Minimizing Margin of Victory for Fair Political and Educational Districting

Ana-Andreea Stoica
Columbia University

Abhijnan Chakraborty
Max Planck Institute for Software Systems

Palash Dey
Indian Institute of Technology Kharagpur

Krishna P. Gummadi
Max Planck Institute for Software Systems

ABSTRACT

In many practical scenarios, a population is divided into disjoint groups for better administration, such as electorates into political districts and students into school districts. However, grouping people arbitrarily may lead to biased partitions, raising concerns of gerrymandering in political districting and racial segregation in schools. To counter such issues, in this paper, we conceptualize such problems in a voting scenario, and given an initial grouping, we propose the Fair Regrouping problem to redistribute a given set of people into k groups, where each person has a preferred alternative and a set of groups they can be moved to, such that the *maximum margin of victory* of any group is minimized. We also propose the Fair Connected Regrouping problem which additionally requires the people within each group to be connected. We show that the Fair Regrouping problem is NP-complete for plurality voting even if we have only 3 alternatives, but admits polynomial time algorithms if everyone can be moved to any group. We further show that the Fair Connected Regrouping problem is NP-complete for plurality voting even if we have only 2 alternatives and $k = 2$. Finally, we propose heuristic algorithms for both problems and show their effectiveness in political districting in the U.K. and in lowering racial segregation in public schools in the U.S.

ACM Reference Format:
Ana-Andreea Stoica, Abhijnan Chakraborty, Palash Dey, and Krishna P. Gummadi. 2020. Minimizing Margin of Victory for Fair Political and Educational Districting. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 10 pages.

1 INTRODUCTION

Dividing a population into smaller groups is often a practical necessity for better administration. For example, in many democratic countries (most notably, in countries following the Westminster System like the U.K., Canada, India, Australia, or the Presidential System like U.S., Brazil, Mexico, Indonesia), electorates are divided into electoral districts; in many organizations, employees are divided into administrative units like departments; students enrolled in public schools in the U.S. are divided into school districts; and so on. However, the population is not homogeneous, as it consists of people with different attributes—gender, race, religion, or ideological leaning. Dividing people arbitrarily may lead to biased grouping, disallowing minorities and underrepresented groups from accessing the same services and opportunities as majority groups.

In electoral districting, given past voting patterns, ruling political parties may draw district boundaries that favor them—a practice termed as *gerrymandering* [38]. For example, they may want to ensure that they enjoy a healthy lead over the opponents in many districts, so that even if a handful of their supporters change sides, it does not hamper the winnability. Alternatively, they can assign the majority of the opposition supporters to others districts, making them minorities in those districts and thus weakening their power. There have been several instances of such manipulations in electoral (re)districting in the U.S., starting as early as in 1812, by then Massachusetts governor Elbridge Gerry (the term *gerrymandering* originated after him) [32]. Since then, efficient (re)distribution of voters into districts remains an open problem due to the complex dynamics involving voter mobility constraints, social and financial burdens, as well as difficulty in testing efficacy of proposed interventions [4, 17, 32].

Public schools in the U.S. are governed by school boards representing local communities and are largely funded by local property taxes [13, 19]. Most of the students go to a school in the district they live, with proximity playing an important role in the school choice [21]. The way in which the students are distributed to schools determines the racial composition of the schools, as well as the revenues they earn. Several reports claim that wealthier, whiter communities have pushed policies so that white families can live in white-majority areas and attend white-majority schools [15, 48]. Despite the desegregating efforts following the landmark Supreme Court verdict in *Brown v. Board of Education case* in 1954 (which ruled racial segregation of children in public schools to be unconstitutional), 63% of classmates of a white student are whites, compared to 48% of all students being whites; similarly, 40% of black and Hispanic students attend schools where over 90% students are people of color [28]. As an economic consequence, a recent report by an educational non-profit EdBuild claimed that *"non-white school districts get \$23 Billion less than white districts, despite serving the same number of students"* [22].

Aside from the offline world, bias in grouping users can exist in online settings as well. For instance, social media platforms like Facebook allow advertisers to target groups of users, raising concerns over the targeting of *political ads* and housing, credit or job *opportunity ads* in such platforms [2, 3, 30, 47, 51]. The option of specifying a narrow target group

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

can be maliciously exploited for political benefits, where the campaigns can tailor their political message solely to a single ideological group, potentially leading to more polarized and extreme messaging, and allowing misinformation to travel and be accepted more easily [23, 41, 50].

To counter such biases in dividing people into groups, we conceptualize these problems in a voting scenario: the goal is to divide n people, each having one preferred alternative (out of m), into k groups. While the mapping of electoral districting, or political ad targeting, into voting is direct and utilizes people's ideological preferences, we can think of context-specific mapping in other scenarios. For instance, in school assignment, we can think of students having preference according to their sensitive attributes (such as gender or race). Once the mapping is done, we utilize the concept of *margin of victory* (defined as the number of people who need to change their preference in order to change the winner) from computational social choice to redistribute the population in groups in a more equitable way, given certain constraints on the groups that a user can be moved to. We propose the Fair Regrouping problem to create k groups such that *the maximum margin of victory of any group is minimized*, and the Fair Connected Regrouping problem which additionally requires people in each group to remain connected in an underlying social graph.

Although by definition the margin of victory looks only at the top two contenders, it could aid in situations where there is a monopolistic advantage, such as in segregated political areas or neighborhoods. Reducing margin of victory conceptually leads to groups in which there is no dominating opinion and thus everyone's opinion is valued, since the consensus of the group can be changed even if a small number of people change their preferences. In political districting, it would lead to higher accountability from the elected candidate, incentivizing them to truly engage with their electorate due to the competition created. Moreover, it could also help avoiding certain gerrymandering practices by disallowing packing voters into districts where one contender wins by a high margin. This allows a minority group to have critical mass, thus indirectly giving a better sense of security and belongingness among people in that group. Similarly, it would lead to lower racial segregation in schools.

For targeting political ads, a lower margin of victory would deter political campaigns to resort to extreme political messaging, given different points of view in the target audience. Note that our proposal is aimed at the online platform (re)grouping individuals when allowing political targeting from external organizations. In the offline setting of local elections or school choice, people's geography and their social connections may constrain the groups they can be a part of, whereas in the online setting such constraints may come from users' expressed interests.

1.1 Contributions

We make the following contributions in this paper.

• We show that the Fair Regrouping problem is NP-complete even when we have only 3 alternatives and there is no constraint on the size of individual groups [Theorem 4.1]. We complement this intractability result by proving existence of polynomial time (more specifically XP) algorithm, when (i) every voter can be moved to any group (which we term as the Fair Regrouping_X problem) [Theorem 4.5], and (ii) we have a constant number of groups [Theorem 4.6].

• We further show that the Fair Connected Regrouping problem is NP-complete even when there are only 2 alternatives, 2 districts, the maximum degree of any vertex in the underlying graph is 5, and no constraint on the size of districts [Theorem 4.4]. This shows that, although both Fair Regrouping and Fair Connected Regrouping problems are NP-complete, Fair Connected Regrouping is computationally harder than Fair Regrouping.

• We propose heuristic algorithms for both Fair Regrouping and Fair Connected Regrouping problems, and show their effectiveness in reducing margin of victory in electoral districts in the U.K., as well as in lowering racial segregation in public schools in the U.S.

2 BACKGROUND & RELATED WORK

Voting mechanisms have been at the center of historical, political and sociological studies [5, 24, 39]. The problem of *unfair distribution of voters into districts*, i.e. *gerrymandering*, has received significant attention [4, 11, 32, 34], setting geographical [38] and social constraints [9, 17, 33] to population mobility. Puppe and Tasnádi [46], and Van Bevern et al. [52] proved the problem to be NP-complete. Central to this problem is the concept of *representation*: does a collective represent the choices or attributes of those comprising it? While recent papers conceptualize different measures of representation in district-based elections [4, 14, 26, 27, 29, 34], to our knowledge, we are the first to use the concept of margin of victory for redistricting voters to achieve better representation.

Computing margin of victory for different voting rules has been studied in [54]. Several works [8, 12, 20, 40] have attempted to infer it in real elections, and showed that even estimation becomes difficult in establishing robust elections. A closely-related problem is *bribery and robustness*, studied in [6, 10, 42, 43]. Yet, to our knowledge, the problem of minimizing margin of victory has not attracted much attention from both a theoretical and an application point of view.

Another related area is achieving *proportional representation* as a fairness goal in clustering, where every cluster contains the same proportion of users from certain demographics as in the general population [16, 37, 49, 55]. Our paper differs from traditional clustering as the objective here is not to maximize a similarity measure, but to use geographical or social similarity as a constraint in minimizing the margin of victory. Note that this is not equivalent to proportional representation, but represents a different metric aiming at better equity in elections. While desirable in certain cases, proportional representation is also NP-complete in general settings [45], and may dilute voter power across many electoral districts.

These works show that the problem of grouping is a complex one and that there is a gap between such measures of robustness, proportional representation, and other gerrymandering metrics. Minimizing margin of victory aims to close this gap in the current state of affairs that involve a deep inequality among existing groups. In our setting, minimizing the margin of victory serves the purpose of eliminating *monopolistic advantage*, creating a healthy competitive environment for the top contenders. While we acknowledge the possibility of a higher risk for manipulation due to small difference in votes, we hope that election audits and other prevention measures can deter such practices. Given classic gerrymandering concerns such as cracking and packing, we argue that a minimal margin of victory prevents such practices in political elections, while a proportional representation constraint may correlate with cracking a minority voter population into all districts, disallowing them from winning any district.

In case of school segregation, the current state of public schools in many cities, including our dataset from Detroit, is of extreme racial segregation, which can be abstractly viewed as a case of *monopoly*, which we aim to alleviate through minimizing the margin of victory. In an ideal world, all schools would be composed of demographics in a proportional and representative way, but we are yet far from achieving such an ideal. What we can do is enact policies that are effective in establishing more equitable access to education and opportunity while being aware of the current state of social inequality. Minimizing margin of victory aims to facilitate the enactment of such policies (which have historical precedent in the form of the desegregation busing decision of the Supreme Court in 1971) with awareness of such inequality through process fairness, considering existing social constraints.

3 PRELIMINARIES

Voting Setting: For a positive integer k, we denote the set $\{1, 2, \ldots, k\}$ by $[k]$. Let $\mathcal{A} = \{a_i : i \in [m]\}$ be a set of m *alternatives* and $\mathcal{V} = \{v_i : i \in [n]\}$ a set of n *voters*. Each voter has a most preferred alternative whom the voter votes for. The *plurality voting rule* chooses the set of winners as the set of alternatives who are the most preferred alternative by the maximum number of voters. The number of voters who prefer an alternative most is called the plurality score of that alternative.

Margin of Victory: The margin of victory is the minimum number of votes that needs to be changed to change the election outcome. With the exception of the case where the top two alternatives are tied (where the margin of victory is 1), it easily follows that the margin of victory of a plurality election is the ceiling of half the difference between the two highest plurality scores of the alternatives. We now define our basic problem of Fair Regrouping.

Definition 3.1 (Fair Regrouping). Given a set $\mathcal{A}$ of m alternatives, a set $\mathcal{V}$ of n voters along with their corresponding preferences, initial partition of k groups $\mathcal{H} = \{H_i, i \in [k]\}$ along with the set $\mathcal{V}_i$ of voters corresponding to each group H_i for $i \in [k]$ such that $(\mathcal{V}_i)_{i\in[k]}$ forms a partition of $\mathcal{V}$, a function $\pi : \mathcal{V} \longrightarrow 2^{\mathcal{H}} \setminus \{\emptyset\}$ denoting the set of groups that each voter can be part of, minimum size s_{min} and maximum size s_{max} of every group, and a target t of maximum margin of victory of any group, compute if there exists a partition $(\mathcal{V}'_i)_{i\in[k]}$ of $\mathcal{V}$ into these k groups such that

(i) For every $i \in [k]$ and $v \in \mathcal{V}'_i$, we have $H_i \in \pi(v)$
(ii) For every $i \in [k]$, we have $s_{min} \leqslant |\mathcal{V}'_i| \leqslant s_{max}$
(iii) The margin of victory in the group H_i is at most t for every $i \in [k]$

We denote an arbitrary instance of this problem by $(\mathcal{A}, \mathcal{V}, k, \mathcal{H} = (H_i)_{i\in[k]}, (\mathcal{V}_i)_{i\in[k]}, \pi, s_{min}, s_{max}, t)$.

An important special case of Fair Regrouping is when every voter can be moved to any group; i.e., $\pi(v) = \mathcal{H}$ for every voter $v \in \mathcal{V}$. We call this problem Fair Regrouping_X (where the subscript X denotes that there is no user specific mobility constraint). We denote an arbitrary instance of Fair Regrouping_X by $(\mathcal{A}, \mathcal{V}, k, \mathcal{H} = (H_i)_{i\in[k]}, (\mathcal{V}_i)_{i\in[k]}, s_{min}, s_{max}, t)$.

The Fair Regrouping problem is generalized to define the Fair Connected Regrouping problem, where the input also have a social graph defined on the set of voters, the given groups are all connected, and we require the new groups to be connected as well. We denote an arbitrary instance of Fair Connected Regrouping by $(\mathcal{A}, \mathcal{V}, \mathcal{G}, k, \mathcal{H} = (H_i)_{i\in[k]}, (\mathcal{V}_i)_{i\in[k]}, \pi, s_{min}, s_{max}, t)$. In this paper, we study the above problems only for the plurality voting rule and thus omit specifying it every time. The following observation is immediate from the problem definitions.

Observation 1. Fair Regrouping_X *many-to-one reduces to* Fair Regrouping *which again many-to-one reduces to* Fair Connected Regrouping*, both in polynomial-time.*

4 THEORETICAL RESULTS

In this section, we present our basic theoretical results. Our first result shows that Fair Regrouping is NP-complete even with 3 alternatives. For that we reduce from the well known SAT problem which is known to be NP-complete.

Theorem 4.1. *The* Fair Regrouping *problem is* NP*-complete even if we have only* 3 *alternatives and there is no constraint on the size of any group.*

Proof. Fair Regrouping clearly belongs to NP. To prove NP-hardness, we reduce from the SAT problem. Let $(\mathcal{X} = \{x_i : i \in [n]\}, \mathcal{C} = \{C_j : j \in [m]\})$ be an arbitrary instance of SAT. Let us consider the following instance $(\mathcal{A}, \mathcal{V}, k, \mathcal{H} = (\mathcal{V}_i)_{i\in[k]}, \pi, s_{min} = 0, s_{max} = \infty, t = 2)$ of Fair Regrouping.

$$\mathcal{A} = \{a, b, c\}, k = 3n + m'$$
$$\mathcal{H} = \{\mathcal{X}_i, \bar{\mathcal{X}}_i, \mathcal{Z}_i : i \in [n]\} \cup \{\mathcal{Y}_j : j \in [m']\}$$
$$\forall i \in [n], \text{Votes in } \mathcal{X}_i : m' + 2 \text{ votes for } a$$
$$m' \text{ votes for } b, m' - 1 \text{ votes for } c$$
$$\forall i \in [n], \text{Votes in } \bar{\mathcal{X}}_i : m' + 2 \text{ votes for } a$$
$$m' \text{ votes for } b, m' - 1 \text{ votes for } c$$
$$\forall i \in [n], \text{Votes in } \mathcal{Z}_i : m' + 2 \text{ votes for } a$$
$$m' + 1 \text{ votes for } c$$
$$\forall j \in [m'], \text{Votes in } \mathcal{Y}_j : m' + 3 \text{ votes for } a$$
$$m' \text{ votes for } b$$
$$t = 2$$

Let f be a function defined on the set of literals as $f(x_i) = \mathcal{X}_i$ and $f(\bar{x}_i) = \bar{\mathcal{X}}_i$ for every $i \in [n]$. We now describe the π function. For $i \in [n]$, no voter in $\mathcal{Z}_i$ can move to any other group except one voter who votes for the alternative c and she can move to $\mathcal{X}_i$ and $\bar{\mathcal{X}}_i$. For $i \in [n]$, no voter voting for the alternatives a or c in both $\mathcal{X}_i$ and $\bar{\mathcal{X}}_i$ leave their current groups; any number of voters in $\mathcal{X}_i$ ($\bar{\mathcal{X}}_i$ respectively) who vote for the alternative b can move to the group $\mathcal{Y}_j$ for some $j \in [m']$ if the variable x_i ($\bar{x_i}$ respectively) appears in the clause C_j. Finally no voter in the group $\mathcal{Y}_j, j \in [m']$ leave their current group. This finishes the description of π and the description of the instance of Fair Regrouping. We claim that the two instances are equivalent.

In one direction, let us assume that the SAT instance is a yes instance. Let $g : \mathcal{X} \longrightarrow \{0, 1\}$ be a satisfying assignment for the SAT instance. Let us consider the following movement of the voters: for $i \in [n]$, if $g(x_i) = 1$, then one voter in the group $\mathcal{Z}_i$ who votes for the alternative c moves to the group $\mathcal{X}_i$; otherwise one voter in the group $\mathcal{Z}_i$ who votes for the alternative c moves to the group $\bar{\mathcal{X}}_i$. For $j \in [m']$, let the clause C_j be $\ell_1 \vee \ell_2 \vee \ell_3$ and g sets the literal ℓ_1 to be 1 (we can assume this without loss of generality). Then one voter from the group $f(\ell_1)$ who votes for b moves to the group $\mathcal{Y}_j$. Since the assignment g satisfies all the clauses, the margin of victory in the group $\mathcal{Y}_j$ is 2 for every $j \in [m']$. For $i \in [n]$, if $g(x_i) = 0$ ($g(x_i) = 1$ respectively), then the margin of victory in the group $\bar{\mathcal{X}}_i$ ($\mathcal{X}_i$ respectively) is 2 since it receives a voter voting for the alternative c. The rest of the groups (for $i \in [n], \mathcal{X}_i$ if $g(x_i) = 0$ and $\bar{\mathcal{X}}_i$ if $g(x_1) = 1$) remain same and their margin of victory remains to be 2. Hence the Fair Regrouping instance is also a yes instance.

In the other direction, let's assume that the Fair Regrouping instance is a yes instance. We define an assignment $g : \mathcal{X} \longrightarrow \{0, 1\}$ to the variables in the SAT instance as follows. For $i \in [n]$, if a voter in the group $\mathcal{Z}_i$ who votes for c moves to $\mathcal{X}_i$, then we define $g(x_i) = 1$; otherwise we define $g(x_i) = 0$. We claim that g is a satisfying assignment for the SAT instance. Suppose not, then there exists a clause $C_j = \ell_1 \vee \ell_2 \vee \ell_3$ for some $j \in [m']$ which g does not satisfy. To make the margin of victory of the group $\mathcal{Y}_j$ at most 2, one voter who votes for b must move into $\mathcal{Y}_j$ either from group $f(\ell_1)$ or from $f(\ell_2)$ or from $f(\ell_3)$. However, since g does not set any of ℓ_1, ℓ_2, or ℓ_3 to 1, none of these groups receive any voter who votes for the alternative c. Consequently, none of the groups can send a voter who votes for the alternative b to the group $\mathcal{Y}_j$ since otherwise the margin of victory of group which sends a voter who votes for the alternative b becomes at least 3 contradicting our assumption that the Fair Regrouping_X instance is a yes instance. Hence, the SAT instance is a yes instance. □

Due to Observation 1, it follows immediately from Theorem 4.1 that Fair Connected Regrouping problem for plurality voting rule is also NP-complete. We next show that Fair Connected Regrouping is NP-complete even if we simultaneously have 2 alternatives and 2 groups. For that, we reduce from 2-Disjoint Connected Partitioning, defined as:

Definition 4.2 (2-Disjoint Connected Partitioning). Given a connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and two disjoint nonempty sets $\mathcal{Z}_1, \mathcal{Z}_2 \subset \mathcal{V}$, compute if there exists a partition $(\mathcal{V}_1, \mathcal{V}_2)$ of $\mathcal{V}$ such that $\mathcal{Z}_1 \subseteq \mathcal{V}_1, \mathcal{Z}_2 \subseteq \mathcal{V}_2, \mathcal{G}[\mathcal{V}_1]$ and $\mathcal{G}[\mathcal{V}_2]$ are both connected. We denote an arbitrary instance of 2-Disjoint Connected Partitioning by $(\mathcal{G}, \mathcal{Z}_1, \mathcal{Z}_2)$.

It is already known that the 2-Disjoint Connected Partitioning problem is NP-complete [53, Theorem 1]. However the proof of Theorem 1 in [53] can be imitated as a reduction from the version of SAT where every literal appears in exactly two clauses; this restricted version of SAT is also known to be NP-complete [7]. This proves the following.

Proposition 4.3. *The* 2-Disjoint Connected Partitioning *problem is* NP*-complete even if the maximum degree of the input graph is* 5.

Theorem 4.4. *The* Fair Connected Regrouping *problem is* NP*-complete even if we have only* 2 *alternatives,* 2 *groups, the maximum degree of any vertex in the underlying graph is* 5*, and we do not have any constraint on the size of groups.*

Proof. The Fair Connected Regrouping problem is clearly in NP. To prove NP-hardness, we reduce from 2-Disjoint Connected Partitioning to Fair Connected Regrouping. Let $(\mathcal{G}' = (\mathcal{U}, \mathcal{E}'), \mathcal{Z}_1, \mathcal{Z}_2)$ be an arbitrary instance of Fair Connected Regrouping. Without loss of generality, let's assume that the degree of every vertex in $\mathcal{Z}_2$ is 2; z_2 be any arbitrary (fixed) vertex of $\mathcal{Z}_2$. Let's consider the following instance $(\mathcal{A}, \mathcal{V}, \mathcal{G} = (\mathcal{V}, \mathcal{E}), k = 2, \mathcal{H} = (H_i)_{i \in [2]}, (\mathcal{V}_i)_{i \in [2]}, \pi, s_{min} = 0, s_{max} = \infty, t = 1)$ of Fair Connected Regrouping.

$$\begin{aligned}
&\mathcal{A} = \{x, y\}\\
&\mathcal{V} = \{v_z : z \in \mathcal{Z}_2\}\\
&\quad \cup \{v_u, w_u : u \in \mathcal{V} \setminus \mathcal{Z}_2\}\\
&\quad \cup \mathcal{D}, \mathcal{D} = \{d_i : i \in [|\mathcal{Z}_2| + 1]\}\\
&\mathcal{E} = \{\{v_a, v_b\} : \{a, b\} \in \mathcal{E}'\}\\
&\quad \cup \{\{v_u, w_u\} : u \in \mathcal{V}[\mathcal{G}'] \setminus \mathcal{Z}_2\}\\
&\quad \cup \{\{d_i, d_j\} : i, j \in [|\mathcal{Z}_2| + 1], j = i + 1\}\\
&\quad \cup \{\{z_2, d_1\}\}\\
&\mathcal{H}_2 = \{d_i : i \in [|\mathcal{Z}_2| + 1]\}\\
&\mathcal{H}_1 = \mathcal{V} \setminus \mathcal{H}_2\\
&\quad \text{Vote of } v_u, u \in \mathcal{V} : x \succ y\\
&\quad \text{Vote of } w_u, u \in \mathcal{V} \setminus \mathcal{Z}_2 : y \succ x\\
&\quad \text{Vote of } d_i, i \in [|\mathcal{Z}_2| + 1] : y \succ x\\
&\quad \pi(v_z) = \{\mathcal{H}_1\}, z \in \mathcal{Z}_1\\
&\quad \pi(d_i) = \{\mathcal{H}_2\}, i \in [|\mathcal{Z}_2| + 1]\\
&\quad \pi(v) = \{\mathcal{H}_1, \mathcal{H}_2\} \text{ for every other vertex } v
\end{aligned}$$

This finishes the description of the instance of Fair Connected Regrouping. We now claim that the Fair Connected Regrouping instance is equivalent to the 2-Disjoint Connected Partitioning instance.

In one direction, let us assume that the 2-Disjoint Connected Partitioning instance is a yes instance. Let $(\mathcal{V}_1, \mathcal{V}_2)$ be a partition of $\mathcal{U}$ such that $\mathcal{Z}_1 \subseteq \mathcal{V}_1, \mathcal{Z}_2 \subseteq \mathcal{V}_2, \mathcal{G}'[\mathcal{V}_1]$ and

$\mathcal{G}'[\mathcal{V}_2]$ are both connected. We consider the following new partition of the voters:

$$\text{Voters of } \mathcal{H}_1 : \{v_u, w_u : u \in \mathcal{V}_1\}; \text{ voters of } \mathcal{H}_2 : \text{ others}$$

Since $\mathcal{G}'[\mathcal{V}_1]$ is connected, it follows that $\mathcal{G}[\mathcal{H}_1]$ is also connected. Similarly, since $\mathcal{G}'[\mathcal{V}_2]$ is connected, $\mathcal{G}[\mathcal{D}]$ is connected, and $\{z_2, d_1\} \in \mathcal{E}[\mathcal{G}]$, it follows that $\mathcal{G}[\mathcal{H}_2]$ is also connected. In $\mathcal{H}_1$, both the alternatives x and y receive the same number of votes and thus the margin of victory of $\mathcal{H}_1$ is 1. In $\mathcal{H}_2$, the alternatives x receives 1 less vote than the alternatives y and thus the margin of victory of $\mathcal{H}_2$ is 1. Thus the Fair Connected Regrouping instance is also a yes instance.

In the other direction, let's assume that there exists a valid partition $(\mathcal{H}_1', \mathcal{H}_2')$ of the voters such that both $\mathcal{G}[\mathcal{H}_1']$ and $\mathcal{G}[\mathcal{H}_2']$ are connected and the margin of victory of both $\mathcal{H}_1'$ and $\mathcal{H}_2'$ are 1. Let us define $\mathcal{V}_1 = \{u \in \mathcal{V}[\mathcal{G}'] : v_u \in \mathcal{H}_1'\}$ and $\mathcal{V}_2 = \mathcal{V}[\mathcal{G}'] \setminus \mathcal{V}_1$. It follows from the function π that we have $\mathcal{Z}_1 \subseteq \mathcal{V}_1', \mathcal{Z}_2 \subseteq \mathcal{V}_2'$. Also $\mathcal{G}'[\mathcal{V}_1']$ is connected since the voters in $\mathcal{H}_1'$ are connected. We also have $\mathcal{G}'[\mathcal{V}_2']$ is connected since the voters in $\mathcal{H}_2'$ are connected, the vertices in $\mathcal{D}$ forms a path, and there exists a pendant vertex in $\mathcal{D}$. We also have $\mathcal{Z}_2 \in \mathcal{V}_2''$ since the voters in $\{v_u :\in \mathcal{Z}_2\}$ belongs to $\mathcal{H}_2$; otherwise the margin of victory of $\mathcal{H}_2$ would be strictly more than 1. Hence $(\mathcal{V}_1', \mathcal{V}_2')$ is a solution of the 2-Disjoint Connected Partitioning instance and thus the instance is a yes instance. □

We now complement our hardness results with polynomial time algorithm for a particular case for the Fair Regrouping_X problem:

Theorem 4.5. *The* Fair Regrouping_X *problem is polynomial time solvable if the number of alternatives is a constant.*

Proof. Let an arbitrary instance of Fair Regrouping_X be $(\mathcal{A}, \mathcal{V}, k, \mathcal{H} = (H_i)_{i\in[k]}, (\mathcal{V}_i)_{i\in[k]}, s_{min}, s_{max}, t)$. For an alternative $a \in \mathcal{A}$, let n_a be the number of votes that a receives. We present a dynamic programming based algorithm for the Fair Regrouping_X problem. The dynamic programming table $\mathcal{T}\left((i_a \in \{0, 1, \ldots, n_a\})_{a\in\mathcal{A}}, \ell \in [k]\right)$ is defined as follows – $\mathcal{T}\left((i_a)_{a\in\mathcal{A}}, \ell\right)$ is the minimum integer λ such that the voting profile consisting i_a number of voters voting for the alternative a can be partitioned into ℓ groups such that the margin of victory of any group is at most λ. For every $i_a \in \{0, 1, \ldots, n_a\}, a \in \mathcal{A}$, we initialize $\mathcal{T}\left((i_a)_{a\in\mathcal{A}}, 1\right)$ to be the margin of victory of the voting profile which consists of i_a number of voters voting for the alternative a for $a \in \mathcal{A}$. We update the entries in the table $\mathcal{T}$ as follows for every $\ell \in \{2, 3, \ldots, k\}$.

$$\mathcal{T}\left((i_a)_{a\in\mathcal{A}}, \ell\right) = \min_{\substack{(i_a')_{a\in\mathcal{A}}, i_a' \geqslant 0\ \forall a\in\mathcal{A} \\ s_{min} \leqslant \sum_{a\in\mathcal{A}} i_a' \leqslant s_{max}}} \max \left\{ \begin{array}{l} mv\left(\left(i_a'\right)_{a\in\mathcal{A}}\right), \\ \mathcal{T}\left(\left(i_a - i_a'\right)_{a\in\mathcal{A}}, \ell - 1\right) \end{array} \right\}$$

In the above expression $mv\left(\left(i_a'\right)_{a\in\mathcal{A}}\right)$ denotes the plurality margin of victory of the profile which consists of i_a' number of voters voting for the alternative a for $a \in \mathcal{A}$. Updating each entry of the table takes $O\left(\prod_{a\in\mathcal{A}} n_a\right) \text{poly}(m, n)$ time. The table has $k \prod_{a\in\mathcal{A}} n_a$ entries. Hence the running time of our algorithm is $O\left(\prod_{a\in\mathcal{A}} n_a^2\right) \text{poly}(m, n) = O\left(n^{2m} \text{poly}(m, n)\right)$ which is $n^{O(1)}$ when we have $m = O(1)$. □

We next present a polynomial time algorithm for Fair Regrouping if we have a constant number of groups.

Theorem 4.6. *The* Fair Regrouping *problem is polynomial time solvable if the number of groups is a constant.*

Proof. An arbitrary instance of Fair Regrouping be $(\mathcal{A}, \mathcal{V}, k, \mathcal{H} = (H_i)_{i\in[k]}, (\mathcal{V}_i)_{i\in[k]}, \pi, s_{min}, s_{max}, t)$. We guess a winner and a runner up of every group – there are $\binom{m}{2}^k = O(m^{2k})$ possibilities. We also guess the plurality score of a winner of every group – there are $O(n^k)$ possibilities. Given a guess of a winner, its plurality score, and a runner up alternative of every group, we reduce the problem of computing if there exists a partition of $\mathcal{V}$ (respecting the given guesses) which achieves the maximum margin of victory of at most t to a s' to t' flow problem (with demand on edges) instance $\left(\mathcal{G} = (\mathcal{U}, \mathcal{E}), c : \mathcal{E} \longrightarrow \mathbb{R}^+, d : \mathcal{E} \longrightarrow \mathbb{R}^+\right)$ as follows.

$$\begin{aligned}
\mathcal{U} &= \mathcal{U}_L \cup \mathcal{U}_M \cup \mathcal{U}_R \cup \{s', t'\} \text{ where} \\
\mathcal{U}_L &= \{u_v : v \in \mathcal{V}\} \\
\mathcal{U}_M &= \{u_{a,i} : a \in \mathcal{A}, i \in [k]\} \\
\mathcal{U}_R &= \{u_i : i \in [k]\} \\
\mathcal{E} &= \left\{(s', u_v) : v \in \mathcal{V}\right\} \\
&\cup \left\{(u_v, u_{a,i}) : v \in \mathcal{V}, i \in [k],\right. \\
&\quad \left. v\text{'s vote is } a \succ \cdots, \mathcal{H}_i \in \pi(v)\right\} \\
&\cup \left\{(u_{a,i}, u_i) : a \in \mathcal{A}, i \in [k]\right\} \\
&\cup \left\{(u_i, t') : i \in [k]\right\}
\end{aligned}$$

The capacity c of every edge from s' to $\mathcal{U}_L$ and from $\mathcal{U}_L$ to $\mathcal{U}_M$ is 1. For every $i \in [k]$, if x and y are respectively the (guessed) winner and runner up of $\mathcal{H}_i$ and n_i is the (guessed) plurality score of a winner in $\mathcal{H}_i$, then we define the capacity and demand of the edge $(u_{x,i}, u_i)$ to be n_i and the capacity and demand of the edge $(u_{y,i}, u_i)$ to be $(n_i - t)$; if $(n_i - t)$ is not positive, then we discard the current guess. We define the capacity of the edge $(u_{z,i}, u_i)$ to be n_i for every alternative z who is not the guessed winner in $\mathcal{H}_i$ for $i \in [k]$. Finally we define the capacity and demand of every edge from U_R to t' to be s_{max} and s_{min} respectively. We claim that the given Fair Regrouping instance is a yes instance if and only if there exists a guess of a winner, its plurality score, and a runner up alternative of every group whose corresponding flow instance has an s' to t' flow of value n.

In one direction, suppose the Fair Regrouping instance is a yes instance. Let x_i and y_i be a winner and a runner up respectively in $\mathcal{H}_i$ and n_i be the plurality score of a winner in $\mathcal{H}_i$ for $i \in [k]$. For the guess corresponding to the solution of Fair Regrouping, we send 1 unit of flow from s' to $u_v, v \in \mathcal{V}$, from u_v to $u_{a,i}$ if the voter v belongs to $\mathcal{H}_i$ in the solution and v votes for a. Since every vertex in $\mathcal{U}_M$ has exactly one outgoing neighbor, all the incoming flow at every vertex in $\mathcal{U}_M$ move to their corresponding neighbor in $\mathcal{U}_R$. Similarly, the outgoing neighbor of every vertex in $\mathcal{U}_R$ is t', all the incoming flow at every vertex in $\mathcal{U}_R$ move to t'. Obviously the flow conservation property is satisfied at every vertex. Also capacity and demand constraints are also satisfied at every edge since the guess corresponds to a solution of the Fair Regrouping instance. Finally since the total outgoing flow at s' is n, the total flow value is also n.

In the other direction, assuming x_i and y_i being a guessed winner and a runner up respectively in $\mathcal{H}_i$ and n_i being the plurality score of a winner in $\mathcal{H}_i$ for $i \in [k]$, the corresponding flow network has a flow value of n, we claim that the Fair Regrouping instance is a yes instance. We can assume without loss of generality that the flow value on every edge in a maximum flow is an integer since the demand and capacity of every edge are integers. We define a voter $v \in \mathcal{V}$ to be in the group $\mathcal{H}_i, i \in [k]$ if there exists an alternative a such that there is one unit of flow in the edge $(u_v, u_{a,i})$. It follows from the construction of the maximum flow instance that the above partitioning the voters into the groups $\mathcal{H}_i$ is valid (that is, it respects π, s_{min}, and s_{max}) and the maximum margin of victory of any group is at most t. Hence the Fair Regrouping instance is also a yes instance. □

We observe that the polynomial time algorithms in Theorem 4.5 and 4.6 show that this problem belongs to the complexity class known as XP parameterized by the number of alternatives and by the number of groups, respectively. On the other hand, the existence of tractable cases for the Fair Connected Regrouping problem remains an open problem, together with the general set-up of the Fair Regrouping_X problem, when the number of alternatives is unbounded. In future, we plan to study the existence of polynomial-time algorithms for these settings where the order of the polynomial is independent of the number of alternatives and the number of groups, respectively. In the next section, we develop fast heuristics for our problems and exhibit their effectiveness in real-world and synthetic data.

5 EXPERIMENTAL EVALUATION

Given the high complexity of the Fair Regrouping problems, in this section, we propose faster greedy heuristics to minimize the margin of victory by moving people between an initial partition into groups, while respecting mobility and connectedness constraints.

5.1 Greedy Algorithms

We develop three variants of a greedy heuristic, each including different constraints related to one of the three defined problems. Each heuristic starts from an initial grouping of people, meant to mimic either a natural tendency of people with common interests/geography to group, or an already existing administrative division.

The algorithms then greedily 'move' people from the group with maximum margin of victory ($\mathcal{V}_{max}$) to others iteratively: it loops through all other district and checks if it can move any number of people from those districts to $\mathcal{V}_{max}$ (or vice-versa) in order to decrease the maximum margin of victory, as Algorithm 1 illustrates in a general greedy framework. Note that $\mathcal{V}$ is the set of voters, $\mathcal{A}$ the set of alternatives, $\pi : \mathcal{V} \longrightarrow 2^{\mathcal{H}} \setminus \{\emptyset\}$ denotes the set of groups that each voter can be part of, s_{min} and s_{max} denote minimum and maximum size, respectively, of every group, $mv(\cdot)$ denotes the margin of victory of a group, and $\mathcal{P}(voters_{\mathcal{V}_i}(A))$ the power set of the voters in group $\mathcal{V}_i$ whose top preference is A. The choice of π models the different constraints we enforce:

Result: A partition of $\mathcal{V}$ into groups $(\mathcal{V}_i)_{i\in[k]}$ with minimal maximum margin of victory.
Input $(\mathcal{V}_i)_{i\in[k]}$, $\pi : \mathcal{V} \longrightarrow 2^{\mathcal{H}} \setminus \{\emptyset\}$, s_{min}, s_{max};
Choose $\mathcal{V}_{max} = \arg\max_{i\in[k]} mv(\mathcal{V}_i)$.
for $v \in \mathcal{P}(voters_{\mathcal{V}_{max}}(A))$ **do**
  if $\mathcal{V}_i \in \pi(v)$ **then**
    if $|\mathcal{V}_{max} \backslash v| > s_{min}$ *and* $|\mathcal{V}_i \cup v| < s_{max}$ **then**
      if $\max(mv(\mathcal{V}_{max} \backslash v), mv(\mathcal{V}_i \cup v)) < \max(mv(\mathcal{V}_i), mv(\mathcal{V}_{max}))$ **then**
        move v from $\mathcal{V}_{max}$ to $\mathcal{V}_i$;
      end
    end
  end
end
for $\mathcal{V}_i \in \mathcal{V}$ *and* $A \in \mathcal{A}$ **do**
  for $v \in \mathcal{P}(voters_{\mathcal{V}_i}(A))$ **do**
    if $\mathcal{V}_{max} \in \pi(v)$ **then**
      if $|\mathcal{V}_i \backslash v| > s_{min}$ *and* $|\mathcal{V}_{max} \cup v| < s_{max}$ **then**
        if $\max(mv(\mathcal{V}_i \backslash v), mv(\mathcal{V}_{max} \cup v)) < \max(mv(\mathcal{V}_i), mv(\mathcal{V}_{max}))$ **then**
          move v from $\mathcal{V}_i$ to $\mathcal{V}_{max}$;
        end
      end
    end
  end
end

Algorithm 1: Greedy Algorithm

- ▷ In Greedy Regrouping_X, π is unconstrained, as people can move to any other group.
- ▷ In Greedy Regrouping, π models e.g. geographical constraints, allowing people to move to closest groups.
- ▷ In Greedy Connected Regrouping, π models the connections between people, allowing them to move such that no group becomes a disconnected subgraph.[1]

Greedy Regrouping_X is most-suited for an online setting given the lack of physical constraints. However, we are not advocating for moving people from one online community to another, but rather offer a framework that platforms can take into account for creating audience groups with low margin of victory for advertisers to target, as motivated in the beginning for settings such as targeting of political or opportunity ads. In the offline world, Greedy Regrouping selects people to move based on their mobility constraints (i.e., checking the list of groups each person can be moved). Finally, Greedy Connected Regrouping ensures both groups to remain connected

[1]Note that in this case the function π can be updated as the algorithm is running.

in the underlying social graph when moving people from one to another. Since the maximum margin of victory is a positive number, all algorithms terminate, with a termination condition defined through no movement between groups for ten consecutive rounds (experimenting with different termination conditions yields similar results).

5.2 Datasets & Experimental Results

To evaluate the applicability of greedy algorithms in real-world scenarios, we consider three main datasets: a synthetic dataset using graph models and two real ones, consisting of data from the U.K. parliament elections in 2017 and demographic information of students in public schools of Detroit, U.S. The synthetic dataset can be thought of emulating both offline and online scenarios, for which we evaluate all three greedy algorithms. In the political and school datasets, we evaluate Greedy Regrouping_X and Greedy Regrouping, but *not* Greedy Connected Regrouping, as we lack the social graph.

5.2.1 Synthetic Data.
We used the line model to simulate *voters, alternatives*, and *voters' political affiliations* [17]. For every node, we generated the preference over alternatives according to the distance between the voter and the alternatives. In addition to this, we simulated a set of graphs based on the Erdős-Renyi (ER) graph model [25]. We then created 50 instances of the ER graph model, where each node represents a voter and the edges are formed according to the model with an added homophily factor based on the distance between nodes (as simulated by the line model). Inputs to such graphs are the number of voters n (100), number of alternatives m (5), number of groups k (5), and homophily parameter α. We split the created graphs in equally-sized groups assigning people to groups in the order given by the line model, representing the initial partition from their political affiliation indices (in sorted order, the first $\lfloor n/k \rfloor$ are in the first district, and so on). The greedy algorithm requires an initial partitioning of voters into districts that it then attempts to improve. Since we start with quite homophilic districts, the greedy algorithm improves the sum of margins of victory in each graph instance, bringing it quite close to the baseline value.

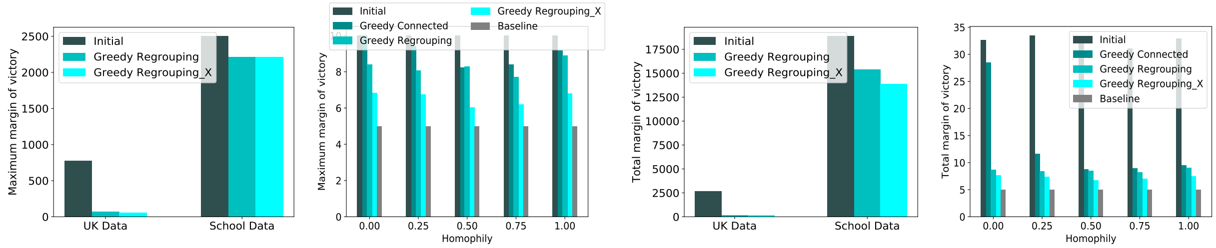
Such models capture the network and clustering effects exhibited by voter districts in real world [1, 18, 35], and can serve as a simulation of online networks as well. We further add a **baseline algorithm** that computes the optimal partition of people into groups with minimal margin of victory through a brute-force approach, given a network, the groups' size constraints, and mobility constraints. Thus, given a population with their preferences, it takes all possible groupings into consideration. This makes it computationally infeasible to scale at the size of the real data, but we use it in synthetic scenarios for comparison with the greedy heuristics.

We simulated the greedy algorithms for each graph instance, averaging over 10 iterations (yielding similar results as for a higher number of iterations) the minimal maximum margin of victory that it can reach and compared that to the baseline value. Whether we aim to solve this problem in the online world or the offline one, all these algorithms are effective in improving the maximum margin of victory aggregated for all graph instances (Figure 1(b)) and the total margin of victory (Figure 1(d)). For minimizing the total margin of victory, we adapt Algorithm 1 to optimize over the sum of margin of victory of all groups rather than the maximum, iterating over all pairs of districts. For the graph creation process, we vary the homophily factor between 0 and 1 (from totally non-homophilic to fully homophilic. We allow groups to change up to 20% in size, for a mobility constraint that allows people to move to the closest 2 groups, noting that each of the 5 groups starts with approximately 20 people (we experiment with different values and report this one, as for $k = 5$ groups allowing people to move to the closest 2 represents an average case). We observe that no matter how homophilic the initial graph is, the greedy heuristic is able to successfully reduce the maximum margin of victory for all three algorithms: Greedy Regrouping_X performs the best as it contains no constraints on mobility, being evaluated close to the baseline value and reducing maximum margin of victory by 35% on average (from 10 to 6–7), Greedy Regrouping performs second-best, reducing it by 20% on average (from 10 to 8), while Greedy Connected Regrouping reduces it by 10% on average (from 10 to 9), performing worse than the other two due to a tighter connectivity constraint. Figure 1(d) shows the overall decrease in margin of victory, where the effect is more significant: Greedy Regrouping_X and Greedy Regrouping achieve a result close to the baseline, reducing the total margin of victory by 75–80% (from 31–32 to 7–8), while Greedy Connected Regrouping performs slightly worse, reducing the total margin of victory of 46% on average (from 31–32 to an average of 13.6).

5.2.2 UK General Elections Data.
We collected data about the U.K. Parliament elections in 2017 from The Electoral Commission (`electoralcommission.org.uk`), using constituencies as *groups* and parties as *alternatives*. Although the votes are cast for individuals, the number of seats for each party is the number that counts in the Parliament, and thus we are interested in the effect of grouping on the distribution of votes over parties rather than over individuals. Knowing the number of votes each party got in each constituency, we simulated the top preference of the voters given a plurality voting mechanism.

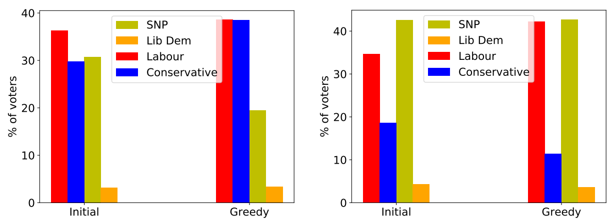
We tested our heuristics on 10 neighboring constituencies out of the 650 in the region of Scotland bordering Edinburgh, which represents a very diverse area in terms of voting preferences. Indeed, each of these constituencies had a clear majority (Figure 2).[2] We subsampled this dataset, working with a randomized sample of approximately 50,000 people and we recorded the center location of each constituency. While we experimented with different mobility constraints, results are qualitatively similar and thus we report an average case, enforcing in Greedy Regrouping that *voters can be incentivized to move or to vote only in their closest two constituencies.*

[2]The 10 constituencies we sampled are: Dumfriesshire, Clydesdale and Tweeddale, Berwickshire, Roxburgh and Selkirk, East Lothian, Midlothian, Edinburgh South, Edinburgh East, Edinburgh North and Leith, Edinburgh South West, Edinburgh West, and Livingston, for which an interactive map with the vote distribution can be found at https://www.bbc.com/news/election-2017-40176349.



(a) Real Data **(b) Synthetic Data** **(c) Real Data** **(d) Synthetic Data**

Figure 1: Maximum margin of victory for all algorithms in (a) real data and (b) synthetic data. Total (sum) of margin of victory in (c) real data and (d) synthetic data.



(a) East Lothian **(b) Edinburgh East**

Figure 2: Voters' distribution in UK constituencies before and after applying GREEDY REGROUPING.

Figure 1(a) shows that both GREEDY REGROUPING and GREEDY REGROUPING_X are able to reduce the maximum margin of victory of this dataset by approximately 91–92%, from an initial 776 to 67 and 55, respectively. Figure 1(c) shows the effect greedy had on minimizing the total margin of victory, showing an even larger decrease by almost 95%, from an initial 2,652 to 148 and 135, respectively. Note that the total of 50,000 sampled people are distributed approximately equally among the 10 constituencies. Since GREEDY REGROUPING represents the more realistic application given its embedded mobility constraints, we show in Figure 2 its effect on the voters' distribution in East Lothian and Edinburgh East, showing that it created a stronger opposition for the leading parties (Labour in East Lothian and SNP in Edinburgh East).

5.2.3 *US Public School Data.*

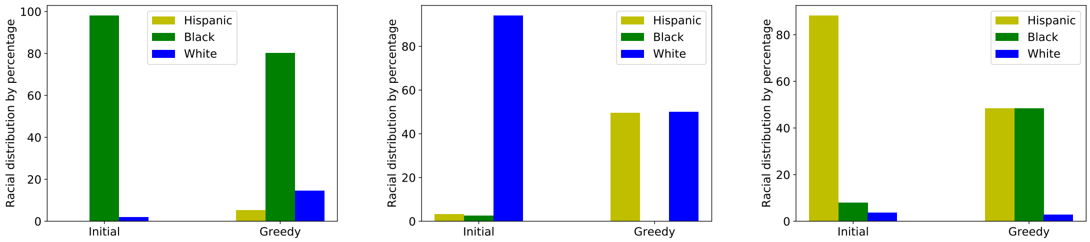
Neighborhood racial segregation is still widespread in many places in the US, trickling down to segregation in schools [28, 48]. Here, we attempt to show that our algorithms can be used to increase racial diversity in schools, if accompanied by government policies that facilitate movement of students between schools [44].

We collected school data from the National Center for Education Statistics (NCES: `nces.ed.gov/ccd`) about public schools in Detroit, MI, which is still one of the cities with highest rate of segregation, and most economic and social struggles encountered by minorities [31, 36]. We gathered data from 61 schools in Detroit, each containing between 40 and 5000 students, summing up to 41,834 students and their reported race. We modeled this data in the form of an election, where the *voters* are the students, the *alternatives* are their reported race (NCES data has 7 reported races: Asian, Native American, Hispanic, Black, White, Hawaiian, and Mixed-race), and the *groups* are the schools. Given each student's race, we modeled this 'election' as a plurality voting scenario, where each student only 'votes' for their reported race. Furthermore, we recorded the location of each school, enforcing in GREEDY REGROUPING that *students can only go to their closest five schools.*

Both algorithms decrease the maximum margin of victory by 11–12% on average, from an initial 2,501 to 2,213 and 2,311, respectively (Figure 1(a)), showing a significant decrease in the overall margin of victory by 18–24%, from an initial 18,870 to 15,360 and 14,376, respectively (Figure 1(c)). As GREEDY REGROUPING represents the more realistic scenario, we observe that schools containing students from one predominant racial group become more equilibrated (Figure 3), from the initial racial distribution in these schools that has a clear majority of a certain race (Black for Dove Academic, White for Universal Academy, and Hispanic for Cesar Chavez Academy).

Of course, since minimizing margin of victory only considers the most predominant two races, we may need to enforce an additional diversity constraint to preserve a minimum fraction of students from other races in a school (e.g., the 2.5% Black students in Universal Academy may need to stay), which we leave for future work. What we do notice, however, is that the demographic variance (which shows how the student demographics of individual schools are spread out from the underlying population distribution, with smaller variance denoting more equal demographic spread) decreases after applying GREEDY REGROUPING algorithm, by approximately 50% for all three schools, from 54,400 to 26,324 for Universal Academy, from 27,355 to 11,266 for Cesar Chavez, and from 593,367 to 313,439 for Dove Academy. Other measures may also be used to compute racial disparity effect of GREEDY REGROUPING.

Choice of mobility: We have experimented with different settings in these datasets, allowing students and voters to move between various distances; our heuristics perform better if we allow movement to farther distances. Ultimately, however, the choice of mobility lies in the hands of the policy-makers who implement measures for redistributing the population. We hope that our conceptual framework enables decision-makers in proposing effective policies for more equitable outcomes, regardless of the exact choice of mobility constraints.



(a) Dove Academy **(b) Universal Academy** **(c) Cesar Chavez Academy**

Figure 3: Racial distribution of students in selected schools before and after applying Greedy Regrouping.

Group size: Since political districts as well as school sizes can not be arbitrary, we set a parameter τ which models group resizing, allowing the algorithm to move individuals in and out of a district without changing its size by more than a τ factor of its original size. Note that the choice of τ lies again in the hands of the policy-makers. We experimented with values in $[0.1, 0.5]$, and observed qualitatively similar results (we report results for $\tau = 0.2$). Furthermore, the size constraints can also be utilized to ensure balance in group sizes.

In summary, Greedy Regrouping and Greedy Connected Regrouping are designed to portray the effect of governmental policies that incentivize physical regrouping of people, whether it is through electoral redistricting or busing (moving students to different schools), while Greedy Regrouping_X is intended as a version of these with no mobility constraints, applicable in online settings like political ad targeting.

6 CONCLUSION & FUTURE DIRECTIONS

In this paper, we considered the problem of fairly dividing people into groups through a voting scenario, with the goal of minimizing the maximum margin of victory across groups. In doing so, we provide a rigorous framework to reason about the complexity of the problem, showing that redistributing people with constraints on their mobility is NP-complete in general, and admits XP algorithms for particular cases.

Furthermore, our fast greedy heuristics show significant improvement of the margin of victory in electoral districting, school assignments, and synthetic experiments. In case of elections, minimizing margin of victory leads to better representation of opposition parties in electoral districts. For school assignment, we show that our greedy algorithms are able to provide more diversity in highly segregated schools. While government policies are ultimately crucial in reducing segregation, we hope that this quantitative analysis can show their potential efficacy.

Multiple directions remain open for future work, such as extending synthetic experiments to real-world effects of political advertising and analyzing the social connections in real datasets, which may change people's mobility constraints. Finally, it would be worthwhile to measure the effect of minimizing the margin of victory on different gerrymandering metrics as well as the effect of decreasing racial segregation on school revenues.

Acknowledgments

We thank the anonymous reviewers for their valuable suggestions. This research was supported in part by a European Research Council (ERC) Advanced Grant for the project "Foundations for Fair Social Computing", funded under the European Union's Horizon 2020 Framework Program (grant agreement no. 789373). A.-A. Stoica was supported by National Science Foundation (grant no. 1514437 and 1761810) and a J. P. Morgan PhD Fellowship. P. Dey acknowledges the support of DST INSPIRE Faculty Fellowship (grant no. DST/INSPIRE/04/2016/001479) and ISIRD grant by IIT Kharagpur.

REFERENCES

[1] Lada A Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*. ACM, 36–43.
[2] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
[3] Athanasios Andreou, Márcio Silva, Fabrício Benevenuto, Oana Goga, Patrick Loiseau, and Alan Mislove. 2019. Measuring the Facebook Advertising Ecosystem. In *NDSS*.
[4] Yoram Bachrach, Omer Lev, Yoad Lewenberg, and Yair Zick. 2016. Misrepresentation in District Voting. In *IJCAI*. 81–87.
[5] Salvador Barberà, Hugo Sonnenschein, Lin Zhou, et al. 1991. Voting by committees. *Econometrica* 59, 3 (1991), 595–609.
[6] Dorothea Baumeister, Tobias Hogrebe, and Lisa Rey. 2019. Generalized Distance Bribery. In *AAAI*. 1764–1771.
[7] Piotr Berman, Marek Karpinski, and Alex D. Scott. 2003. Approximation Hardness and Satisfiability of Bounded Occurrence Instances of SAT. *Electronic Colloquium on Computational Complexity* 10, 022 (2003).
[8] Michelle Blom, Peter J Stuckey, and Vanessa J Teague. 2018. Computing the Margin of Victory in Preferential Parliamentary Elections. In *International Joint Conference on Electronic Voting*. Springer, 1–16.
[9] Allan Borodin, Omer Lev, Nisarg Shah, and Tyrone Strangway. 2018. Big City vs. the Great Outdoors: Voter Distribution and How It Affects Gerrymandering.. In *IJCAI*. 98–104.
[10] Robert Bredereck, Piotr Faliszewski, Andrzej Kaczmarczyk, Rolf Niedermeier, Piotr Skowron, and Nimrod Talmon. 2017. Robustness among multiwinner voting rules. In *SAGT*. 80–92.
[11] David Butler. 1992. The redrawing of parliamentary boundaries in Britain. *British Elections and Parties Yearbook* 2, 1 (1992), 5–12.
[12] David Cary. 2011. Estimating the Margin of Victory for Instant-Runoff Voting.. In *EVT/WOTE*.
[13] Abhijnan Chakraborty, Negar Mohammadi, Nuno Mota, and Krishna P Gummadi. 2020. Fair Partitioning of Public Resources: Redrawing District Boundary to Minimize Spatial Inequality in School Funding.
[14] Abhijnan Chakraborty, Gourab K Patro, Niloy Ganguly, Krishna P Gummadi, and Patrick Loiseau. 2019. Equality of voice: Towards fair representation in crowdsourced top-k recommendations. In *ACM FAT**.
[15] Alvin Chang. 2018. We can draw school zones to make classrooms less segregated. This is how well your district does. (2018).
[16] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. 2017. Fair clustering through fairlets. In *NeurIPS*. 5029–5037.

[17] Amittai Cohen-Zemach, Yoad Lewenberg, and Jeffrey S Rosenschein. 2018. Gerrymandering over graphs. In *AAMAS*. 274–282.

[18] Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *AAAI ICWSM*.

[19] Antonella Corsi-Bunker. 2015. Guide to the education system in the United States. *University of Minnesota* 23 (2015).

[20] Palash Dey and Y Narahari. 2015. Estimating the margin of victory of an election using sampling. In *IJCAI*.

[21] Matthew F. Larsen Douglas N. Harris. 2015. The identification of schooling preferences: methods and evidence from post-Katrina New Orleans. (2015).

[22] EdBuild. 2019. Non-white school districts get $23 Billion less than white districts, despite serving the same number of students. (2019). https://edbuild.org/content/23-billion

[23] Thomas B. Edsall. 2019. Trump Is Winning the Online War. (2019). https://www.nytimes.com/2019/10/16/opinion/trump-digital-campaign.html

[24] Gábor Erdélyi, Edith Hemaspaandra, and Lane A Hemaspaandra. 2015. More natural models of electoral control by partition. In *ADT*. 396–413.

[25] Paul Erdős and Alfréd Rényi. 1960. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* 5, 1 (1960), 17–60.

[26] Marc R Feix, Dominique Lepelley, Vincent Merlin, Jean-Louis Rouet, and Laurent Vidu. 2008. Majority efficient representation of the citizens in a federal union. (2008).

[27] Dan S Felsenthal and Nicholas R Miller. 2015. What To Do about Election Inversions under Proportional Representation? *Representation* 51, 2 (2015), 173–186.

[28] Erica Frankenberg. 2019. What school segregation looks like in the US today, in 4 charts. (2019).

[29] Andrew Gelman, Jonathan N Katz, and Francis Tuerlinckx. 2002. The mathematics and statistics of voting power. *Statist. Sci.* (2002), 420–435.

[30] HUD. 2018. HUD v. Facebook, Inc. *FHED 01-18-0323-8* (2018). https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf

[31] Urban Institute. 2018. Segregated Neighborhoods, Segregated Schools? (2018). urban.org/features/segregated-neighborhoods-segregated-schools

[32] Samuel Issacharoff. 2002. Gerrymandering and political cartels. *Harvard Law Review* 116 (2002).

[33] Takehiro Ito, Naoyuki Kamiyama, Yusuke Kobayashi, and Yoshio Okamoto. 2019. Algorithms for Gerrymandering over Graphs. In *AAMAS*. 1413–1421.

[34] Ron Johnston, David Rossiter, and Charles Pattie. 2006. Disproportionality and bias in the results of the 2005 general election in Great Britain: evaluating the electoral system?s impact. *Journal of Elections, Public Opinion and Parties* 16, 1 (2006), 37–54.

[35] Jon Keegan. 2016. Blue feed, red feed. *The Wall Street Journal* 18 (2016).

[36] Alexander Kent and Thomas C. Frohlich. 2015. The 9 Most Segregated Cities In America. *Huffington Post* (2015).

[37] Matthäus Kleindessner, Pranjal Awasthi, and Jamie Morgenstern. 2019. Fair k-center clustering for data summarization. *arXiv preprint arXiv:1901.08628* (2019).

[38] Yoad Lewenberg, Omer Lev, and Jeffrey S Rosenschein. 2017. Divide and conquer: Using geographic manipulation to win district-based elections. In *AAMAS*. 624–632.

[39] David Lublin. 1999. *The paradox of representation: Racial gerrymandering and minority interests in Congress.* Princeton University Press.

[40] Thomas R Magrino, Ronald L Rivest, Emily Shen, and David Wagner. 2011. Computing the Margin of Victory in IRV Elections.. In *USENIX EVTWOTE.*

[41] Tahiat Mahboob. 2019. How Facebook was harnessed to micro-target voters and promote Donald Trump. (2019). cbc.ca/passionateeye/features/how-facebook-was-harnessed-to-micro-target-voters-and-promote-do-\nald-trump

[42] Cynthia Maushagen, Marc Neveling, Jörg Rothe, and Ann-Kathrin Selker. 2018. Complexity of shift bribery in iterative elections. In *AAMAS*. 1567–1575.

[43] Neeldhara Misra and Chinmay Sonar. 2019. Robustness radius for chamberlin-courant on restricted domains. In *International Conference on Current Trends in Theory and Practice of Informatics*. Springer, 341–353.

[44] JW Montgomery III. 1970. Swann v. Charlotte-Mecklenburg Board of Education: Roadblocks to the Implementation of Brown. *Wm. & Mary L. Rev.* 12 (1970), 838.

[45] Clemens Puppe and Attila Tasnádi. 2008. A computational approach to unbiased districting. *Mathematical and Computer Modelling* 48, 9-10 (2008), 1455–1460.

[46] Clemens Puppe and Attila Tasnádi. 2009. Optimal redistricting under geographical constraints: Why ?pack and crack? does not work. *Economics Letters* 105, 1 (2009), 93–96.

[47] Filipe N Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P Gummadi, and Elissa M Redmiles. 2019. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *ACM FAT**. 140–149.

[48] Meredith P Richards. 2014. The gerrymandering of school attendance zones and the segregation of public schools: A geospatial analysis. *American Educational Research Journal* 51, 6 (2014), 1119–1157.

[49] Melanie Schmidt, Chris Schwiegelshohn, and Christian Sohler. 2018. Fair coresets and streaming algorithms for fair k-means clustering. *arXiv preprint arXiv:1812.10854* (2018).

[50] Philip Seargeant and Caroline Tagg. 2019. Social media and the future of open debate: A user-oriented approach to Facebook's filter bubble conundrum. *Discourse, Context & Media* 27 (2019), 41–48.

[51] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Potential for discrimination in online targeted advertising. In *ACM FAT**.

[52] Rene Van Bevern, Robert Bredereck, Jiehua Chen, Vincent Froese, Rolf Niedermeier, and Gerhard J Woeginger. 2015. Network-based vertex dissolution. *SIAM Journal on Discrete Mathematics* 29, 2 (2015), 888–914.

[53] Pim van 't Hof, Daniël Paulusma, and Gerhard J. Woeginger. 2009. Partitioning graphs into connected parts. *Theor. Comput. Sci.* 410, 47-49 (2009), 4834–4843.

[54] Lirong Xia. 2012. Computing the margin of victory for various voting rules. In *ACM EC*. 982–999.

[55] Imtiaz Masud Ziko, Eric Granger, Jing Yuan, and Ismail Ben Ayed. 2019. Clustering with Fairness Constraints: A Flexible and Scalable Approach. *arXiv preprint arXiv:1906.08207* (2019).