considering the Discrete distance and the Swap distance. We showed a wide diversity of the (parameterized) complexity of EA, with respect to several natural parameters.

There are numerous opportunities for future research, some of which we briefly discuss next. Most immediately, there are still some open questions, as Table 1 suggests, and one might also consider other parameterizations as well as approximation algorithms for coping with the NP-hard cases. It is also natural to extend this line of research to other voting rules (for example, can the polynomial-time algorithm presented in Theorem 1 be extended to the Borda rule? what happens for Approval voting? what happens for multi-winner rules?). Similar in spirit, it is natural to consider other distances besides the Discrete distance and the Swap distance. While we considered a *minmax* approach, as we compute the distance between two elections as the sum of distances between their voters (and allow to permute the voters), it might also be interesting to study a *minmax* approach, where, roughly speaking, we would not allow any individual voter to change its vote by too much. This could be of particular interest as it might preserve the original election more closely. On a similar note, it is worth studying how to preserve more properties of the original election; while we only require to preserve the winner, one might require to preserve the full relative ranking of the alternatives (that is, fixing some scoring rule, and considering two alternatives c' and c'' such that c' is achieving a greater score than c'' in the original election, we might require c' to achieve greater score than c'' in the resulting election as well). Somehow related, one might consider some stronger notions of k-anonymity, for example l-diversity, which might be interesting to explore in the context of elections. Finally, one might experiment with real-world preference-data (for example, from PrefLib [21]) to evaluate the quality of the algorithms presented in this paper.

Bibliography

[1] T. Ashur and O. Dunkelman. Poster: On the anonymity of Israel's general elections. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, pages 1399–1402. ACM, 2013.

[2] T. Biedl, F. J. Brandenburg, and X. Deng. On the complexity of crossings in permutations. *Discrete Mathematics*, 309(7):1813–1823, 2009.

[3] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. *Multiagent systems*, pages 213–283, 2012.

[4] R. Bredereck, A. Nichterlein, and R. Niedermeier. Pattern-guided k-anonymity. *Algorithms*, 6(4):678–701, 2013.

[5] R. Bredereck, V. Froese, S. Hartung, A. Nichterlein, R. Niedermeier, and N. Talmon. The complexity of degree anonymization by vertex addition. In *Algorithmic Aspects in Information and Management*, pages 44–55. Springer, 2014.

[6] R. Bredereck, A. Nichterlein, R. Niedermeier, and G. Philip. The effect of homogeneity on the computational complexity of combinatorial data anonymization. *Data Mining and Knowledge Discovery*, 28(1):65–91, 2014.

[7] D. L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–90, 1981.

[8] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan. Truthful mechanisms for agents that value privacy. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 215–232. ACM, 2013.

[9] C. Clifton and T. Tassa. On syntactic anonymity and differential privacy. In *Data Engineering Workshops (ICDEW), 2013 IEEE 29th International Conference on*, pages 88–93. IEEE, 2013.

[10] E. Cuvelier, O. Pereira, and T. Peters. Election verifiability or ballot privacy: Do we need to choose? In *Computer Security–ESORICS 2013*, pages 481–498. Springer, 2013.

[11] R. G. Downey and M. R. Fellows. *Fundamentals of Parameterized Complexity*. Springer, 2013.

[12] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 9(3-4):211–407, 2013.

[13] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web*, pages 613–622. ACM, 2001.

[14] E. Elkind, P. Faliszewski, and A. Slinko. On the role of distances in defining voting rules. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 375–382. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[15] J. Flum and M. Grohe. *Parameterized Complexity Theory*. Springer, 2006.

[16] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[17] S. Hartung, A. Nichterlein, R. Niedermeier, and O. Suchỳ. A refined complexity analysis of degree anonymization in graphs. *Information and Computation*, 2014.

[18] H. W. Lenstra. Integer programming with a fixed number of variables. *Math. Oper. Res.*, 8:538–548, 1983.

[19] K. Liu and E. Terzi. Towards identity anonymization on graphs. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pages 93–106. ACM, 2008.

[20] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[21] N. Mattei and T. Walsh. Preflib: A library of preference data. In *Proceedings of Third International Conference on Algorithmic Decision Theory (ADT 2013)*, Lecture Notes in Artificial Intelligence. Springer, 2013.

[22] A. Meyerson and R. Williams. On the complexity of optimal k-anonymity. In *Proceedings of the Twenty-third ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 223–228. ACM, 2004.

[23] R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Oxford University Press, 2006.

[24] H. Nurmi, A. Salomaa, and L. Santean. Secret ballot elections in computer networks. *Computers & Security*, 10(6):553–560, 1991.

[25] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05): 557–570, 2002.