Divide and Conquer: Using Geographic Manipulation to Win District-Based Elections

Yoad Lewenberg and Omer Lev

Abstract

District-based elections, in which voters vote for a district representative and those representatives ultimately choose the winner, are vulnerable to gerrymandering, i.e., manipulation of the outcome by changing the location and borders of districts. Many countries aim to limit blatant gerrymandering, and thus we introduce a geographically-based manipulation problem, where voters must vote at the ballot box closest to them.

We show that this problem is NP-complete in the worst case. However, we present a greedy algorithm for this problem, and using it on the real-world data from the 2015 Israeli and British elections, we show that many parties are potentially able to make themselves victorious using district manipulation. Moreover, we show that the relevant variables here are not vote share, nor even seat share, but more likely some form of geographic dispersion.

1 Introduction

Voting mechanisms are commonly used to select a single option from a multitude of options. However, in some cases, an intermediary step is used. In many parliamentary democracies, the public votes for a representative of their district,[1] and those representatives choose the executive authority. For example, the electoral college in US presidential elections and Westminster-system parliaments (UK, Canada, Australia, etc.) work in this way. The technique is not unique to electoral contexts, but manifests itself in many systems which have an organizational structure: companies divided into divisions, in which each division makes a recommendation, and then division heads reach a final decision; collections of sensors interpreting input, in which each subgroup of sensors reports its understanding to a central processing unit, etc.

One of the major issues facing district-based parliamentary systems is the ability of participants in the system to manipulate it by determining the districts, influencing the outcome (so one's opponents are either a minority in many districts, or their majorities are very centralized in very few districts containing a high concentration of them). In US political jargon, this is commonly termed gerrymandering, after Massachusetts governor Elbridge Gerry, who was accused in 1812 of creating a salamander shaped district in the Boston area to benefit his party. US political parties have used this technique to manipulate elections for years [21, 8], and due to its use to disenfranchise African-American voters in some states, the US Voter Rights Act of 1965 included provisions that required district changes in several states to be approved by federal authorities [32].

In response to accusations of such manipulations, a call for more "rational" districting has been heard from many quarters [20, 34]. This is commonly understood to include a system in which voters are close to the rest of their district [34]. In a sense, voters should always go to a ballot box (or central area) that is closest to them, not one that is further

[1] Various terms are used for this: districts in the US, constituencies in the UK, ridings in Canada, etc. In this work we shall use the term district.

away. Moreover, this is a recurring problem, since setting of district boundaries is not a one-time event—due to population movement, district boundaries are constantly changing, and countries adopt mechanisms to make sure they are updated (in the US, a constitutionally mandated census triggers this; in the UK, parliament asks the Electoral Commission to do this, etc.).

However, as this paper will make clear, even in settings that seek a "rational" district division, manipulations are still possible. We consider the problem using both theoretical and empirical tools. As one of the first papers in computational social choice to include a spatial component, we examine the complexity of designing voting districts in which all voters vote in the ballot box nearest to them. We show that the complexity of finding a geographical division to make a preferred candidate win is NP-complete.

However, we present a greedy algorithm that is able to find some of the possible manipulations. We use this algorithm, combined with ballot box information from both the 2015 Israeli election and the 2015 UK election, to show how different geographic divisions would result in different winners.

2 Related Work

The effects of voting districts on election outcomes has been widely debated, particularly in the US, where gerrymandering became an issue in the early 19th century, and in the UK, where the existence of "rotten boroughs"[2] caused an outcry from the 18th century onwards, which was remedied in successive reform bills, from 1832 and onward.

Academic research in this field has dealt with the historical aspect [6, 4], the sociological aspect [25], and the legal aspect, in particular following the Voting Rights Act of 1965, focusing on particular countries (though mainly the US) [32, 21, 14].

However, the main area where this topic has been explored is in political science [8]. This analysis mainly delved into data of past elections [10, 26, 27, 22, 23, 18], along with statistical assumptions [33]; it tried to determine when gerrymandering occurs, and how to calculate some of its properties in the case of two parties. Some work was done to examine the difference between fully proportional representation vs. the outcome under winner-takes-all districts, and to find distance metrics between these two results [15, 12, 16, 11]. This included analysis using Banzhaf index and voter power.

The main topic of interest in the computational social choice community, since its initial groundbreaking papers [3, 2], was the issue of voter manipulation [36, 37, 29] and dealing with the implications of the Gibbard-Satterthwaite theorem [17, 31]. The issue of institutional manipulation has been explored to a far lesser degree. Control problems, where a central planner may influence the outcome using its power over the voting process, have been explored to some extent (e.g., [19] and the survey [9]), which included some preliminary work on dividing voters into groups [7]. Focusing only on a two-party scenario (as in the US), [28, 13] examined optimal gerrymandering strategies. More recent work [1] defined a ratio to indicate how unrepresentative a district election is, and showed a few bounds on this value. In any case, to our knowledge, no paper in the field has approached the problem from a spatial, geographic, point of view.

Closest in spirit to this paper is the work of Puppe and Tasnádi [30], which shows that dividing voters into districts in a way such that the number of representatives of a party will be proportional to its share of the votes, under some general geographic constraints, is computationally intractable.

[2]Voting districts based on centuries-old divisions which after some time contained a very small electorate, usually controlled by very few people. Most notorious of which was Old Sarum, set up in 1295 with 2 members of parliament, yet by 1831 contained only 11 voters, none of whom lived in it.

3 Preliminaries

An election $\mathcal{E}_f = (V, C)$ is comprised of a set V of n voters (possibly weighted) and a set of candidates C. Let $\pi(C)$ be the set of orders over the elements of C. Each voter $v \in V$ has a preference order $\succ_v \in \pi(C)$. A voting rule is a function $f : \pi(C)^n \to C$.

In this work we will focus on the most common voting rule, *plurality*. Under this voting rule, each voter awards a point to a single candidate, and the candidates with the maximal number of points are the winners. A tie-breaking rule $t : 2^C \to C$ is then used to select the ultimate winner of the election.

In a district-based election, $\mathcal{E}_{f-g}$ voters are divided into disjoint sets $V_1, \ldots, V_s$ such that $\cup_{i=1}^{s} V_i = V$. These sets define a set of s elections $\mathcal{E}_f^i = (V_i, C)$. The ultimate winner from amongst the winners of $\mathcal{E}_f^i$ is determined by g, which in the analysis below will be plurality combined with a threshold function (i.e., the winner will need to win a plurality of the districts, and, potentially, above a certain number of districts).

We are now ready to define the problem with which we will be dealing:

Definition 1. *The input of the* GERRYMANDERING$_f$ *problem is:*

- *A set of candidates C.*
- *A set of voters $V = \{v_1, \ldots, v_n\} \subset \mathbb{R}^2$, where every voter $v \in V$ is identified by their location on the plain, a weight $w_v > 0$ and a strict preference $\succ_v \in \pi(C)$ over C.*
- *A set of possible ballot boxes $B = \{b_1, \ldots, b_m\} \subset \mathbb{R}^2$. Each ballot box is a district.*
- *Parameters $l, k \in \mathbb{N}$, such that $l \leq k \leq m$.*
- *A target candidate $p \in C$.*

In the GERRYMANDERING$_f$ *problem, we are asked whether there is a subset of k ballot boxes $B' \subset B$, such that they define a district-based election, in which every voter votes at their closest ballot box in B', the winner at every ballot box is determined by voting rule f, then p wins in at least l ballot boxes.*

Remark 1. *Note that while we use the term gerrymandering, this is not gerrymandering as the term is commonly used: we require voters to vote at their closest ballot box, and prevent designing "unnatural" districts, that force voters to vote far from their local area.*

This prevents mathematically possible manipulations such as setting each supporter of p in their own district, and grouping all the rest in a contiguous single district.

Example 1. *Consider a* GERRYMANDERING$_{plurality}$ *instance with two candidates, 8 voters, and 4 possible ballot boxes from which we are asked to choose 3, as illustrated in Figure 1a. The voters who support candidate a are represented as circles, the voters who support candidate b are represented as triangles, and the ballot boxes are the squares. Figure 1b shows a possible selection of 3 ballot boxes (the filled squares are the selected ballot boxes) that induces a partition into three districts (the boundaries are the dashed lines) in which candidate a wins two out of the three districts and thus wins the election, while Figure 1c shows a possible selection of 3 ballot boxes such that candidate b wins two out the three districts and thus wins the election.*

4 Complexity of GERRYMANDERING$_{plurality}$

We are now ready to present the main result of this paper.

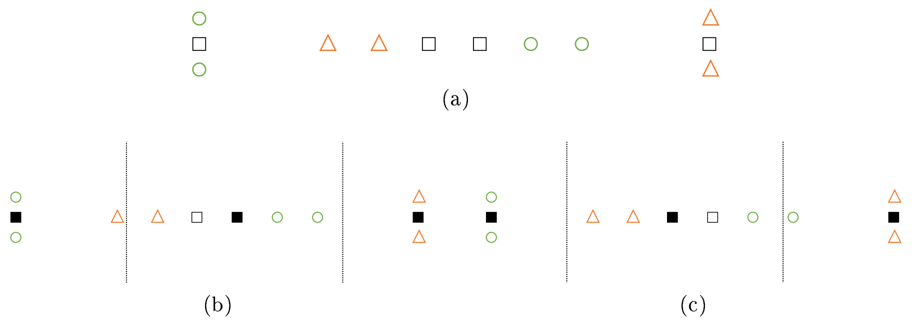


Figure 1: An example of a GERRYMANDERING$_{plurality}$ instance with two candidates (a) and two possible outcomes that result in different winners (b) and (c).

Theorem 1. *GERRYMANDERING$_{plurality}$ is NP-Complete, even when the number of candidates is a constant.*

To show GERRYMANDERING$_{plurality}$ is NP-Complete we will reduce Planar X3C, a known NP-Complete problem [5], to GERRYMANDERING$_{plurality}$.

Definition 2. *In the* Planar Exact Cover by 3-Sets (X3C) *we are given a bipartite planar graph $G = (X \cup \mathcal{S}, E)$, where $X = \{x_1, \ldots, x_{3n}\}$, $\mathcal{S} = \{S_1, \ldots, S_m\}$, every $S \in \mathcal{S}$ is $S \subseteq X$ and $|S| = 3$, and $(x, S) \in E \iff x \in S$. We are asked whether there is a subset $\bar{\mathcal{S}} \subset \mathcal{S}$ such that $\left|\bar{\mathcal{S}}\right| = n$ and for every $S, S' \in \bar{\mathcal{S}}$ it holds that $S \cap S' = \emptyset$.*

In what follows, when we are given a planar graph G, we will associate every node of G as a point in $\mathbb{R}^2$.

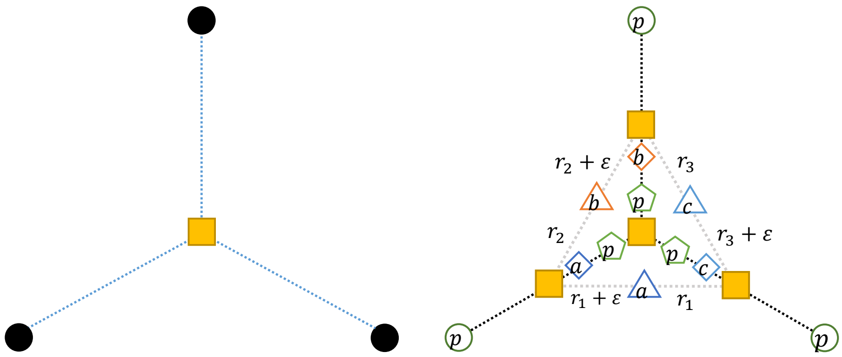
Before we can begin showing that GERRYMANDERING$_{plurality}$ is NP-Complete by reduction from Planar X3C, we must add some constraints to the Planar X3C problem.

Definition 3. *In the* Planar X3C* *problem we are given a bipartite graph G as in Planar X3C. However, the graph G when embedded on the plain has the following properties (similar to what is portrayed in Figure 2a):*

1. *For every $x \in X$, $S, S' \in \mathcal{S}$ such that $x \in S'$ and $x \notin S$: $d(x, S) < d(x, S')$ where $d(\cdot, \cdot)$ is the Euclidean distance.*
2. *For every $x, x' \in X$, $S \in \mathcal{S}$ such that $x \in S$ and $x' \notin S$: $d(x, S) < d(x', S)$.*
3. *For every $x \in X$, $S, S' \in \mathcal{S}$ such that $x \in S \cap S'$: $d(x, S) < 2d(x, S')$.*
4. *For every $x, x' \in X$, $S \in \mathcal{S}$ such that $x, x' \in S$: $d(x, S) < 2d(x', S)$.*
5. *For very two elements x, x' that belong to the same set $S \in \mathcal{S}$, the angle $\angle xSx'$ is at least $\frac{\pi}{3}$ and at most $\frac{5\pi}{6}$.*
6. *Every three elements that share a set induce a triangle, and the triangles do not overlap.*

Next, we show that adding these constraints on the planar graph do not make the problem easier.

Lemma 1. *Planar X3C* is NP-Complete.*



(a) (b)

Figure 2: (a) The Planar X3C* gadget; the circles are the elements and the square is the set. And (b) the GERRYMANDERING$_{plurality}$ gadget; the squares are the ballot boxes, the circles, pentagons, rhombi and triangles are voters V^1, V^2, V^3 and V^4, respectively, with their preferences.

Proof sketch. We start with the 3,4-SAT problem. This problem is a special case of the well-known SAT problem where each variable appears in at most four clauses, and every clause contains at most three variables. This problem is known to be NP-Complete [35]. A 3,4-SAT instance will be reduced to a Planar 3-SAT instance [24]. Next, the Planar 3-SAT instance will be reduced to a Planar 1-3-SAT [5]. Finally, the Planar 1-3-SAT instance will be reduced to Planar X3C* instance (via the reduction in [5]).

By following this reduction chain we are guaranteed that the degree of every vertex is bounded by a constant, and that all the constraints are satisfied. □

of Theorem 1. We reduce an arbitrary Planar X3C* instance, $G = (X \cup \mathcal{S}, E)$ to the following GERRYMANDERING$_{plurality}$ instance. First, we may assume that there is a map function $\pi : \mathcal{S} \times X \to \{1, 2, 3\}$ such that if $x, x' \in S$, $x \neq x$ for some $S \in \mathcal{S}$ then $\pi(S, x) \neq \pi(S, x')$. Hereafter, when we address a set $S = \{x_i, x_j, x_k\}$, we assume that $\pi(S, x_i) = 1$, $\pi(S, x_j) = 2$ and $\pi(S, x_k) = 3$.

In the reduced GERRYMANDERING$_{plurality}$ instance there are 4 candidates $C = \{p, a, b, c, d\}$; 4 sets of voters $V = V_1 \cup V_2 \cup V_3 \cup V_4$; and 2 sets of ballot-boxes $B = B_1 \cup B_2$.

For every element in X we will have a voter in V_1, that is, $V_1 = \left\{v_1^1, \dots, v_{3n}^1\right\}$. For every (x, S), $x \in S$ pair we will have a voter in V_2, a voter in V_3 and a voter in V_4. That is, $V_i = \left\{v_{x,S}^i : x \in S\right\}$ for $i = 2, 3, 4$. The weight of the voters in V_3 is 3, and the weight of the other voters is 2. Voters in V_1 and V_2 prefer candidate p; and for every set $S = \{x_i, x_j, x_k\}$, voters $v_{x_i,S}^3$ and $v_{x_i,S}^4$ prefer candidate a; voters $v_{x_j,S}^3$ and $v_{x_j,S}^4$ prefer candidate b; and voters $v_{x_k,S}^3$ and $v_{x_k,S}^4$ prefer candidate c.

For every set in $\mathcal{S}$ we will have a ballot box in B_1, that is, $B_1 = \left\{b_1^1, \dots, b_m^1\right\}$. For every (x, S), $x \in S$ pair we will have a ballot box in B_2. That is, $B_2 = \left\{b_{x,S}^2 : x \in S\right\}$. To conclude, there are $3n + 9m$ voters and $4m$ ballot boxes.

The location of a voter in V_1 will be as the location of the corresponding element, and the location of a ballot box in B_1 will be as the location of the corresponding set. Ballots and voters are organized as shown in Figure 2b:

- For every ballot box $b_{x,S}^2 \in B_2$, the location of $b_{x,S}^2$ will be on the line between x and S such that $d\left(b_{x,S}^2, S\right) = d\left(b_{x,S}^2, x\right)$.

- Location of a voter $v^3_{x,S} \in V_3$ will be on the line between $b^2_{x,S}$ and b^1_S such that $d\left(v^3_{x,S}, b^2_{x,S}\right) = \varepsilon$, for some small $\varepsilon > 0$.
- Voter $v^2_{x,S} \in V_2$ location is on the line between $v^3_{x,S}$ and b^1_S such that $d\left(v^2_{x,S}, v^3_{x,S}\right) = d\left(v^2_{x,S}, b^1_S\right)$.
- For every set $S = \{x_i, x_j, x_k\}$, set the location of $v^4_{x_i,S}$ on the line between $b^2_{x_i,S}$ and $b^2_{x_j,S}$ such that $d\left(v^4_{x_i,S}, b^2_{x_i,S}\right) = d\left(v^4_{x_i,S}, b^2_{x_j,S}\right) + \varepsilon$.
- Similarly, location of $v^4_{x_j,S}$ is on the line between $b^2_{x_j,S}$ and $b^2_{x_k,S}$ such that $d\left(v^4_{x_j,S}, b^2_{x_j,S}\right) = d\left(v^4_{x_j,S}, b^2_{x_j,S}\right) + \varepsilon$; and set the location of $v^4_{x_k,S}$ on the line between $b^2_{x_k,S}$ and $b^2_{x_i,S}$ such that $d\left(v^4_{x_k,S}, b^2_{x_k,S}\right) = d\left(v^4_{x_k,S}, b^2_{x_i,S}\right) + \varepsilon$.

At this point we should note that:

- For every $x \in S$, voter v^1_x is closer to ballot box $b^2_{x,S}$, than to ballot box b^1_S.
- For every $x \in S \cap S'$, voter v^1_x is closer to ballot box $b^2_{x,S}$, than to ballot box $b^1_{S'}$ (this holds due to requirement 3 in Definition 3).
- For every $x \in S$, $x' \in S'$ such that $x \notin S'$, voter v^1_x is closer to ballot box $b^2_{x,S}$, than to ballot box $b^1_{S'}$ and ballot box $b^2_{x',S'}$ (this holds due to requirements 1 and 2 in Definition 3).
- For every $x \in S$, voter $v^2_{x,S}$ is closest to ballot box b^1_S, then to $b^2_{x_i,S}$, and then to all other ballot boxes; and voter $v^3_{x,S}$ is closest to ballot box $b^2_{x,S}$ then to all other ballot boxes (this holds due to requirements 2 and 5 in Definition 3).
- For every $S = \{x_i, x_j, x_k\}$, voter $v^4_{x_i,S}$ is closer to ballot box $b^2_{x_j,S}$ than to ballot box $b^2_{x_i,S}$; voter $v^4_{x_j,S}$ is closer to ballot box $b^2_{x_k,S}$ than to ballot box $b^2_{x_j,S}$; and voter $v^4_{x_k,S}$ is closer to ballot box $b^2_{x_i,S}$ than to ballot box $b^2_{x_k,S}$;

The objective is to decide whether there is a subset of $2n + m$ ballot boxes such that p will win in all of them.

First, assume $G = (X \cup \mathcal{S}, E)$ is a "yes" X3C* instance. Let $\bar{\mathcal{S}} \subset \mathcal{S}$ such that $|\bar{\mathcal{S}}| = n$ and $\bigcup_{S \in \bar{\mathcal{S}}} S = X$. Let $B'_1 = \left\{b^1_i : S_i \notin \bar{\mathcal{S}}\right\}$, $B'_2 = \left\{b^2_{x,S} : S \in \bar{\mathcal{S}}, x \in S\right\}$ and finally $B' = B'_1 \cup B'_2$.

For every $v^1_i \in V_1$ there exists only one $S \in \bar{\mathcal{S}}$ such that $x_i \in S$. Therefore, $b^2_{x_i,S} \in B'$ and voter v^1_i will go to vote there.

For every $S = \{x_i, x_j, x_k\} \in \bar{\mathcal{S}}$, we have that v^1_i, $v^2_{x_i,S}$, $v^3_{x_i,S}$ and $v^4_{x_k,S}$ will vote in $b^2_{x_i,S}$, therefore p will win as they would receive 4 votes and the other candidates would get at most 3 votes. In the same way, p will also win in $b^2_{x_j,S}$ and $b^2_{x_k,S}$.

For every $S = \{x_i, x_j, x_k\} \notin \bar{\mathcal{S}}$, we have that $v^2_{x_i,S}$, $v^3_{x_i,S}$, $v^4_{x_i,S}$, $v^2_{x_j,S}$, $v^3_{x_j,S}$, $v^4_{x_j,S}$,$v^2_{x_k,S}$, $v^3_{x_k,S}$, and $v^4_{x_k,S}$ will vote in b^1_S. p will get 6 votes and every other candidate will get 5, and p will thus win.

Hence p will win in every ballot box, and there are a total of $3n+m-n = 2n+m$ ballot boxes. Thus the reduced GERRYMANDERING$_{plurality}$ instance is a "yes" instance as well.

Now, assume that the resulting GERRYMANDERING$_{plurality}$ instance is a "yes" instance, so there is $B' \subset B$ such that $|B'| = 2n + m$ and p wins in all of the ballot boxes.

Let $b^2_{x_i,S} \in B'$; it must hold that $b^1_S \notin B'$, otherwise, p could not win at this ballot box, as $v^2_{x_i,S}$ will vote at b^1_S and $v^3_{x_i,S}$ will vote at $b^2_{x_i S}$. Furthermore, assume $S = \{x_i, x_j, x_k\}$,

$v^4_{x_i,S}$ is closer to $b^2_{x_j,S}$ than to $b^2_{x_i,S}$; hence it must hold that $b^2_{x_j,S} \in B'$. In the same way, $b^2_{x_k,S} \in B'$.

Therefore, for every $S = \{x_i, x_j, x_k\}$ either $\left\{b^2_{x_i,S}, b^2_{x_j,S}, b^2_{x_k,S}\right\} \subset B'$ and $b^1_S \notin B'$, or $\left\{b^2_{x_i,S}, b^2_{x_j,S}, b^2_{x_j,S}\right\} \cap B' = \emptyset$.

For every $S = \{x_i, x_j, x_k\} \in \mathcal{S}$, let $l(S) = \left|\left\{b^2_{x_iS}, b^2_{x_jS}, b^2_{x_kS}\right\} \cap B'\right| \in \{0, 3\}$; in addition, let $t = \left|\left\{S \in \mathcal{S} : b^1_S \in B'\right\}\right|$, and finally let $r = |\{S \in \mathcal{S} : l(s) = 3\}|$.

We have that $|B'| = t + 3r \le m + 2r$, as $t + r \le m$. Moreover, $|B'| = m + 2n$, hence $n \le r$. Falsely assume that $n < r$; then there is $v^1_i \in V_1$ such that $x_i \in S \cap S'$ and $l(S) = l(S') = 3$. However, v^1_i can vote only at one ballot box, say $b^2_{x_i,S}$, thus p will loss at $b^2_{x_i,S'}$ which contradicts the assumption that p wins at every ballot box. Therefore $n = r$. Finally, let $\bar{\mathcal{S}} = \{S \in \mathcal{S} : l(S) = 3\}$; we have that $\left|\bar{\mathcal{S}}\right| = n$, and for every $S, S' \in \bar{\mathcal{S}}$ $S \cap S = \emptyset$ — otherwise as before, for $x \in S \cap S'$ where $S, S' \in \bar{\mathcal{S}}$ it holds that $b^2_{x,S}, b^2_{x,S'} \in B'$, yet p cannot win at $b^2_{x,S}$ and at $b^2_{x,S'}$. Therefore, the X3C* instance is a "yes" instance. □

Remark 2. *In the reduction we required that a specific candidate win in all ballot boxes. However, any other bound can be similarly proven, by adding dummy voters and ballot boxes "far far away".*

5 GERRYMANDERING$_{plurality}$ in the Real World

While showing GERRYMANDERING$_{plurality}$ is NP-complete in the worst case, we wish to examine its difficulty and applicability in the real world. To do so, we show a greedy algorithm, and check its applicability by running it using the 2015 Israeli and UK elections, seeing what parties we can make victorious just by adjusting the number and border of districts. Our results attest to the primacy of geographical dispersion as a key aspect, apart from voting share or parliamentary seat share.[3]

5.1 A Greedy Algorithm

As the GERRYMANDERING$_{plurality}$ problem is NP-Complete, we design a greedy algorithm that takes as input a set of candidates C, a set of voters $V = \{v_1, \ldots, v_n\} \subset \mathbb{R}^2$, each with a preference order over C, a set of ballot boxes $B = \{b_1, \ldots, b_m\} \subset \mathbb{R}^2$, a target candidate $p \in C$, and a parameter $k \le m$. The algorithm tries to find a subset of the ballot boxes B' sized k such that when every voter in V votes at their closest ballot box in B', the target candidate wins a plurality of the districts.

The greedy algorithm initially sets B' to be the full ballot boxes set, and then eliminates ballot boxes from B' one after the other, until $|B'| = k$.

The objective of the greedy algorithm, in every elimination step, is to keep the ratio between the number of ballot boxes that p wins to the number of ballot boxes that any other candidate wins as high as possible. The pseudocode of the greedy algorithm is given in Algorithm 1.

Remark 3. *The objective of the greedy algorithm described in Algorithm 1 is to find a partition to k districts such that p wins a plurality of districts, while the decision problem is to decide whether there is a partition to k districts such that p wins at least l of them. As*

[3] The elections we looked into use plurality, but, in general, this technique can still be used, in a way, to analyze GERRYMANDERING with other voting rules, by considering each data point of a set of voters as a single weighted voter with the voter's full preference being according to the the vote distribution at that particular data point.

Algorithm 1 Greedy Gerrymandering$_{plurality}$

procedure GREEDYGERRYMANDERING(V, B, k, p)
  $B' \leftarrow B$
  while $|B'| > k$ **do**
    for all $b \in B'$ **do**
      $f_b \leftarrow$ FINDRATIO(B', b, V, p)
    end for
    $b \leftarrow \arg\max_{b \in B'} \{f_b\}$
    $B' \leftarrow B' \setminus \{b\}$
  end while
  if p wins a plurality of ballot boxes **then**
    return $True$
  else
    return $False$
  end if
end procedure

procedure FINDRATIO(B, b, V, p)
  $B' = B \setminus \{b\}$
  return $\frac{|\{\tilde{b} \in B' : p \text{ wins in } \tilde{b}\}}{\max_{c \in C, c \neq p} |\{\tilde{b} \in B' : c \text{ wins in } \tilde{b}\}}$
end procedure

noted in Remark 2, the two problems are essentially equivalent; furthermore, Algorithm 1 can be easily modified to meet the other objective.

5.2 Israeli Results

The dataset of the 2015 Israeli legislative election[4] contains the number of voters and vote distribution in every Israeli city, town, village, and hamlet. We considered each location both a voter and as a possible ballot box. All voters in a particular locale were considered as if they were living in the same central location. The location of the voters and the ballot boxes is the geographic location of the place itself. We had in our dataset 1098 locales.

In Figure 3 we show parties that won at least some of the 120 seats in the Israeli 20th Knesset, the percentage of the popular vote, and the percentage of Knesset seats won in the election. Moreover, for every party the graphs show the maximal and minimal number of districts such that Algorithm 1 finds a partition to that number of districts such that the party wins the plurality of districts.

As Figure 3 shows, neither the percentage of votes nor the number of seats have a monotonic effect on the partition that the greedy algorithm finds. For example, the *Kulanu* party won 7.49% of the votes, and the greedy algorithm could not find a partition to districts such that this party would have won a plurality of districts; while the *Jewish Home* party and the *United Torah Judaism* party both won below 7% of the popular vote, yet are able to win in a significant number of district allocations, attesting to their particular geographical voter dispersal pattern, which maintains majorities in geographically significant parts of the country.

We can also examine how close elections were, using the algorithm's goal function, as can be seen in Figure 4. Notice how, despite its lower overall vote share, the Zionist Union is

[4] http://votes20.gov.il/

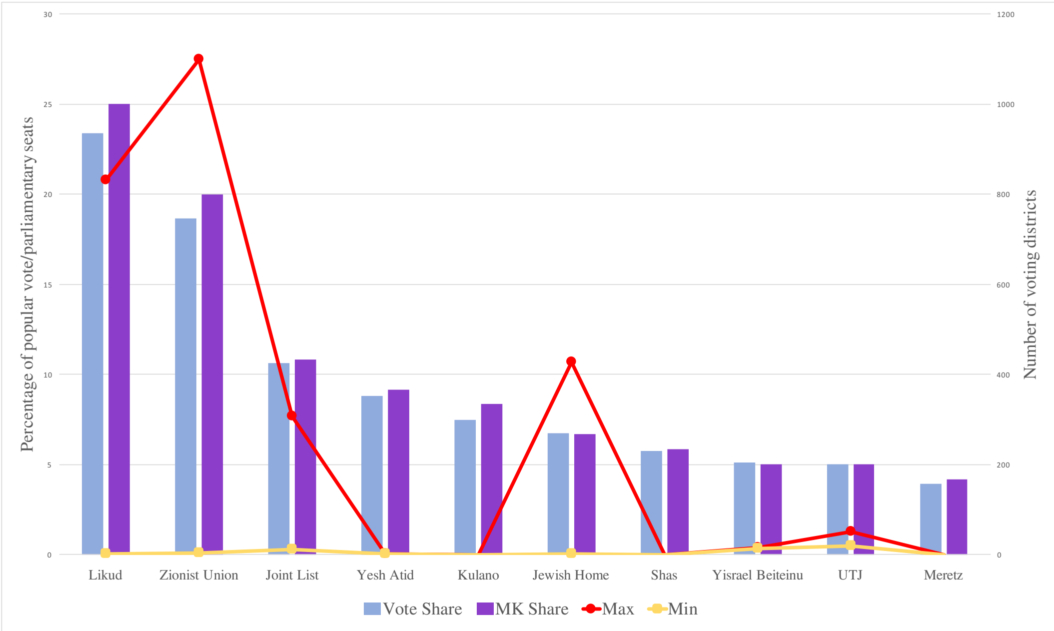


Figure 3: Maximal and minimal district number that can enable a party to become the plurality winner (line graphs) compared to share of popular vote and MK share (bar graphs)

able to gain a massive number of seats over other parties, thanks to its particular spread-out distribution in the country.

5.3 United Kingdom Results

The dataset of the 2015 British general election[5] consists of the number of votes for every party in every constituency. As in the dataset of the Israeli elections, all voters in a particular constituency were considered as if they were living in the same central location. We had in our dataset 650 constituencies.

Figure 5 shows for every party that won one of the 650 seats in the 56th Parliament of the United Kingdom, its percent of the popular vote, its percent of parliamentary seats, and the number of districts the UK can be divided into, that result in the party winning a plurality of districts.

As in the Israeli case, Figure 5 shows that neither the percentage of votes nor the number of parliamentary seats has a monotonic effect on the partition that the greedy algorithm finds. Note that while geographically focused parties—in Scotland (SNP), Wales (Plaid Cymru), and Northern Ireland (Sinn Féin, SDLP and UUP)—are able to find winning gerrymandering, more widely-supported, yet widely-dispersed parties, such as UKIP and the Green Party, were not able to do so.

Once more, Figure 6 allows us to look at how close elections were, using the algorithm's goal function. Because it is well-spread over a swath of the UK, the SNP is able at a lower district level (where its further-away competitors are condensed into fewer districts) to achieve a significant margin of victory.

[5] http://www.electoralcommission.org.uk/our-work/our-research/electoral-data

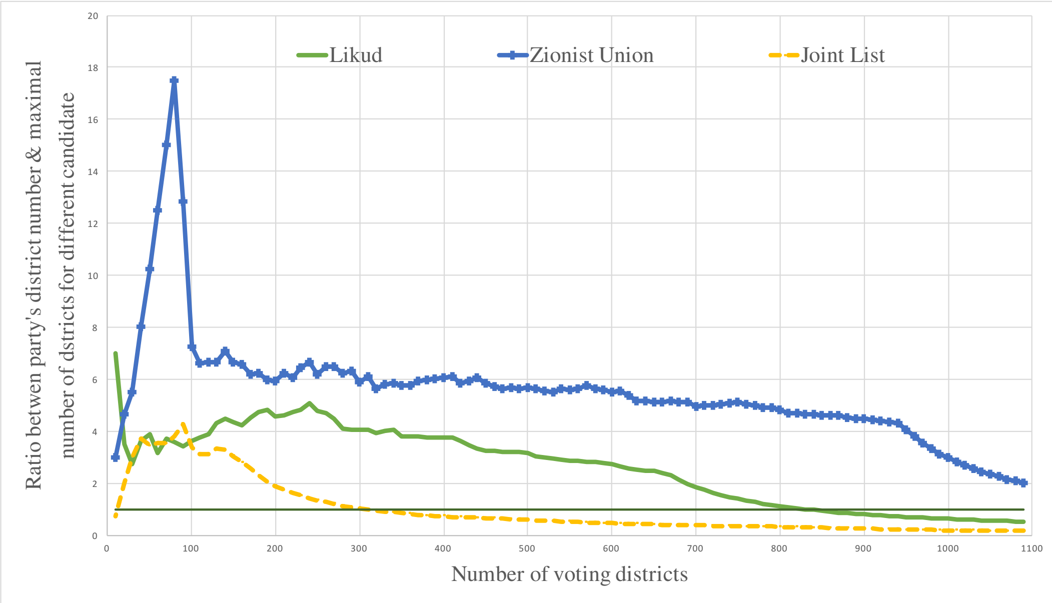


Figure 4: Ratio between the number of districts the party won and the maximal number of districts that any other party won; in every iteration of Algorithm 1. When the values are above 1, the greedy algorithm finds a partition where the party wins a plurality of districts.

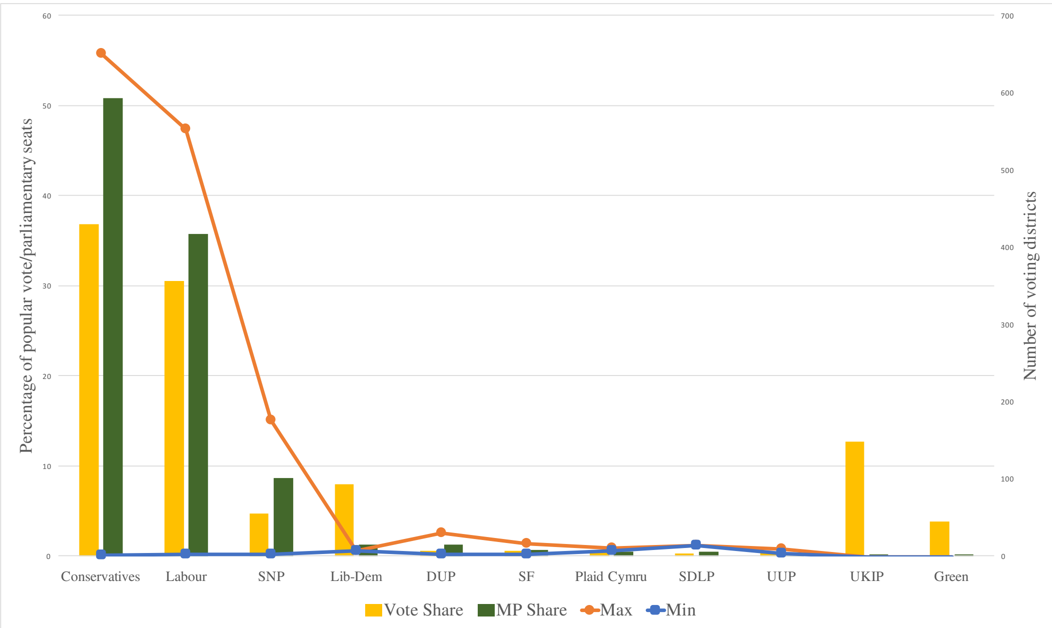


Figure 5: Maximal and minimal district number that can enable a party to become the plurality winner (line graphs) compared to share of popular vote and MP share (bar graphs)

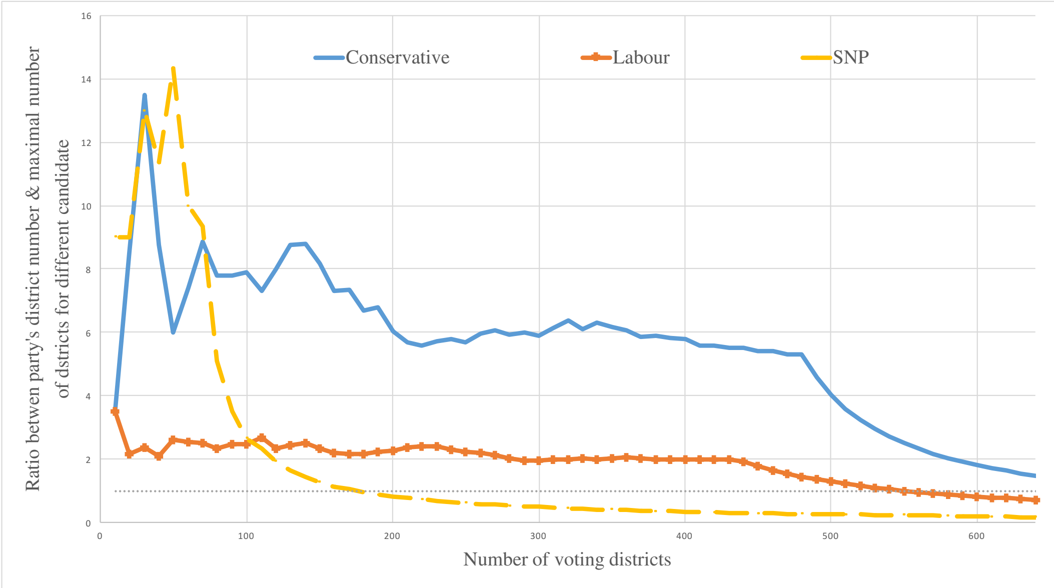


Figure 6: Ratio between the number of districts the party won and the maximal number of districts that any other party won; in every iteration of Algorithm 1.

6 Discussion

In this work we introduced the GERRYMANDERING problem, a control manipulation problem that is based on winning district-based elections by using a particular division of a spatial area into districts. We then examined solving this problem with a greedy algorithm using real-world data from the 2015 Israeli and UK elections.

It is obvious that if district lines were completely arbitrary, the problem would be trivial: as was possible in pre-1832 Britain, with its multitude of rotten boroughs, one could define particular voters as a district on their own, while putting a mass of voters into a single district, thus ensuring victory. However, our requirement that voters vote in the nearest geographical district to them prevents such barefaced gerrymandering (in fact, Elbridge Gerry's own salamander shaped district would not be allowed in our setting...), in line with current efforts to limit the possibility of gerrymandering.

Our empirical work shows that the issue determining possibility of manipulation is not directly related to voting share or parliamentary weight (which, particularly as Britain has a district-based election system, is surprising). The geographical dispersion of the voters plays a major part in this, and we hope further research will do more to investigate the various variables that come into play when manipulation is geographically based.

An issue we have put aside in this first analysis of spatial voting manipulations is the size of each district. While there are certainly disparities in many countries' district sizes (in the US Senate, the "district" of California contains more than 70 times the population of the "district" of Wyoming, both having equal representation), many countries do aim to achieve a rough equivalence in district sizes. Our theoretical result shows the complexity is NP-complete when allowing districts different from each other by a fixed factor of only $2\frac{1}{3}$ (or more). Our empirical results did not consider this issue; though, for example, one of the divisions in which Labour won a plurality of seats in the British election had the maximal district only $2\frac{2}{3}$ larger than the smallest one.

References

[1] Yoram Bachrach, Omer Lev, Yoad Lewenberg, and Yair Zick. Misrepresentation in district voting. *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, 2016.

[2] John J. Bartholdi III and James B. Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

[3] John J. Bartholdi III, Craig A. Tovey, and Michael A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[4] D. Butler. The electoral process the redrawing of parliamentary boundaries in britain. *British Elections and Parties Yearbook*, 2(1), 1992.

[5] Martin E. Dyer and Alan M. Frieze. Planar 3dm is np-complete. *Journal of Algorithms*, 7(2):174–184, 1986.

[6] E. J. Engstrom. Stacking the states, stacking the house: The partisan consequences of congressional redistricting in 19th century. *APSR*, 100:419–427, 2006.

[7] G. Erdélyi, E. Hemaspaandra, and L. A. Hemaspaandra. More natural models of electoral control by partition. In *ADT*, pages 396–413, 2015.

[8] R. S. Erikson. Malapportionment, gerrymandering, & party fortunes in congressional elections. *APSR*, 66(4):1234–1245, 1972.

[9] P. Faliszewski, E. Hemaspaandra, and L. A. Hemaspaandra. Using complexity to protect elections. *CACM*, 53(11):74–82, 2010.

[10] M. R. Feix, D. Lepelley, V. R. Merlin, and J. Rouet. The probability of conflicts in a U.S. presidential type election. *Economic Theory*, 23(2):227–257, 2004.

[11] M.R. Feix, D. Lepelley, V.R. Merlin, J. Rouet, and L. Vidu. Majority efficient representation of the citizens in a federal union. Technical report, Université de la Réunion, de Caen, d'Orleans, 2008.

[12] D. S. Felsenthal and M. Machover. Minimizing the mean majority deficit: The 2nd square-root rule. *Math. Soc. Sci.*, 37(1):25–37, 1999.

[13] J. N. Friedman and R. T. Holden. Towards a theory of optimal partisan gerrymandering. Technical report, Harvard University, 2005.

[14] John N. Friedman and Richard T. Holden. The rising incumbent reelection rate: What's gerrymandering got to do with it? *The Journal of Politics*, 71:593–611, 4 2009.

[15] M. Gallagher. Proportionality, disproportionality and electoral systems. *Electoral Studies*, 10(1):33–51, 1991.

[16] Andrew Gelman, Jonathan N. Katz, and Francis Tuerlinckx. The mathematics and statistics of voting power. *Statistical Science*, 17(4):420–435, 2002.

[17] Allan Gibbard. Manipulation of voting schemes. *Econometrica*, 41(4):587–602, July 1973.

[18] B. Grofman, W. Koetzle, and T. Brunell. An integrated perspective on the three potential sources of partisan bias: Malapportionment, turnout differences, & the geographic distribution of party vote shares. *Electoral Studies*, 16(4):457–470, 1997.

[19] E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Anyone but him: The complexity of precluding an alternative. *AIJ*, 171(5–6):255–285, 2007.

[20] Christopher Ingraham. One easy way to end gerrymandering: Stop letting politicians draw their own districts. Washington Post: Wonkblog, June 2014.

[21] S. Issacharoff. Gerrymandering & political cartels. *Harvard Law Review*, 116(2):593–648, 2002.

[22] R. Lahrach and V. Merlin. *Electoral Systems: Paradoxes, Assumptions, and Procedures*, chapter Which Voting Rule Minimizes the Probability of the Referendum Paradox? Lessons from French Data, pages 129–150. Springer, 2011.

[23] R. Lahrach and V. R. Merlin. Assessing the probability of the referendum paradox: The french local election case. Voting Power in Practice workshop, 2010.

[24] David Lichtenstein. Planar formulae and their uses. *SIAM journal on computing*, 11(2):329–343, 1982.

[25] D. Lublin. *The Paradox of Representation: Racial Gerrymandering and Minority Interests in Congress.* Princeton University Press, 1997.

[26] N.R. Miller. *Electoral Systems: Paradoxes, Assumptions & Procedures*, chapter Election Inversions by the U.S. Electoral College, pages 93–127. Springer, 2011.

[27] N.R. Miller. The house size effect and the referendum paradox in u.s. presidential elections. *Electoral Studies*, 35:265–271, September 2014.

[28] G. Owen and B. Grofman. Optimal partisan gerrymandering. *Political Geography Quarterly*, 7(1):5–22, January 1988.

[29] A. D. Procaccia, J. S. Rosenschein, and A. Zohar. Multi-winner elections: Complexity of manipulation, control & winner-determination. In *IJCAI*, pages 1476–1481, 2007.

[30] Clemens Puppe and Attila Tasnádi. A computational approach to unbiased districting. *Mathematical and Computer Modelling*, 48(9–10):1455 – 1460, 2008. Mathematical Modeling of Voting Systems and Elections: Theory and Applications.

[31] Mark Allen Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, April 1975.

[32] P.H. Schuck. The thickest thicket: Partisan gerrymandering and judicial regulation of politics. *Columbia Law Review*, 87(7):1325–1384, 1987.

[33] A. Tangian. Computational application of the mathematical theory of democracy to Arrow's impossibility theorem (how dictatorial are Arrow's dictators?). *SCW*, 35(1):129–161, 2010.

[34] Jon Terbush. How to fix gerrymandering and stop future shutdowns. The Week, October 2013.

[35] Craig A Tovey. A simplified np-complete satisfiability problem. *Discrete Applied Mathematics*, 8(1):85–89, 1984.

[36] L. Xia, M. Zuckerman, A. D. Procaccia, V. Conitzer, and J. S. Rosenschein. Complexity of unweighted coalitional manipulation under some common voting rules. In *IJCAI*, pages 348–353, 2009.

[37] M. Zuckerman, A. D. Procaccia, and J. S. Rosenschein. Algorithms for the coalitional manipulation problem. In *SODA*, pages 277–286, 2008.

Yoad Lewenberg
The Rachel and Selim Benin School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem, Israel
Email: `yoadlew@cs.huji.ac.il`

Omer Lev
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada
Email: `omerl@cs.toronto.edu`