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

Yoad Lewenberg[1]
yoadlew@cs.huji.ac.il

Omer Lev[2]
omerl@cs.toronto.edu

Jeffrey S. Rosenschein[1]
jeff@cs.huji.ac.il

[1]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel
[2]University of Toronto, Ontario, Canada

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 the problem; testing it both on simulation data as well as on 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 go beyond share of the vote; the form of geographic dispersion also plays a crucial role.

CCS Concepts

•Theory of computation → Solution concepts in game theory; •Applied computing → Sociology;

Keywords

Voting; Gerrymandering; Districts

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 that 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 [22, 9], 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 [36].

In response to accusations of such manipulations, a call for more "rational" districting has been heard from many quarters [21, 39]. This is commonly understood to include a system in which voters are close to the rest of their district [39]. In a sense, voters should always go to a ballot box (or central area) that is closest to them, not one that is further 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 every decade; 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.

[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.

Appears in: *Proc. 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.*

However, we present a greedy algorithm that is able to find some of the possible manipulations, and can be extended to satisfy various constraints, appropriate for specific settings. We use this algorithm with two data sets:

Simulations A set of simulations that try to mimic a distribution of population and political views, and examining whether relatively weak candidates can still become winners, given a particular district structure.

Real world data 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 [7, 5], the sociological aspect [27], and the legal aspect, in particular following the Voting Rights Act of 1965, focusing on particular countries (though mainly the US) [36, 22, 15].

However, the main area where this topic has been explored is in political science [9]. This analysis mainly delved into data of past elections [12, 28, 29, 24, 23, 19], along with statistical assumptions [37]; 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 versus the outcome under winner-take-all districts, and to find distance metrics between these two results [16, 13, 17, 11]. This included analysis using the Banzhaf index and voter power.

The computational social choice community, since the seminal initial papers [3, 2], has dedicated a significant effort to the issue of voter manipulation [42, 43, 31] and dealing with the implications of the Gibbard-Satterthwaite theorem [18, 35]. 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., [20] and the survey [10]), which included some preliminary work on dividing voters into groups [8]. Focusing only on a two-party scenario (as in the US), [30, 14] examined optimal gerrymandering strategies (see overview of computational social choice literature, including control in [4, 34], and more particularly on gerrymandering in [38]). More closely related to our concerns, Bachrach et al. [1] defined a ratio to indicate how unrepresentative a district election is, and showed a few bounds on this value (and used voter simulations based on Mallow's model of preferences, which we use as well). In any case, to our knowledge, no paper in the field has approached the problem from a spatial, geographic, point of view.

[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 these 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.

Closest in spirit to this paper is the work of Puppe and Tasnádi [32, 33], 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. Several techniques on how to "fairly" divide areas into districts (or how to dissolve a district [41]) have been examined [25].

3. PRELIMINARIES

An election $\mathcal{E}_f = (C, V)$ 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 $\mathcal{E}_f^i = (C, V_i)$. 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 plane, 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.*

We use weights in order to ease handling of multiple voters in the same location (which, therefore, all vote at the same place) in order to ease handling of future proofs.

REMARK 2. *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.*

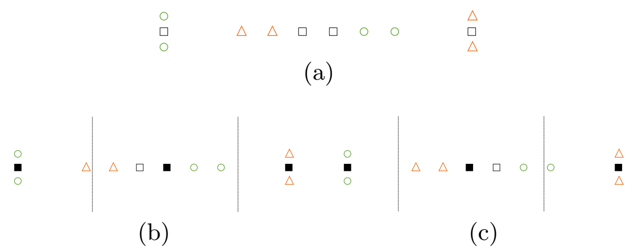


Figure 1: An example of a GERRYMANDERING$_{plurality}$ instance with two candidates (a) and two possible outcomes, (b) and (c), that result in different winners.

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.

In real life, not only are non-contiguous districts very rare, but in many democratic countries (e.g., the UK) district division is done according to objective criteria, like compactness (a desirable property also noted by [38]); we believe our "vote at nearest ballot" approximates many such attempts. The United States does not have an independent process for division of districts, but some states have set up independent bodies to do this.

EXAMPLE 3. *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 squares represent potential ballot boxes. 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. THE COMPLEXITY OF GERRYMANDERING$_{PLURALITY}$

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

THEOREM 4. 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 [6], to GERRYMANDERING$_{plurality}$.

DEFINITION 5. *In the* Planar Exact Cover by 3-Sets (X3C) *we are given a bipartite planar graph $G = (X \cup \mathcal{S}, E)$, where $X = \{x_1, \dots, x_{3n}\}$, $\mathcal{S} = \{S_1, \dots, S_m\}$, and every element in $\mathcal{S}$ in connected to exactly three elements in X, that is, for every $S \in \mathcal{S}$ it holds that $|\{x \in X : (x, S) \in E\}| = 3$. We are asked whether there is a subset $\bar{\mathcal{S}} \subset \mathcal{S}$ such that every element of X is connected to exactly one member of $\bar{\mathcal{S}}$, i.e., for all $x \in X$ it holds that $\left|\{S \in \bar{\mathcal{S}} : (x, S) \in E\}\right| = 1$.*

In what follows, when we are given a planar graph G, we will associate every node of G as a point in $\mathbb{R}^2$. In addition, we identify every element in $\mathcal{S}$ as a subset of X that contains only the elements to which it is connected.

Before we can begin showing that our problem is NP-complete by reduction from Planar X3C, we must add some constraints to the Planar X3C problem.

DEFINITION 6. *In the* Planar X3C* *problem we are given a bipartite graph G as in Planar X3C. However, the graph G when embedded on the plane 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 does not make the problem easier.

LEMMA 7. *Planar X3C* is NP-complete.*

PROOF SKETCH. In order to show that Planar X3C* is NP-complete, we start with a 3,4-SAT instance. The 3,4-SAT 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 [40]. A 3,4-SAT instance will be reduced to a Planar 3-SAT instance [26]. Next, the Planar 3-SAT instance will be reduced to a Planar 1-3-SAT instance [6]. Finally, the Planar 1-3-SAT instance will be reduced to Planar X3C* instance (via the reduction in [6]).

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. □

PROOF OF THEOREM 4. 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 $C = \{p, a, b, c\}$;
- 4 sets of voters $V = V^p_o \cup V^p_c \cup V^o_l \cup V^o_t$; and
- 2 sets of ballot-boxes $B = B_{\mathcal{S}} \cup B_O$.

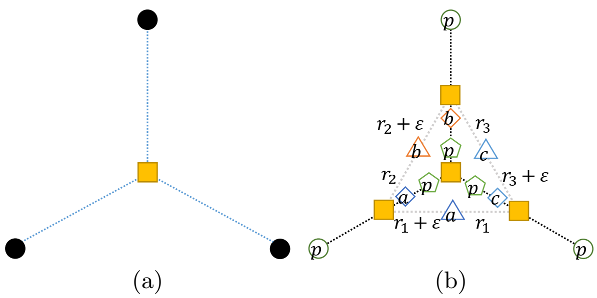


Figure 2: (a) The Planar X3C* gadget; the circles are the elements and the square is the set. And (b) the GERRYMANDERING$_{plurality}$ gadget; the squares are the ballot boxes, the circles, pentagons, rhombi and triangles are voters V_o^p, V_c^p, V_l^o and V_t^o, respectively, along with their preferences. Voters in V_o^p are located in the *outer* part of the gadget; voters in V_o^p are located near the *center* of the gadget; voters in V_l^o are located on the *line* between the central ballot box and an outer ballot box; and the voters in V_t^o are located on the inner *triangle* of the gadget.

For every element in X we will have a voter in V_o^p, that is, $V_o^p = \{v_{x_1}, \ldots, v_{x_{3n}}\}$. For every (x, S), $x \in S$ pair we will have a voter in V_c^p, a voter in V_l^o and a voter in V_t^o. That is, for every (x, S), $x \in S$ pair we have three different voters, we will distinguish between those voters by the set they belong to. The weight of the voters in V_t^o is 3, and the weight of all other voters is 2.

Voters in V_o^p and V_c^p prefer candidate p, and voters in V_l^o and V_t^o prefer some *other* candidate as follows: for every set $S = \{x_i, x_j, x_k\}$,

- The voters in V_l^o and V_t^o that represent (x_i, S) prefer candidate a;
- The voters in V_l^o and V_t^o that represent (x_j, S) prefer candidate b; and
- The voters in V_l^o and V_t^o that represent (x_k, S) prefer candidate c;

The set of ballot boxes $B_{\mathcal{S}}$ will have one ballot box for every set in $\mathcal{S}$. That is, $B_{\mathcal{S}} = \{b_{S_1}, \ldots, b_m\}$. B_O is the set of *outer*-ring ballot boxes; B_O will consist of one ballot box for every (x, S), $x \in S$ pair. That is, $B_O = \{b_{x,S} : x \in S\}$. To conclude, there are $3n + 9m$ voters and $4m$ ballot boxes.

Figure 2b illustrates one GERRYMANDERING$_{plurality}$ gadget. The three outer-ring ballot boxes are members of B_O, and the one central ballot box is a member of $B_{\mathcal{S}}$.

The location of a voter in V_o^p will be as the location of the corresponding element, on the *outer* triangle that the gadget creates. The location of a ballot box in $B_{\mathcal{S}}$ will be as the location of the corresponding set.

Ballots and voters are organized as shown in Figure 2b:

- For every ballot box $b_{x,S} \in B_O$, the location of $b_{x,S}$ will be on the line between x and S such that $d(b_{x,S}, S) = d(b_{x,S}, x)$.
- Location of a voter $v_{x,S} \in V_l^o$ will be on the *line* between $b_{x,S}$ and b_S such that $d(v_{x,S}, b_{x,S}) = \varepsilon$, for some small $\varepsilon > 0$.
- Voter $v_{x,S} \in V_c^p$ location is on the line between $v_{x,S} \in V_l^o$ and b_S, near the *center* of the gadget, such that
$$d\left(v_{x,S(\in V_c^p)}, v_{x,S(\in V_l^o)}\right) = d\left(v_{x,S(\in V_c^p)}, b_S\right).$$
- For every set $S = \{x_i, x_j, x_k\}$, set the location of $v_{x_i,S} \in V_t^o$ on the line between $b_{x_i,S}$ and $b_{x_j,S}$, in the inner *triangle* of the gadget, such that
$$d(v_{x_i,S}, b_{x_i,S}) = d(v_{x_i,S}, b_{x_j,S}) + \varepsilon.$$
- Similarly, the location of $v_{x_j,S} \in V_t^o$ is on the line between $b_{x_j,S}$ and $b_{x_k,S}$ such that
$$d(v_{x_j,S}, b_{x_j,S}) = d(v_{x_j,S}, b_{x_k,S}) + \varepsilon.$$
- Finally, set the location of $v_{x_k,S} \in V_t^o$ on the line between $b_{x_k,S}$ and $b_{x_i,S}$ such that
$$d(v_{x_k,S}, b_{x_k,S}) = d(v_{x_k,S}, b_{x_i,S}) + \varepsilon.$$

An example of a reduced GERRYMANDERING$_{plurality}$ gadget is given in Figure 2b.

At this point we should note that:

- For every $x_i \in S$, voter $v_{x_i} \in V_o^p$ is closer to ballot box $b_{x_i,S}$, than to ballot box b_S.
- For every $x \in S \cap S'$, voter $v_x \in V_o^p$ is closer to ballot box $b_{x,S}$, than to ballot box $b_{S'}$ (this holds due to requirement 3 in Definition 6).
- For every $x \in S$, $x' \in S'$ such that $x \notin S'$, voter $v_x \in V_o^p$ is closer to ballot box $b_{x,S}$, than to ballot box $b_{S'}$ and ballot box $b_{x',S'}$ (this holds due to requirements 1 and 2 in Definition 6).
- For every $x \in S$, voter $v_{x,S} \in V_c^p$ is closer to ballot box b_S, than to $b_{x,S}$, and than to all other ballot boxes; and voter $v_{x,S} \in V_l^o$ is closer to ballot box $b_{x,S}$ than to all other ballot boxes (this holds due to requirements 2 and 5 in Definition 6).
- For every $S = \{x_i, x_j, x_k\}$, voter $v_{x_i,S} \in V_t^o$ is closer to ballot box $b_{x_j,S}$ than to ballot box $b_{x_i,S}$; voter $v_{x_j,S} \in V_t^o$ is closer to ballot box $b_{x_k,S}$ than to ballot box $b_{x_j,S}$; and voter $v_{x_k,S} \in V_t^o$ is closer to ballot box $b_{x_i,S}$ than to ballot box $b_{x_k,S}$;

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 $\left|\{S \in \bar{\mathcal{S}} : (x, S) \in E\}\right| = 1$, for every $x \in X$. It must hold that $S \cap S' = \emptyset$ for every $S \neq S' \in \mathcal{S}$, $\bigcup_{S \in \bar{\mathcal{S}}} S = X$ and thus $\left|\bar{\mathcal{S}}\right| = n$. Now, let $B'_{\mathcal{S}} = \left\{b_S : S \notin \bar{\mathcal{S}}\right\}$, $B'_O = \left\{b_{x,S} : S \in \bar{\mathcal{S}}, x \in S\right\}$ and finally $B' = B'_{\mathcal{S}} \cup B'_O$.

For every $v_x \in V_o^p$ there exists only one $S \in \bar{\mathcal{S}}$ such that $x \in S$. Therefore, $b_{x,S} \in B'$ and voter v_x will go to vote there.

For every $S = \{x_i, x_j, x_k\} \in \bar{\mathcal{S}}$, we have that $v_{x_i} \in V_o^p$, $v_{x_i,S} \in V_c^p$, $v_{x_i,S} \in V_l^o$ and $v_{x_k,S} \in V_t^o$ will vote in $b_{x_i,S}$, 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 $b_{x_j,S}$ and $b_{x_k,S}$.

For every $S = \{x_i, x_j, x_k\} \notin \bar{\mathcal{S}}$, we have that voters $v_{x_i,S}, v_{x_j,S}, v_{x_k,S} \in V_o^p$, voters $v_{x_i,S}, v_{x_j,S}, v_{x_k,S} \in V_l^o$ and voters $v_{x_i,S}, v_{x_j,S}, v_{x_k,S} \in V_t^o$, will vote in b_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 $b_{x,S} \in B'$; it must hold that $b_S \notin B'$, otherwise, p could not win at this ballot box, as $v_{x,S} \in V_c^p$ will vote at b_S and $v_{x,S} \in V_l^o$ will vote at $b_{x,S}$.

Furthermore, assume $S = \{x_i, x_j, x_k\}$, $v_{x_i,S} \in V_t^o$ is closer to $b_{x_j,S}$ than to $b_{x_i,S}$; hence it must hold that $b_{x_j,S} \in B'$. In the same way, $b_{x_k,S} \in B'$.

Therefore, for every $S = \{x_i, x_j, x_k\}$ we have that either $\{b_{x_i,S}, b_{x_j,S}, b_{x_j,S}\} \cap B' = \emptyset$, or $\{b_{x_i,S}, b_{x_j,S}, b_{x_k,S}\} \subset B'$ and $b_S \notin B'$.

For every $S = \{x_i, x_j, x_k\} \in \mathcal{S}$, let

$$l(S) = \left|\{b_{x_i,S}, b_{x_j,S}, b_{x_k,S}\} \cap B'\right| \in \{0, 3\},$$

in addition, let $t = |\{S \in \mathcal{S} : b_S \in B'\}|$, and finally let $r = |\{S \in \mathcal{S} : l(s) = 3\}|$.

We have that $|B'| = t + 3r \leq m + 2r$, as $t + r \leq m$. Moreover, $|B'| = m + 2n$, hence $n \leq r$. Falsely assume that $n < r$; then there is $v_x \in V_o^p$ such that $x \in S \cap S'$ and $l(S) = l(S') = 3$. However, v_x can vote only at one ballot box, say $b_{x,S}$, thus p will lose at $b_{x,S'}$, 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 $|\bar{\mathcal{S}}| = n$, 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 $b_{x,S}, b_{x,S'} \in B'$, yet p cannot win at $b_{x,S}$ and at $b_{x,S'}$. Therefore, the X3C* instance is a "yes" instance. $\square$

REMARK 8. *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$_{PLURALITY}$ GREEDY ALGORITHM

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 both simulations, which allow us to play with various variables, as well as by analyzing the 2015 Israeli and UK elections, seeing what parties we can make victorious just by adjusting the number and border of districts.

We construct 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 $B = \{b_1, \ldots, b_{S_m}\} \subset \mathbb{R}^2$, a target candidate $p \in C$, and a parameter $k \leq m$. 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 $|B'| = k$.

Algorithm 1 Greedy Gerrymandering$_{plurality}$

```
procedure GreedyGerrymandering(V, B, k, p)
    B' ← B
    while |B'| > k do
        for all b ∈ B' do
            f_b ← findRatio(B', b, V, p)
        end for
        b ← arg max_{b∈B'} {f_b}
        B' ← B' \ {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 \ {b}
    return |{b̃∈B' : p wins in b̃}| / max_{c∈C,c≠p} |{b̃∈B' : c wins in b̃}|
end procedure
```

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 9. *The objective of the greedy algorithm as 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 noted in Remark 8, the two problems are essentially equivalent; furthermore, Algorithm 1 can be easily modified to meet the other objective.*

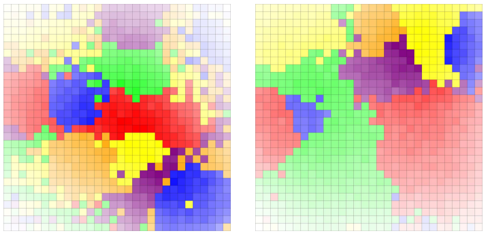
This algorithm can produce very lopsided districts—some with many voters, some with far fewer. However, it can easily incorporate a bound on this, by not eliminating ballot boxes that result in too small (or large) districts and striving for equality of size if no candidate can win (in findRatio). In our running of the algorithm we chose the current bound of the US Senate, which allows equal representation for each state, so the districts we analyze will never be more lopsided than the US Senate is. Naturally, this number can be tweaked for particular settings.

6. GERRYMANDERING$_{PLURALITY}$ SIMULATIONS

We start by exploring the possibilities of gerrymandering in a set of simulations, to give us a better understanding of how effective it may be. Simulations have an advantage over real world data in that the former can include many more variables and considerations to which we do not have access in real-world data.

6.1 Simulation Setup

We chose a geographic grid of 30x30 (i.e., with 900 initial ballots boxes), to distribute people on, in order to keep simulation size manageable. We randomly assigned 10 points as major "cities", defining for each a normal (i.e., Gaussian) distribution with a mean of the city point (and a variance



(a) (b)

Figure 3: Sample elections with variance of ϕ distribution 0.5 (a) and 0.25 (b). The color of each cell corresponds to the candidate with the highest plurality score. The brightness represents the number of voters in that cell.

of 7, for a wide metropolitan area). Each of the 1,000,000 voters is first randomly assigned to belong to one of the cities, and then assigned a particular location according to the city's normal distribution. Each setting was run several thousand times.

To set voter preferences we wanted to both create diversity, as well as to approximate the generally observed urban-rural vote divide. Following much previous research, we chose to use multiple Mallows distributions. A Mallows distribution assumes a preference order $\sigma \in \pi(C)$ which is a ground truth, and a value $\phi \in [0, 1]$, which indicates how probable are votes that are far from σ (when $\phi = 0$, all voters vote σ; when $\phi = 1$, voters' votes are uniformly random).

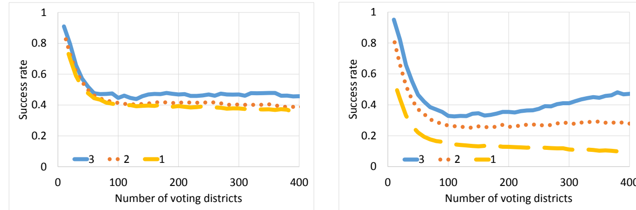
In order to create a more realistic distribution of votes, which commonly included an urban-rural split, we used the normal distribution to create a dispersion of votes aligned with the distance from the "city". We assign each city a ground truth (to be explained below), and each voter's preference order is chosen from a different distribution: use the same normal distribution that allocated voters' location, and use the absolute distance from the city to determine the value of ϕ for each voter, with the variance of the normal distribution[3] being of 0.5 or 0.25 (we present both results). Therefore, if a voter is distant from the city, they also have a higher probability of differing in their views from the city's ground truth.

Finally, we have 6 candidates, but we did not want to create them equal—if they were equally popular, their expected chance of success would be equal. Therefore, we intentionally created an unequal situation, in which some candidates are more powerful than others. We did this by assigning 3 of the cities a ground truth in which the same candidate is leading; an additional 2 cities had a ground truth of a second candidate, and another 2 had a ground truth of a third. The remaining 3 cities each had a ground truth with different candidates winning. Therefore, we had one initially strong candidate (with 3 cities' ground truth), 2 medium strong candidates (each with 2 cities' ground truth), and 3 weak candidates (each with a single city). Figure 3 shows sample elections with variance of ϕ distribution 0.5 and 0.25.

6.2 Simulation Results

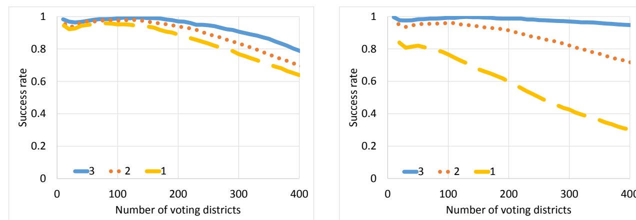
We expected weak candidates (that have been allocated only a single city where they lead the ground truth preference order) to struggle to win. Somewhat surprisingly,

[3]We use the distance from the city, so actually half of a normal distribution.



(a) (b)

Figure 4: Share of simulations in which a candidate who led the ground truth ranking in 3, 2, or 1 cities managed to win the election, depending on how many districts were possible. Ratio of largest to smallest district was capped by that of the US Senate; variance of ϕ distribution 0.5 (a) and 0.25 (b), so voters are more concentrated geographically.



(a) (b)

Figure 5: Share of simulations in which a candidate who led the ground truth ranking in 3, 2, or 1 cities managed to win the election, depending on how many districts were possible. Ratio between districts was not bounded; variance of ϕ distribution 0.5 (a) and 0.25 (b).

they did not do very badly—as can be seen in Figure 4a. Moreover, as the number of districts drops, the ability of less supported candidates to win actually grows, as they are able to divide other candidates' voters between their different districts, essentially overwhelming their vote (thanks to using plurality in each district).

We also examine what happens when we decrease the spread of opinions by having a smaller variance of opinions—we reduced the variance of potential values of ϕ, so most voters have significantly reduced noise compared to the ground truth. As can be seen in Figure 4b, this essentially strengthened the power of the cities, as even "rural" voters were closer to them in their preference order. This meant that it was much harder for weak candidates to succeed, and the gap between stronger and weaker candidates grows.

The dynamic shown here is not significantly different when we examine the outcome without limiting the relative size of districts (Figure 5). The rise in less-supported candidates success rate as number of districts decrease is much more rapid, since candidates are able to include their opponents in a few large districts in many more cases (this also explains the higher rate of success). Not limiting the difference between district sizes also seems to cause a "bump" when there are extremely few districts, in which it is sometimes harder for weaker candidates to push the powerful, 3-city, candidate to a few districts.

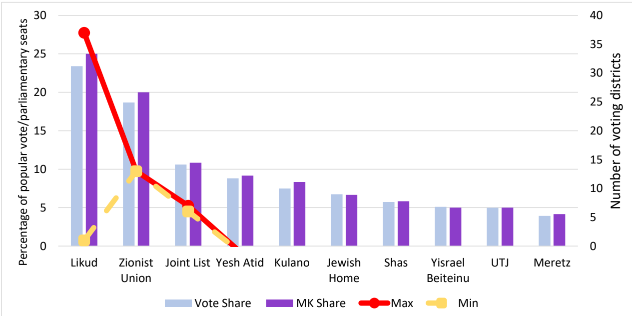


Figure 6: Maximal and minimal district number that can enable a party to become the plurality winner (line graphs) compared to share of popular vote and MK share (bar graphs). Ratio between the population of the most populous district and that of the least populous one was bounded.

7. GERRYMANDERING$_{PLURALITY}$ IN THE REAL WORLD

We now turn to real-sworld data. While it does not allow us to easily explore the influence of various variables on the results, we are able to see in recent election results (2015 UK and Israeli elections) how different districting could affect the outcome. Our results attest to the primacy of geographical dispersion as a key aspect, apart from voting share or parliamentary seat share.[4]

7.1 Israeli Results

The dataset of the 2015 Israeli legislative election[5] 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 6 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.

Figure 6 shows that with Israeli voters' distribution, the ability to win is, in general, monotonic with the percent of the vote. Interestingly, both *Zionist Union* and *Joint List* parties have a very narrow band of district number that allows them to win—they would need a rather tailored district structure to win (examining Figure 8 shows they are quite close to enlarging their possibility, so this may be an artifact of these particular results).

When we do not constrain relative district sizes, as can be seen in Figure 7, things are quite different. Unexpectedly,

[4]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 vote distribution at that particular data point.

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

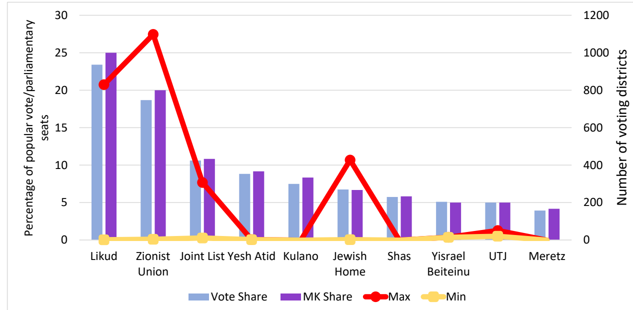


Figure 7: Results when not bounding the ratio between largest and smallest districts.
Maximal and minimal district number that can enable a party to become the plurality winner (line graphs) compared to share of popular vote and MK share (bar graphs).

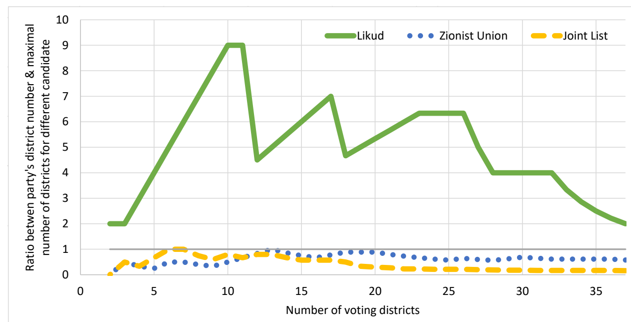


Figure 8: Ratio between the number of districts the party won and the maximal number of districts that any other party won; in every iteration of Algorithm 1. When the values are above 1, the greedy algorithm finds a partition where the party wins a plurality of districts.

winning is no longer as tied to voting block, but much more dependent on particular geographic distribution. The *Kulanu* party won 7.49% of the votes, and the 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, due to their particular geographic voter dispersal patterns, which are focused in ways that allow them to relegate other candidates' supporters into few districts.

7.2 United Kingdom Results

The dataset of the 2015 British general election[6] 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 9 shows for every party that won one of the 650 seats in the 56th Parliament of the United Kingdom, its

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

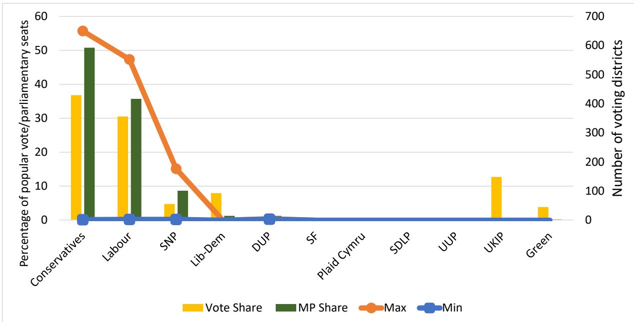


Figure 9: Maximal and minimal district number that can enable a party to become the plurality winner (line graphs) compared to share of popular vote and MP share (bar graphs)

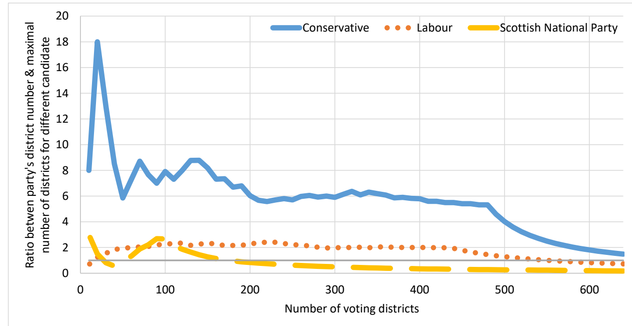


Figure 10: Ratio between the number of districts the party won and the maximal number of districts that any other party won; in every iteration of Algorithm 1.

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.

Figure 9 shows that beyond the 2 large parties, the other parties that were able to win by gerrymandering are geographically concentrated, despite parties with a larger share of the vote being unable to do so. In Scotland (SNP) and Northern Ireland (DUP) parties were able to find winning gerrymandering, though parties with larger support (Liberal-Democrats, UKIP, Greens), and even with more MPs (Liberal-Democrats) were not able to do so.

Figure 10 shows the SNP had multiple district structures that could make it victorious (unlike the DUP). Note that while the SNP received a far smaller share of the UK vote than the Zionist Union party did in Israel, thanks to its geographical spread, it is much more flexible in the number of districts in which it can become a winner.

8. 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 simulations as well as 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). This is in line with current efforts to limit the possibility of gerrymandering.

It is important to note that as we use compact and contiguous districts, several of the suggested "fixes" to politically biased gerrymandering do not apply here. Despite using this restricted type of gerrymandering we show in our simulations and empirical work how different parties can become winners with a carefully chosen district structure (and this is beyond the structural issue detailed in [1]).

Our simulations show that even relatively weak candidates can reach victory with well-structured gerrymandering, and together wth the empirical work they clarify how, beyond voting share and parliamentary weight, the significant issue influencing the possibility of manipulation is the geographical dispersion of the voters. We hope further research will do more to investigate the various variables that come into play when manipulation is geographically based, and what is an optimal gerrymandering strategy for various voter concentration patterns. Moreover, examining sharper urban-rural divides (as exists in the US) may be of interest, as it increases the possibility of gerrymandering.

The relative size of districts is an important issue which we took into account in our simulations and real-world analysis in a fairly simple way—enforcing a fixed cap on the ratio between their sizes (and we showed how different results look when we do not enforce such a cap). Obviously, some of the outcomes were much better than this cap—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. While it is obvious that smaller parties will struggle to gerrymander when a lower ratio is required, the more specific relationship between geographic concentration, district size ratio, and voter number, and how one can offset any one attribute with another, is an interesting topic to continue to explore in this area.

Acknowledgment

This research has been partly funded by Microsoft Research through its PhD Scholarship Program, the HUJI Cyber Security Research Center in conjunction with the Israel National Cyber Bureau in the Prime Minister's Office, by Israel Science Foundation grant #1227/12, and by NSERC grant 482671.

REFERENCES

[1] Y. Bachrach, O. Lev, Y. Lewenberg, and Y. Zick. Misrepresentation in district voting. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 81–87, New York City, New York, July 2016.

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

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

[4] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, March 2016.

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

[6] M. E. Dyer and A. M. Frieze. Planar 3DM is NP-complete. *Journal of Algorithms*, 7(2):174–184, 1986.

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

[8] G. Erdélyi, E. Hemaspaandra, and L. A. Hemaspaandra. More natural models of electoral control by partition. In *International Conference on Algorithmic DecisionTheory*, pages 396–413. Springer, 2015.

[9] R. S. Erikson. Malapportionment, gerrymandering, and party fortunes in congressional elections. *The American Political Science Review*, 66(4):1234–1245, 1972.

[10] P. Faliszewski, E. Hemaspaandra, and L. A. Hemaspaandra. Using complexity to protect elections. *Communications of the ACM*, 53(11):74–82, 2010.

[11] M. Feix, D. Lepelley, V. 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] 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.

[13] D. S. Felsenthal and M. Machover. Minimizing the mean majority deficit: The second square-root rule. *Mathematical Social Sciences*, 37(1):25–37, 1999.

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

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

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

[17] A. Gelman, J. N. Katz, and F. Tuerlinckx. The mathematics and statistics of voting power. *Statistical Science*, 17(4):420–435, 2002.

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

[19] 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.

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

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

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

[23] R. Lahrach and V. Merlin. Assessing the probability of the referendum paradox: the french local election case, 2010.

[24] 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.

[25] Z. Landau and F. E. Su. Fair division and redistricting. In *AMS Special Sessions on The Mathematics of Decisions, Elections, and Games*, pages 17–36, January 2013.

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

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

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

[29] N. Miller. The house size effect and the referendum paradox in U.S. presidential elections. *Electoral Studies*, 35:265–271, September 2014.

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

[31] A. D. Procaccia, J. S. Rosenschein, and A. Zohar. Multi-winner elections: Complexity of manipulation, control, and winner-determination. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 1476–1481, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.

[32] C. Puppe and A. 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.

[33] C. Puppe and A. Tasnádi. Optimal redistricting under geographical constraints: Why "pack and crack" does not work. *Economic Letters*, 105(1):88–73, October 2009.

[34] J. Rothe, editor. *Economics and Computation.* Springer, 2015.

[35] M. A. 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.

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

[37] A. Tangian. Computational application of the mathematical theory of democracy to arrow's impossibility theorem (how dictatorial are arrow's dictators?). *Social Choice and Welfare*, 35(1):129–161, 2010.

[38] A. Tasnádi. The political districting problem: A survey. *Society and Economy*, 33(3):543–554, 2011.

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

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

[41] R. Van Bevern, R. Bredereck, J. Chen, V. Froese, R. Niedermeier, and G. J. Woeginger. Network-based vertex dissolution. *SIAM Journal on Discrete Mathematics*, 29(2):888–914, 2015.

[42] L. Xia, M. Zuckerman, A. D. Procaccia, V. Conitzer, and J. S. Rosenschein. Complexity of unweighted coalitional manipulation under some common voting rules. In *Proceedings of the 21st International Joint Conference on Artifical Intelligence*, IJCAI'09, pages 348–353, San Francisco, CA, USA, 2009.

[43] M. Zuckerman, A. D. Procaccia, and J. S. Rosenschein. Algorithms for the coalitional manipulation problem. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, pages 277–286, Philadelphia, PA, USA, 2008.