Multi-Objective Optimization for Political Districting: A Scalable Multilevel Approach

Rahul Swamy∗1, Douglas King†1, and Sheldon Jacobson‡2

1Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, 104 S Mathews Ave, Urbana, IL 61801.
2Department of Computer Science, University of Illinois Urbana-Champaign, 201 N Goodwin Ave, Urbana, IL 61801.

Abstract

Political districting in the United States is a decennial process of redrawing the boundaries of congressional and state legislative districts. The notion of fairness in political districting has been an important topic of subjective debate, with district maps having consequences to multiple stakeholders. Even though districting as an optimization problem has been well-studied, existing models primarily rely on non-political fairness measures such as the compactness of districts. This paper introduces optimization models for districting with political fairness criteria. The criteria considered are based on fundamental fairness principles such as vote-seat proportionality (efficiency gap), partisan (a)symmetry, and competitiveness. Given the large sizes of practical instances, a multilevel algorithm is presented, which first reduces instance sizes by a series of graph contractions, and then solves an exact multi-objective problem at the inner level using the $\epsilon$-constraint method. The non-linearity of the partisan asymmetry objective is efficiently tackled with the branch-and-cut method. Case studies on congressional districting in Wisconsin are presented. The results demonstrate that district plans constituting the Pareto-front are equitable, symmetric, competitive, and compact, and that a multi-objective approach is integral to paving the way towards a holistic districting process that addresses all the stakeholders.

1 Introduction

Once every ten years, state and federal institutions in the United States undertake the task of districting (or partitioning) a state into political districts to account for migrations in population. The precise way in which these districts are drawn affects how the voters are spatially divided, which in turn influences electoral representation and the political landscape of a legislature. Given its potential impact, gerrymandering, the act of intentionally manipulating district boundaries to achieve certain political objectives (that certain stakeholders may consider “unfair”), has become a common phenomenon. In the recent past, gerrymandering cases have been appeared in the Supreme Court disputing the district plans of Wisconsin, Maryland and North Carolina (Royden and Li, 2017).

The analysis of fairness in political districting has been a topic of subjective debate, with multiple stakeholders favoring different notions of what constitutes fairness. For example, even when political parties can arrive at a consensus, they may prefer districts that have majorities that
are “proportional” to their share of voters, while the voters themselves may prefer districts that are “competitive” so that their votes have greater impact in representation. It is imperative to take a holistic multi-objective approach to districting.

This paper presents a multi-objective optimization-based districting framework that includes multiple political fairness objectives, extending existing compactness-based models which are exclusively based on non-political data. In addition to compactness, three fairness metrics are considered that explicitly capture political or electoral information. The first metric is the efficiency gap, which measures the difference in the wasted votes for the two major parties. Minimizing the efficiency gap ensures that partisan gerrymandering characterized by “packing” and “cracking” of voters is limited. The second notion of fairness is to ensure that when voter preferences shift in the future, the rate at which the two parties gain or lose seats is symmetric between the two parties (assuming the shift is uniform across all the districts). This paper quantifies and seeks to minimize the extent of asymmetry, called partisan asymmetry. Third, competitiveness of districts is essential to minimize incumbent gerrymandering, and for greater accountability of representatives to their voters. With each of these fairness metrics along with the compactness of districts, the multi-objective problem is solved using an $\epsilon$-constraint method to generate Pareto-optimal solutions.

To generate problem instances sizes that are solvable by an exact method, the solution framework first solves a series of graph contraction problems forming multiple levels of problem instances of increasing coarseness. The coarsest level in this multilevel scheme is used to solve the exact multi-objective problem to produce Pareto-optimal solutions. A case study based on congressional districting in Wisconsin with Census 2010 data is presented. The results demonstrate that district plans can be produced that are equitable, symmetric, competitive, and compact. The implication of these results is that algorithms that optimize across fairness measures are essential to designing a districting process that integrates the conflicting objectives of multiple stakeholders.

The rest of the paper is organized as follows. Section 2 reviews the key literature related to this work. Section 3 introduces the political fairness criteria in the multi-objective problem, the optimization formulations and complexity analysis. Section 4 presents a multi-level solution framework that tackles the related computational challenges. Section 5 presents computational results based on a case study on congressional districting in Wisconsin, and Section 6 provides concluding remarks.

2 Related Literature

This study links two bodies of literature from (i) political sciences and related fields, and (ii) operations research and management sciences. The former covers extensive work on the socio-political requirements and implications of districting, while the latter focuses on efficient methodologies for finding optimal district plans.

2.1 Quantifying Fairness

Fairness is approached from both the partisan perspective and the voters’ perspective. Metrics that quantify partisan fairness typically revolve around proportionality and symmetry. Proportionality is the notion that a party gets more seats (i.e. wins elections in more districts) if more voters favor that party, while symmetry is a more general idea where the outcomes from a given district plan favor both the parties symmetrically with respect to every possible share of voters. Partisan fairness metrics aim at measuring the extent of partisan gerrymandering. Incumbent gerrymandering is a commonly observed (bi)partisan effort at drawing districts that intentionally increases the likelihood that certain (incumbent) candidates win (Silverberg, 1995). Hence, from the voters’ perspective,
one school of thought is that the *competitiveness* of a district is a proxy measure against incumbency. Additionally, competition increases accountability of representatives to their constituents (Jones, 2013).

The seminal work by Gallagher (1991) introduces proportional representation as a fairness criterion, and reviews six ways to measure proportionality. Linear regression models that fit vote-seat curves based on election results have also been analyzed (Gelman and King, 1994). Recently, Stephanopoulos and McGhee (2015) introduce the *efficiency gap* (EG), which measures the difference in the wasted votes between the two major parties. A fair district plan according to EG has a constant of proportionality (i.e. slope) of two (Warrington, 2018). EG has emerged as the most prominent measure of partisan fairness owing to its elegant definition and simplicity in computation, and has even been used by the Supreme Court in Wisconsin’s partisan gerrymandering case. The introduction of EG has further spurred a plethora of research into alternative metrics to quantify partisan gerrymandering from other angles, in addition to analyzing the limitations and applicability of EG (Chambers et al., 2017; Stephanopoulos and McGhee, 2018).

Despite its popular adoption, proportionality is not a constitutional requirement in the US [REF]. An alternative school of thought seeks the broader notion of partisan symmetry. Grofman and King (2007) note that the symmetry standard requires “the electoral system to treat similarly-situated parties equally, so that each receives the same fraction of legislative seats for a particular vote percentage as the other party would receive if it had received the same percentage [of the vote].” Several ways to quantify partisan asymmetry using vote-seat curves have been explored in the literature, with alternatives in the exact numerical function used (Jackman, 1994; Duchin, 2018). This paper extends prior work on quantifying partisan asymmetry by computing the *area* between the vote-seat curves of the two parties.

While optimizing for partisan fairness criteria creates district plans that are equitable and symmetric to both political parties, it is also important to consider fairness with respect to the voters. District competitiveness is the third metric considered. In highly competitive districts, small changes to the voter preferences lead to larger changes in the seat-share, and are generally considered desirable in preventing the creation of *safe seats*, or incumbent gerrymandering (Tapp, 2018). Past studies have analyzed the correlation between competitive districts and other metrics. For example, McCarty et al. (2009) note that partisan gerrymandering could in some cases create competitive districts, when one party “cracks” the other party’s voters, just enough for the other party to lose. Other works show unclear results on the trends in polarized (or non-competitive) districts based on congressional and presidential elections (Abramowitz et al., 2006; McCarty et al., 2009). Regardless, the general consensus is that more competitive elections establish a formal relationship between the voters and the representatives (Hirano and Snyder, 2012). This paper seeks to maximize competitiveness by minimizing the maximum margin of victory in all the districts. The consequence of doing so is that “packing” of voters is minimized.

### 2.2 Political Fairness-Based Algorithms

Simulation-based algorithms form the predominant approach among districting algorithms that consider political fairness objectives. A simulation algorithm typically generates a large set of *independent* district plans using a suitable random sampling method, with the intent of studying patterns in political objectives, while also serving as a benchmark for existing district plans. For instance, Chen and Rodden (2013) simulate district plans across US and conclude that Democratic voters being concentrated in certain areas leads to district plans that strongly favor Republicans. Fifield et al. (2015) provides the first step towards a theoretical framework based on Markov chain Monte-Carlo simulations while incorporating feasibility constraints such as contiguity and popula-
tion balance. Recent work by Liu et al. (2016) highlights the advances in utilizing computational tools such as high-performance parallel computing, where the analysis is performed on 1,174,702 randomly generated district plans.

In the optimization side, Puppe and Tasnádi (2008) consider the problem of unbiased (proportional) districting and show that the problem is NP-Hard. King et al. (2015) consider a competitiveness metric as their objective, and design a local search method to find approximate district solutions. However, competitive districting as an exact optimization problem has not been analyzed. Moreover, the efficiency gap and the partisan asymmetry metrics have not been incorporated in an optimization setting. Given these gaps in the literature, there is an opportunity for research in the integration of optimization methods in political fairness-based districting. This paper bridges this gap by incorporating multiple political fairness metrics into optimization modeling, in addition to providing a scalable solution method.

2.3 Optimization Approaches to Districting

Political districting is NP-Complete, with a variety of solution methods proposed and studied (Ricca et al., 2013). For exact methods, contiguity enforcement has been a major challenge. A method that is used in practice has been to solve the problem without contiguity constrains (Gentry et al., 2015), and then using a heuristic to alter a discontiguous solution to satisfy contiguity (Salazar-Aguilar et al., 2011). Drexl and Haase (1999) are the first to provide a set of constraints that ensure contiguity in districts (albeit exponential in number) when using an MIP-based model. Later, Shirabe (2009) provides a flow-based model for contiguity enforcement in a unit-allocation problem (a special case of districting). Recent work by Haase and Müller (2014) extends the formulation to a sales-force districting problem. In addition to the traditional branch-and-bound based methods, several alternative exact approaches have also been explored. Garfinkel and Nemhauser (1970) propose a two-stage strategy, where the first stage enumerates all feasible (contiguous) districts and the second stage solves a set covering problem by column generation. Mehrotra et al. (1998) suggests a different set-partitioning approach considering the compactness objective, and a post-processing stage where over-populated districts transfer some of their population to under-populated districts using network flows.

Approximation algorithms are popular in practical districting applications, when exact methods cannot find optimal solutions in reasonable time (Ricca and Simeone, 2008). Classic heuristic methods include simulated annealing (Browdy, 1990; D'Amico et al., 2002), tabu search (Bozkaya et al., 2003), local search (Ricca and Simeone, 2008), genetic algorithm (Bacao et al., 2005), and polygonal clustering (Joshi et al., 2012), among others. Computational geometry-based techniques using Voronoi regions have also been explored (Ricca et al., 2008). While exact and approximation methods are studied separately, this paper presents a “coarsen-first solve-next” approach that first heuristically reduces the problem data by a coarsening scheme, and exactly solves the reduced-size problem.

2.4 Multilevel Algorithms

Multilevel algorithms are a class of graph-based algorithms that create multiple hierarchical levels of graphs by an iterative coarsening procedure (where adjacent nodes are merged to create a smaller graph) and then mapping a district solution from the reduced graph instance back to the the original graph (Kernighan and Lin, 1970; Hendrickson and Leland, 1995). They are considered the most effective algorithms for graph partitioning problems with applications in VLSI design and image processing (Buluç et al., 2016). The multilevel paradigm has also been adopted to other
combinatorial optimization problems such as the traveling salesman problem (Walshaw, 2002) and graph-drawing (Walshaw, 2000). Districting problems are different from graph partitioning problems (whose objective is to minimize the edge-cut weights) in that districting typically involves more constraints such as contiguity, tighter population balance constraints, among others, in addition to the differences in objectives. Adapting this algorithm to districting, this paper provides innovations in two aspects. First, the coarsening procedure presented introduces a matching problem based on unit populations that minimizes the most populous matched edge. The idea behind this is to produce a homogeneous population distribution in the next level, thereby enabling more feasible districts that satisfy the population balance constraints. Second, the classical multilevel algorithm coarsens a graph till the number of nodes is as small as the number of districts. The algorithm presented coarsens only till the graph is small enough for an exact method (less than 200 nodes), and then solves the MIP to optimality. Thereby, the presented framework embeds an exact method within the multilevel algorithm.

3 Multi-objective Political Districting

This section introduces optimization models for political districting with explicit fairness objectives. The general multi-objective districting model for the problem is introduced, followed by the formulation and analysis of each of the single objective optimization problems.

A Districting Problem (DP) partitions a geographical region into a finite number of districts where each district satisfies certain requirements. The input to DP consists of a set of discrete geographical units $V$ (or simply units), their adjacency information, and the number of districts, $K$. Each unit represents an area of administrative significance such as a county, census tract, or a census block. Two units $i, j \in V$ are said to be adjacent if they share a boundary curve which contains a continuous segment (as opposed to being a set of singleton points). An adjacency graph $G = (V, E)$ is an undirected graph where the set of nodes $V$ represents the set of units, and an edge $(i, j)$ exists in $E \subseteq V \times V$ if and only if units $i$ and $j$ are adjacent. Additionally, each unit $i \in V$ contains a certain population of residents, denoted by $p_i \geq 0$. A district is a subset of $V$ that satisfies certain conditions, and a district plan is a partition of $V$ into $K$ districts, i.e. $z : [K] \rightarrow 2^V$ denotes a district plan where for each $k \in [K]$, $z(k) \subset V$ is a district. A district plan $z$ is said to be feasible for DP if for each district $k \in [K]$, the subgraph induced by $z(k)$ in $G$ is connected, and the population $\sum_{i \in z(k)} p_i$ belongs to $[lb, ub]$, where $lb$ and $ub$ are the lower and upper bounds on the population required in each district, respectively. Typically, the $(lb, ub)$ values are set to be $[\overline{P}(1 - \tau), \overline{P}(1 + \tau)]$, where $\overline{P} := (\sum_{i \in V} p_i)/K$ is an ideal average population expected in each district, and $\tau \in [0, 1]$ is the population deviation tolerance. Given an objective function $\phi : [K] \times 2^V \rightarrow \mathbb{R}$, DP finds a feasible district plan that minimizes $\phi$.

The Political Districting Problem (PDP) is a multi-objective generalization of a DP, with each objective representing a political fairness metric. In addition to the inputs for DP, in order for district plans to be evaluated on their political fairness, PDP also requires information on the political leanings of voters. Consider a two-party electoral system, where parties A and B have a distribution of voters spread among the units, i.e., let $p_i^A, p_i^B \geq 0$ (with $p_i^A + p_i^B \leq p_i$) be the number of voters for parties A and B respectively residing in unit $i \in V$. For a given district plan $z$, let $P_k^r(z) := \sum_{i \in z(k)} p_i^r$ be the number of voters for party $r \in \{A, B\}$ in district $k \in [K]$. In a plurality-based election (such as in US congressional elections), a party A (B) is said to win (lose) district $k$ if $P_k^A(z) > P_k^B(z)$, i.e. if A has more voters in district $k$ than party B. Note that ties rarely occur in practice when the number of voters is large, and the units are discrete with a heterogeneous distribution of voters. For a given district plan, based on the political leanings of the
voters in each district and the districts won/lost by the two parties, one can assess how politically fair the district plan is. There are several ways to assess political fairness. Let \( \{ \phi_q \}_{q=1}^Q \) be a set of \( Q \geq 1 \) political fairness functions. Here, for each \( q \in [Q] \), \( \phi_q : [K] \times 2^V \to \mathbb{R} \) represents a real-valued function that measures the fairness of a district plan \( z \). PDP is a \( Q \)-objective optimization problem that minimizes the set of objective functions \( \{ \phi_q \}_{q=1}^Q \) within the solution space of DP.

PDP can be formulated as follows. The primary decision variables that assigns units to districts is given by,

\[
x_{ij} = \begin{cases} 
  1, & \text{if } j (\neq i) \in V \text{ is assigned to the district with center } i \in V. \\
  0, & \text{otherwise}.
\end{cases}
\]

\[
x_{ii} = \begin{cases} 
  1, & \text{if unit } i \in V \text{ is chosen as a district center.} \\
  0, & \text{otherwise}.
\end{cases}
\]

The contiguity of districts is enforced using flow variables defined as follows. Originating from every unit \( i \in V \), let \( f_{ivj} \geq 0 \) be the amount of flow on edge \((v,j) \in E\), whose value is positive only if \( i \) is a district center, and units \( v \) and \( j \) are assigned to the district with center \( i \). PDP is given by,

\[
\text{(PDP) Minimize } \{ \phi_1(x), \phi_2(x), \ldots, \phi_Q(x) \}
\]

\[
\text{subject to } (1 - \tau)P x_{ii} \leq \sum_{j \in V} x_{ij} p_j \leq (1 + \tau)P x_{ii} \quad \forall \ i \in V, \\
\sum_{i \in V} x_{ii} = K, \\
\sum_{i \in V} x_{ij} = 1 \quad \forall \ j \in V, \\
x_{ij} \leq x_{ii} \quad \forall \ i, j \in V, \\
x_{ij} + \sum_{v \in N(j)} (f_{ijv} - f_{ivj}) - \sum_{v \in V \mid i = j} x_{iv} = 0 \quad \forall \ i, j \in V, \\
|V| x_{ij} - \sum_{v \in N(j)} f_{ivj} \geq 0 \quad \forall \ i, j \in V, \\
x_{ij} \in \{0,1\}, \ f_{ivj} \geq 0 \quad \forall \ i, j \in V, v \in N(j).
\]

The tuple of objectives in (1) are minimized in a PDP. Constraints (2) define the upper and lower bounds on district populations. Constraints (3) ensure that there are exactly \( K \) districts. Constraint (4) assigns a unique district to each unit. Constraints (5) ensure that a unit can be assigned to a district with unit \( i \) as its center only if \( i \) is district center, i.e., \( x_{ii} = 1 \). Originally introduced by Shirabe (2009), flow-based constraints (6)-(7) have recently become popular for contiguity enforcement in MIPs. Here, from every district center \( i \in V \), a certain amount of flow, \( f_{ivj} \geq 0 \), is sent through edge \((v,j) \in E\), and the constraints ensure that flow is conserved if and only if the district is contiguous. Constraints (8) define the variable domains.

This paper considers four fairness objectives: compactness, efficiency gap, partisan asymmetry and competitiveness of districts. The rest of this section provides background for each of these objectives and analyzes the mathematical formulation of the single objective optimization problems resulting from considering each of the four objectives at a time. Computational hardness results are also presented.
3.1 Compact Districting Problem

Compactness of a district is a subjective measure that evaluates a district’s geometrical shape, with several ways to quantify it (Young, 1988). A popular metric is the dispersion of units assigned to each district, measured by the sum of their distances from the geographical center of the district. A unit is said to be the district center of a district if it has the least sum of distances to other units in the district. Let $d_{i,j} \geq 0$ be the distances between every pair of units $i, j \in V$, and let $c_k \in V$ be the center of district $k$. The Compact Districting Problem, or Compact-DP, finds a district plan that minimizes this compactness objective, and can be formulated as,

$$ (\text{Compact-DP}) \ \text{Minimize } \phi_{\text{comp}}(x) := \sum_{i,j \in V} d_{ij} x_{ij} $$

subject to (2) – (8).

The compactness objective $\phi_{\text{comp}}$ in (9) minimizes the sum of distances from the center of each district. Compact-DP is a well studied classical problem that generalizes the NP-complete p-median problem (Megiddo and Supowit, 1984). Even though compact districts are generally perceived to be fair, often times even optimally compact districts could yield unfair results upon consideration of voter information, as illustrated in the following example.

![Figure 1: Example with four districts on a grid graph suggesting that an optimally compact solution may not yield a proportional seats-share to the two parties.](image)

Consider an example with units forming a $10 \times 10$ grid (Figure 1), where each $1 \times 1$ square is a unit. Assume that the shaded units are densely populated with party A voters, whereas the unshaded units are populated with party B voters, i.e. $(p_i, p_i^A, p_i^B)$ for each unit $i$ is: $(40, 40, 0)$ for the shaded units and $(40, 0, 10)$ for the unshaded units. Here, the total number of voters for A and B are 1000 and 750, respectively. This example illustrates a scenario of spatial polarization of voters in a real-life setting where urban and rural areas having different concentrations of party support has been observed (Chen and Rodden, 2013). Consider the four districts defined by the four equal-sized quadrants of the grid in Figure 1a. Clearly, these districts satisfy the contiguity and population balance constraints, and are also optimally compact. However, the seat-share for B is 0.75, while their vote-share is 0.43. From a proportionality stand-point, this district plan is not ideal since B wins three times the number of districts as A does, even though B’s vote share is less than A’s. In contrast, in the district plan in Figure 1b, the seat-share for B is 0.5, which is closer to B’s vote-share than that of the previous district plan despite having less compact districts. Hence, there is a need to explicitly consider voter information in an optimization model in addition to the compactness objective.
### 3.2 Equitable Districting Problem

In a scenario where one party (A) gerrymanders, the phenomenon of packing and cracking is observed. While creating districts, packing is when a large number of the other party (B) voters are assigned to a small number of districts, while cracking is when the remaining party B voters are diluted into the rest of the districts in which party A holds a majority. Even though this phenomenon has been widely discussed in decades of gerrymandering research, it was not until recently that a metric to quantify packing and cracking has been formulated. The efficiency gap introduced by Stephanopoulos and McGhee (2015) uses the concept of wasted votes. A vote in a certain district is said to be wasted if it is either a surplus vote for the winning party, or a vote to the losing party. In each district \( k \in [K] \), let \( w^A_k(z) \) be the wasted votes for party A given by,

\[
w^A_k(z) = \begin{cases} 
  P^A_k(z) - \frac{P^A_k(z) + P^B_k(z)}{2}, & \text{if } P^A_k(z) \geq P^B_k(z) \\
  \frac{P^A_k(z) + P^B_k(z)}{2}, & \text{otherwise}.
\end{cases}
\]

Exactly half the votes in every election are wasted. However, if one party disproportionately wastes more votes than the other, it is considered to be an indication of partisan gerrymandering (Stephanopoulos and McGhee, 2015). Among the wasted votes, surplus votes quantify packing, while the losing votes quantify cracking. The difference in wasted votes in each district \( k \) between the two parties is given by,

\[
w^A_k(z) - w^B_k(z) = \begin{cases} 
  \frac{P^A_k(z) - P^B_k(z)}{2}, & \text{if } P^A_k(z) \geq P^B_k(z) \\
  \frac{3P^A_k(z) - P^B_k(z)}{2}, & \text{otherwise}.
\end{cases}
\]

For a given district plan \( z \), the efficiency gap \( \phi_{EG} \) measures the magnitude of the net difference in wasted votes between two political parties A and B across all the districts normalized by the total number of voters, defined as,

\[
\phi_{EG}(z) := \frac{\sum_{k=1}^{K} |(w^A_k(z) - w^B_k(z))|}{\sum_{k=1}^{K} (P^A_k(z) + P^B_k(z))}.
\]

Note that for any given district plan \( z \), \( \phi_{EG}(z) \) is a value between 0 and 0.5, where a value of 0 implies that both parties have wasted an equal number of votes. In this paper, a district plan is said to be optimally equitable if both the parties waste the same number of voters (i.e., with efficiency gap close to 0). The Equitable Districting Problem is a districting problem whose objective is to minimize the efficiency gap, \( \phi_{EG}(z) \). The complexity of this problem can be shown to be NP-complete when contiguity constraints (6)-(7) are relaxed.

**Theorem 1.** For a general graph \( G = (V,E) \), unit populations \( \{p_i,p^A_i,p^B_i\}_{i \in V} \) and number of districts \( K \geq 2 \), the Equitable Districting Problem is NP-complete.

The proof (presented in Section A of the online appendix) proceeds by relaxing the contiguity and population balance constraints, and then reducing from a variant of the number partitioning problem whose objective is constructed to mirror the efficiency gap function. Even though a relaxation of the problem is considered, the result suggests that Equitable-DP could be intractable for a planar graph instance as well. Equitable-DP can be formulated as a mixed integer program (MIP). Let \( a_i \in \mathbb{R} \) for unit \( i \in V \) be the difference in wasted votes between parties A and B in district with center \( i \); 0 if \( i \) is not a district center. Additional decision variables are given by,

\[
y^A_i = \begin{cases} 
  1, & \text{if } i \in V \text{ is a district center and party A wins that district} \\
  0, & \text{otherwise}.
\end{cases}
\]
\[ v_{ij}^A = \begin{cases} 1, & \text{if party } A \text{ wins the district with center } i, \text{ and } j \in V \text{ is assigned to that district.} \\ 0, & \text{otherwise.} \end{cases} \]

Equitable-DP can be formulated as,

\begin{align*}
\text{(Equitable-DP) Minimize } & \phi_{EG}(x) = \frac{|\sum_{i \in V} a_i|}{\sum_{i \in V} (p_i^A + p_i^B)} \\
\text{subject to} & (2) - (8), \\
& -M \leq \sum_{j \in V} (p_j^A - p_j^B) x_{ij} - M y_i^A \leq 0 \quad \forall i \in V, \\
v_{ij}^A \leq x_{ij} \quad \forall i, j \in V, \\
v_{ij}^A \leq y_i^A \quad \forall i, j \in V, \\
v_{ij}^A \geq x_{ij} + y_i^A - 1 \quad \forall i, j \in V, \\
a_i = \sum_{j \in V} \left( \frac{3p_j^A - p_j^B}{2} \right) x_{ij} - \sum_{j \in V} (p_j^A + p_j^B) v_{ij}^A \quad \forall i \in V, \\
v_{ij}^A, y_i^A \in \{0, 1\}, \quad w_i \in \mathbb{R} \quad \forall i, j \in V. 
\end{align*}

The efficiency gap objective \( \phi_{EG} \) in (11) minimizes the absolute value of the net difference in wasted votes normalized by the total number of voters. Constraints (12) define \( \{y_i^A\}_{i \in V} \) in relation to \( \{x_{ij}\}_{i, j \in V} \). Note that \( M \) can be any value greater than \( P(1 - \tau) \). The first inequality ensures that \( y_i^A = 1 \Rightarrow \sum_{j \in V} (p_j^A - p_j^B)x_{ij} \geq 0 \). The second inequality ensures that \( \sum_{j \in V} (p_j^A - p_j^B)x_{ij} \geq 0 \Rightarrow y_i^A = 1 \). Constraints (13)-(15) are quadratic linearization constraints that ensure that \( v_{ij}^A = 1 \) if and only if \( x_{ij} = y_i^A = 1 \). Constraints (16) define the difference in wasted votes \( \{a_i\}_{i \in V} \) based on Equation (10). Constraints (17) define the variable domains.

### 3.3 Symmetric Districting Problem

The notion of partisan (a)symmetry stems from the principle that if voters shift their preferences from one party (A) to another party (B) uniformly across all the districts, the rate at which B wins districts must be \emph{symmetric}, i.e., it must be the same rate as A wins districts if the voters shifted their preferences from B to A uniformly across all the districts. The corollary could be stated as: given a district plan, A’s share of seats must the same as B’s share of seats if B received the same number of votes that A did (Duchin, 2018). For example, assume that A receives 60% of the votes across all the districts and wins 75% of the seats under a certain district plan. Then, the district plan is symmetric if: had B received 60% vote-share, B would have received 75% of the seats in the same district plan. In order to quantify partisan asymmetry for a given district plan, the first step is to construct a vote-seat curve. Starting from the actual vote-share with respect to one fixed party (A), its vote-share is altered by adding (or subtracting) a small number of voters at a time uniformly across all the districts, in order to observe how the seat-share for party A increases (or decreases). This is recorded for all values of A’s vote-shares between 0 and 1. The next step is to find the symmetric reflection of this curve about the line \( y = 1 - x \), which represents the vote-seat curve for the other party B. Partisan asymmetry is the extent to which the vote-seat curve is not symmetric, which can be measured by the area between the vote-seat curves for A and B. A highly asymmetric district plan is one in which for the same number of votes, one party would receive significantly more seats than the other party.
Consider an example where the vote-seat curve for a district plan in Wisconsin is depicted in Figure 2. The vote-seat curve with respect to the Democratic party (D) is shown in a solid-blue line. The symmetric reflection shown using a dashed-red line shows the vote-seat curve for Republican party (R). This district plan has a partisan asymmetry of 0.11. Note that when D’s vote-share is 60%, D’s corresponding seat-share is 62.5%, whereas when R’s vote-share is 60%, R’s corresponding seat-share is 75%. This asymmetry is captured for all vote-shares by the area between the two curves. In contrast, the optimal district plan when minimizing partisan asymmetry has the vote-seat curves shown in Figure 5c with a partisan asymmetry of 0.0089.

For any given district plan \( z \), partisan asymmetry is now formalized. The overall vote-share for party \( r \in \{A, B\} \) is the total fraction of \( r \) voters, given by \( \alpha_r := \sum_{i \in V} p^r_i / \sum_{i \in V} (p^A_i + p^B_i) \). The overall seat-share for a party is the fraction of districts won by that party. For a given district plan \( z \), let \( P^r_k(z) := \sum_{i \in z(k)} p^r_i \) and \( \alpha^r_k(z) := P^r_k(z) / \sum_{i \in z(k)} (p^A_i + p^B_i) \) be the number and fraction of voters for party \( r \in \{A, B\} \) in district \( k \in [K] \).

In order to simplify the notation, the rest of discussion is with respect to an arbitrary party, say A. Starting with \( \{\alpha^A_k(z)\}_{k \in [K]} \), the idea behind computing partisan asymmetry is to vary A’s district vote-shares homogeneously across all the districts in \( z \) to range from 0 to 1 with the intention of observing its effect on A’s overall seat-share. Let \( \delta \in [-1, 1] \) be the fractional value of A’s voters that is uniformly added (i.e., \( \delta > 0 \)) or subtracted (i.e., \( \delta < 0 \)) from all the districts. For a given \( \delta \), let \( \mu_k(z, \delta) \in [0, 1] \) be A’s corresponding district vote-share in district \( k \) defined by,

\[
\mu_k(z, \delta) := \begin{cases} 
\min\{ (\alpha^A_k(z) + \delta), 1 \}, & 0 < \delta \leq 1 \\
\max\{ (\alpha^A_k(z) + \delta), 0 \}, & -1 \leq \delta \leq 0 
\end{cases}
\]

Note that \( \mu_k(z, 0) = \alpha^A_k(z) \) for each \( k \in [K] \). The district vote-share \( \mu_k \) in each district \( k \) increases linearly as \( \delta \) increases, and is bounded between 0 and 1 by definition. Based on A’s vote-share in each of the districts in \( z \), A’s overall vote-share across all the districts, denoted by \( \mu(z, \delta) \in [0, 1] \), is a function of \( \delta \) given by,

\[
\mu(z, \delta) := \frac{\sum_{k=1}^K (P^A_k(z) + P^B_k(z)) \mu_k(z, \delta)}{\sum_{k=1}^K (P^A_k(z) + P^B_k(z))}.
\]  

(18)
The seat-share for A across all the districts in z is the fraction of districts that A wins, which is also a function of \( \delta \) is given by \( s(z, \delta) := \{k \in [K] : \mu_k(z, \delta) > 0.5\} / K \). As \( \delta \) increases, A’s seat-share follows a step function which increases by a value of \( 1/K \) every time A’s district vote-share in a certain district crosses the 0.5 mark. Based on A’s overall vote-share and the seat-share values for every possible value of \( \delta \), a vote-seat curve can be constructed as follows. The vote-seat curve for A is the set of points given by \( \{(\mu(z, \delta), s(z, \delta))\}_{\delta \in [−1, 1]} \) between A’s overall vote-share and its seat-share generated by parameter \( \delta \in [−1, 1] \). Let \( s_{\mu(z)} \) denote A’s seat-share as a function of A’s overall vote-share. Note that \( 1 − s_{1−\mu(z)} \) is the seat-share for party B when the overall vote-share for B is \( \mu(z) \). Thus, for a district plan \( z \), the partisan asymmetry (PA) is defined by the difference between functions \( s_{\mu(z)} \) and \( 1 − s_{1−\mu(z)} \), formulated as,

\[
\phi_{PA}(z) := \int_0^1 |s_{\mu(z)} - (1 - s_{1-\mu(z)})| d\mu(z). \tag{19}
\]

Here, \( \phi_{PA} \) measures the area between the vote-seat curves for parties A and B, in essence capturing the asymmetry in rewards (number of seats) as the overall vote-share values fluctuate from completely favoring one party to completely favoring the other party. Note that even though the quantities \( (\mu, \delta, s) \) were defined with respect to party A, the value of partisan asymmetry would be the same if the quantities were defined with respect to party B. In a practical setting, since \( s_{\mu(z)} \) follows a step function, the computation of the integral can be avoided as long as the breakpoints of in the step function can be calculated in \( O(K) \) time as detailed in Section 4. The Symmetric Districting Problem is a districting problem whose objective is to minimize the partisan asymmetry function \( \phi_{PA} \), formulated as,

\[
(Symmetric-DP) \text{ Min } \phi_{PA}(x) \text{ subject to } (2) - (8). \tag{20}
\]

The partisan asymmetry objective \( \phi_{PA}(x) \) for a district plan given by \( x \) measures the area between party A’s and party B’s vote-seat curves, as defined in Equation (19). The explicit formulation is omitted since \( \phi_{PA} \) is non-linear with infinite variables (i.e. one for each overall vote-share value \( \mu \in [0, 1] \)), and the proposed solution method tackles this challenge as detailed in the section 4.3.2.

### 3.4 Competitive Districting Problem

Competition in a district can be quantified by the relative difference in the number of voters in that district for each party. Given a district plan \( z \), let the fractional margin in a district \( k \in [K] \) be defined to be the fraction of voters by which the winning party leads the other party. For a given district plan \( z : [K] \rightarrow V \), the competitiveness objective, \( \phi_{cmpthv}(z) \), is defined to be maximum fractional margin among all the districts, i.e.

\[
\phi_{cmpthv}(z) := \max_{k \in [K]} \frac{|P^A_k(z) - P^B_k(z)|}{P^A_k(z) + P^B_k(z)}.
\]

The Competitive Districting Problem is a districting problem whose objective is to minimize the competitiveness function \( \phi_{cmpthv}(z) \). The computational complexity of a related problem is now analyzed. The Aggregate Competitive Districting Problem finds a district solution \( z \) that minimizes the \( \max_{k \in [K]} |P^A_k(z) - P^B_k(z)| \), and it can be shown that this problem is NP-Complete.
Theorem 2. For a general graph \( G = (V, E) \), unit populations \( \{p_i, p_i^A, p_i^B\}_{i \in V} \) and number of districts \( K \geq 2 \), the Aggregate Competitive Districting Problem is NP-Complete.

The proof (presented in Section B of the online appendix) proceeds by relaxing the contiguity and population balance constraints, and showing a reduction from the NP-Hard multi-way number partitioning problem, which partitions set of integers while minimizing the maximum partition size. Even though the complexity are derived for a related problem, the result suggests that Competitive-DP (with a normalized objective) could be intractable. Competitive-DP can be formulated as the following MIP. Let \( h \geq 0 \) be the maximum fractional margin across all the districts. For every pair of units \( i, j \in V \), let \( b_{ij} := h \cdot x_{ij} \) be an artificial quadratic variable introduced to define \( h \) in the MIP; \( b_{ij} \) takes the value \( h \) when \( x_{ij} \) is 1, 0 otherwise. Competitive-DP is given by,

\[
\begin{align*}
(\text{Competitive-DP}) \quad \text{Minimize} & \quad \phi_{\text{cmpttv}}(x) = h \\
\text{subject to} & \quad (2) - (8),
\end{align*}
\]

\[
\sum_{j \in V} (p_j^A + p_j^B) b_{ij} \geq \sum_{j \in V} (p_j^A - p_j^B) x_{ij} \quad \forall i \in V, \quad (22)
\]

\[
\sum_{j \in V} (p_j^A + p_j^B) b_{ij} \geq -\sum_{j \in V} (p_j^A - p_j^B) x_{ij} \quad \forall i \in V, \quad (23)
\]

\[
b_{ij} \leq x_{ij} \quad \forall i, j \in V, \quad (24)
\]

\[
b_{ij} \leq h \quad \forall i, j \in V, \quad (25)
\]

\[
b_{ij} \geq h - (1 - x_{ij}) \quad \forall i, j \in V, \quad (26)
\]

\[
h, b_{ij} \geq 0 \quad \forall i, j \in V. \quad (27)
\]

The competitiveness objective \( \phi_{\text{cmpttv}} \) in (21) minimizes the maximum fractional margin \( h \). Constraints (22)-(23) define \( \{b_{ij}\}_{i,j \in V} \) in relation to fractional margins in each district. Constraints (24)-(26) linearize the quadratic variables \( \{b_{ij}\}_{i,j \in V} \). Constraints (17) defines the non-negativity of the variables.

### 4 Solution Method

This section presents a multilevel solution strategy for solving a PDP. As emphasized earlier, the districting problems considered in this paper are NP-Complete, and an optimization algorithm for districting must address two main challenges. First, the problem instance sizes for district problems are too large to be solved by an exact method. For example, Wisconsin has 1,409 census tracts (units) and 3,857 edges between them, and the MIP (2)-(8) has \( 7.5 \times 10^6 \) variables and \( 6 \times 10^6 \) constraints. Second, the problem has multiple conflicting (non-linear) objectives, and the solution approach must produce solutions that highlight the inherent trade-offs between the objectives. This section provides a multilevel optimization framework which tackles these challenges. In this approach, the problem instance is iteratively reduced in size by solving a series of graph contractions, followed by an exact-solution method to solve PDP using the reduced problem instance.

#### 4.1 Multilevel Framework

Algorithm 1 provides an outline of the multilevel optimization approach used in this study. There are three parts to the multilevel algorithm: (a) coarsening, (b) solving exactly at the coarsest-level, and (c) un-coarsening. The coarsening stage contracts the graph in multiple steps, thereby creating levels of graphs that are progressively coarser. The coarsest level of graph is used as an input to
the exact solution stage, using which the districting problem is solved to optimality. In the un-
coarsening stage, the optimal solution at the coarsest level is mapped backed to the finer levels,
while performing local-search improvements at each of the levels.

Algorithm 1: Schematic of a multilevel districting algorithm

Input: Level 0 adjacency graph \( G_0 = (V_0, E_0) \), population \( p_i \) \( \forall i \in V \), number of levels \( L \)
Output: \( z^*_0 \), the district solution at level 0
1 \( \{G_0, G_1, \ldots, G_L\} \leftarrow \) Coarsen from level 0 through \( L \);
2 \( z^*_L \leftarrow \) Optimal solution to the districting problem at the \( L \)-th level;
3 \textbf{for} level \( l \in \{L, L-1, \ldots, 1\} \) \textbf{do}
4 \hspace{1em} \( z^*_{l-1} \leftarrow \) Execute heuristic improvement at \( G_l \) with \( z^*_l \) as initial solution;
5 \textbf{end}

4.2 Coarsening

Starting from the given adjacency graph \( G_0 \) and the number of levels \( L \), graphs \( G_1, G_2, \ldots, G_L \) are
constructed such that size of \( G_l \) is smaller than \( G_{l-1} \) for all \( l \in [L] \). Even though the primary
purpose of a coarsening procedure is to reduce the size of a problem instance, it is equally important
to ensure that the reduced instances are likely to produce good feasible solutions to the districting
problem. Based on experimental observations, a key challenge here is that combining highly
populated units often leads to an imbalance in the population among the combined units, thereby
increasing the difficulty in finding even a feasible district solution (after combining the units) that
satisfies the population balance constraints.

To explicitly promote the creation of a homogeneous population distribution among the units
after coarsening, a matching problem is solved at each level, with higher preference given to con-
tracting pairs of units with relatively smaller populations. Given a weighted undirected graph
\( G = (V, E) \) and edge weights \( u_e \forall e \in E \), a matching \( M \subseteq E \) is a subset of edges such that no two
edges share the same node, i.e. \( (i, j), (i, j') \in M \Rightarrow j = j' \). A maximal matching \( M \) is a matching
such that no additional edges can be added to \( M \) while \( M \) remaining a matching, i.e. \( \forall e \in E \setminus M, \)
\( M \cup \{e\} \) is not a matching. Then, the minimax weight matching problem finds a maximal matching
\( M \) such that \( \max_{e \in M} u_e \) is minimized. At level \( l \in 0, 1, \ldots, L - 1 \), given \( G_l = (V_l, E_l) \) and the
populations \( p_i \) \( \forall i \in V_l \), an instance of minimax weight matching is created by setting the edge
weight \( u_{i,j} = p_i + p_j \) for all \( (i, j) \in E_l \). With these chosen weights, the minimax matching problem
achieves the goal of minimizing the most populous pair of units when contracted.

The next level \( G_{l+1} = (V_{l+1}, E_{l+1}) \) is created by contracting every matched edge from \( G_l \) into
a single node in \( V_{l+1} \). Let \( C : V_l \rightarrow V_{l+1} \) be the mapping of a node from level \( l \) to level \( l + 1 \) such
that for \( i, j \in V_l, C(i) = C(j) \) if and only if \( (i, j) \in M^*_l \), where \( M^*_l \) is the minimax weight matching
in \( G_l \). Further, for every edge \( (i, j) \in E_l \setminus M^*_l \), there exists an edge \( (C(i), C(j)) \in E_{l+1} \), and the
population of a node in \( V_{l+1} \) is the sum of the populations of the corresponding matched units in
\( V_l \). The number of voters are also similarly aggregated.

All the levels are created using a series of matching problems. The minimax weight matching
problem is solved using a greedy algorithm as follows. Starting with an empty matching, \( M = \emptyset \),
the edge \( e^* := \arg \max_{e \in E \setminus M} u_e \) is added to \( M \) iteratively until \( M \) is a maximal matching. This
algorithm runs in \( O(|E|) \) time. If the maximal matching is close to a perfect matching (i.e. all the
nodes are matched), the next level will have close to half the number of nodes of the previous level.
Further, the number of levels \( L \) can be decided either apriori or on-the-fly such that the size of the
coarsest level is appropriate for the available computing resources for an optimization solver to be able to solve the districting problem exactly.

4.3 Exact Method for PDP

This subsection presents the inner stage of the multilevel algorithm comprising of an exact solution strategy for solving a PDP, using the coarsest level of graph $G_L$ as used as the input. In a multi-objective optimization problem with $Q \geq 2$ objective functions given by $\phi = \{\phi_1, \phi_2, \ldots, \phi_Q\}$ and a solution space $\mathcal{X}$, a solution $\hat{x} \in \mathcal{X}$ is said to dominate another solution $x \in \mathcal{X}$ if $\phi_q(x) \leq \phi_q(\hat{x})$ for all $q \in [Q]$, and $\phi_{q'}(x) < \phi_{q'}(\hat{x})$ for some $q' \in [Q]$. Further, $\hat{x} \in \mathcal{X}$ is said to be Pareto-optimal (or efficient) if for all other $x \in \mathcal{X}$, $\hat{x}$ dominates $x$. Let $\mathcal{P}$ denote the set of all Pareto-optimal solutions, or the Pareto set. The Pareto frontier is the set of values $\{\{\phi_1(\hat{x}), \phi_2(\hat{x}), \ldots, \phi_Q(\hat{x})\}\}_{\hat{x} \in \mathcal{P}}$ in the objective space corresponding to Pareto-optimal solutions.

The $\epsilon$-constraint method is a popular method to compute the Pareto set in multi-objective problems (Chirdumsi et al., 2012). This method first selects one of the objectives as a primary objective, while the other (secondary) objectives are introduced as constraints of the form $\phi_q(x) \leq \epsilon_q \forall x \in \mathcal{X}$, where $\epsilon_q$ is an upper bound on a secondary objective $q$. For different values of $\epsilon_q$, the problem is solved as a single-objective problem, and the dominated solutions are removed to produce the Pareto-optimal solutions. The resulting objective values represent the Pareto frontier.

For a PDP, the compactness objective ($\phi_{\text{comp}}$) is chosen as the primary objective since the dispersion function (sum of distances from district center) is structurally similar to the $p$-median problem, which is known to possess good linear programming lower bounds (Salazar-Aguilar et al., 2011). This was experimentally verified by setting each of the other fairness objectives as the primary objective, which resulted in a lack of convergence to optimality. Hence, the solution approach is to solve Compact-DP supplemented by constraints related to the other three fairness metrics (as formulated in their single-objective problems) and the $\epsilon$-constraints. The corresponding $\epsilon$-constraints for $\phi_{\text{EG}}$ and $\phi_{\text{cmpttv}}$ are given by constraints (28) and (29) where $\epsilon_{\text{EG}}$ and $\epsilon_{\text{cmpttv}}$ are upper bounds on the efficiency gap and competitiveness, respectively.

$$-\epsilon_{\text{EG}} \sum_{i \in V} (p_i^A + p_i^B) \leq \sum_{i \in V} a_i \leq \epsilon_{\text{EG}} \sum_{i \in V} (p_i^A + p_i^B), \quad (28)$$

$$h \leq \epsilon_{\text{cmpttv}}. \quad (29)$$

Here, $a_i$ is the difference in wasted votes between parties A and B in a district with center $i \in V$, and $h$ is the maximum fractional margin across all the districts. For partisan asymmetry $\phi_{PA}$ with upper bound $\epsilon_{PA}$, the corresponding $\epsilon-$constraint is enforced by verifying the feasibility of the incumbent solutions within the branch-and-cut tree as detailed in subsection 4.3.2.

In addition to solving the MIP models presented in Section 3, the efficiency of the exact method can be improved by the quality of an initial feasible solution, and the efficiency of computations within a node in the branch-and-bound tree, as detailed in the following subsections.

4.3.1 Initial solution heuristic

A multistart local-search heuristic is used to generate an initial feasible solution that is introduced at the root of the branch-and-cut tree. A feasible solution is one that satisfies contiguity, population balance, and $\epsilon$-constraints. To generate a feasible solution, the heuristic proceeds as in Bozkaya et al. (2003) as follows. An arbitrary unit is first selected to seed a district, followed by extending the district by assigning the neighboring units to it. This continues until the population in the district exceeds $\mathcal{P}$, after which an unassigned unit is chosen to seed a new district, and so on.
until all the units are assigned to a district. At the end of this procedure, it is possible to have either more than, or less than \( K \) districts. If there are more than \( K \) districts, pairs of neighboring districts are iteratively merged. If there are fewer than \( K \) districts, more populous districts are split into smaller ones while preserving contiguity. This method requires \( O(|V|) \) time to produce \( K \) contiguous districts. If this solution does not satisfy the population balance or fairness constraints, it is discarded, and the method is repeated until a feasible solution is found.

In addition to finding a feasible solution, it is preferable to find one that is highly compact (the primary objective). In a local search method, starting with a feasible solution, the neighborhood of the solution is explored for any improvement in the objective (Ricca et al., 2013). The neighborhood considered in this paper is obtained by re-assigning a unit from one district to another district. After every such re-assignment, the feasibility of the solution is assessed. For assessing contiguity, an efficient geo-graph method (King et al., 2018, 2015) is used, which can be evaluated in \( O(R(v)) \), where \( R(v) \) is an augmented neighborhood of the re-assigned unit \( v \). The solution that leads to the the maximum decrease in the contiguity is chosen as the starting point for the next iteration, and the local search terminates either when the number of iterations reaches a limit, or when there is no further improvement. The quality of the final solution from the local-search method is typically influenced by the initial solution. In order to diversify the solution space, a multistart local-search procedure is used, where multiple feasible solutions are generated as starting points for local search, and the final solution is the one with the least objective.

### 4.3.2 Evaluating partisan asymmetry for incumbent solutions

As noted in Section 3.3, the partisan asymmetry objective is non-linear, and hence solving the Symmetric-DP presents a computational challenge. To tackle this, the approach is to instead solve Compact-DP without the \( \epsilon \)-constraint for \( \phi_{PA} \), and whenever a new incumbent solution \( \pi \) is found at a node in the branch-and-cut tree, the partisan asymmetry of \( \pi \) is computed at that node; if \( \phi_{PA}(\pi) > \epsilon_{PA} \), \( \pi \) is rejected.

The efficiency of computing \( \phi_{PA} \) of a given district plan relies on the observation that the vote-seat curve (e.g. in Figure 2) follows a step function since the number of seats (districts) is finite. Hence, it is sufficient to find the set of breakpoints in the vote-seat curve, and use them to compute the area between the curve and its symmetric reflection, as illustrated below.

**Computing breakpoints:** A breakpoint in the vote-seat curve for party \( A \) is a particular value of \( A \)'s overall vote-share when \( A \) can win one additional district with a small perturbation. As illustrated by the example in Figure 2, the breakpoints correspond to the vertical shifts in the step functions. Starting with an arbitrary party’s (\( A \)’s) actual overall vote-share \( \{a_k\}_{k \in [K]} \) in the given district plan (i.e. the incumbent solution), the breakpoints in \( A \)'s vote-seat curve are determined by calculating at what values of \( A \)'s overall vote-share (\( \mu \)) each district turns from \( A \) losing the district to winning it, or from winning a district to losing it. At any stage in the algorithm, let \( \mu_k \) be \( A \)'s vote-share in district \( k \in [K] \), and let \( S^- := \{ k \in [K] : \mu_k < 0.5 \} \) be the subset of districts in which \( A \) loses. For the current district vote-shares, the algorithm then picks district \( k^* := \text{arg min}_{k \in S^-} (0.5 - \mu_k) \), which is the district which \( A \) can win with the smallest increase in \( A \)'s vote-share. The difference, \( (0.5 - \mu_{k^*}) \), is added uniformly to \( A \)'s vote-shares in all the districts (including \( k^* \)). Using the updated district vote-shares \( \{\mu_k\}_{k \in [K]} \), the overall vote-share (\( \mu \)) is re-evaluated using Equation (18) and is noted as a breakpoint along with its corresponding seat-share. This procedure continues iteratively, and terminates when \( A \) wins all the districts. Similarly, starting once again with \( A \)'s actual vote-share, its vote-share is iteratively decreased while \( A \) loses one new district in every iteration. This procedure terminates when \( A \) loses all the districts.
Computing $\phi_{PA}$: The breakpoints along with the corresponding seat-shares define the vote-seat curve. The breakpoints for the symmetric reflection (party B’s vote-seat curve) is the set of points obtained by subtracting A’s breakpoints from 1, while their corresponding seat-shares are also obtained by subtracting A’s corresponding seat-shares from 1. The area between the curves can be computed by adding the rectangular areas between the successive breakpoints of A’s and B’s vote-seat curves. Note that this procedure takes $O(K)$ time, where $K$ is the number of districts.

4.4 Un-coarsening

The final stage of the multilevel algorithm comprises of mapping the optimal solution from level $L$ back to level 0. Note that the optimal solution at level $L$ may not be optimal at lower levels, and hence this solution could potentially be improved using a heuristic. Let $z_{l+1} : V_{l+1} \rightarrow [K]$ be the district solution at level $l + 1$. First, an initial district solution at level $l$ is first created, i.e. $z_i(i) = z_{l+1}(\mathcal{C}(i))$ for all $i \in V_l$, where $\mathcal{C}$ is the mapping function defined in the coarsening procedure. In $G_l$, using a steepest-descent local-search method by single unit re-assignments, this solution is improved further with respect to the compactness objective while enforcing contiguity, population balance and the $\epsilon$-constraints.

5 Case Study in Wisconsin

This section presents a case study on congressional districting for the state of Wisconsin. Wisconsin is an ideal choice since both the Democratic (D) and Republican (R) parties have an approximately equal share of voters (51.6 and 48.4% respectively) across the state based on the 2012 and 2016 presidential elections. The population data and adjacency information are obtained from the U.S. Census Bureau (2010), and processed using Quantum-GIS. Wisconsin has $|V| = 1409$ census tracts (units) and $K = 8$ congressional districts with average district population $\bar{P} = 710,873$. At the pre-processing stage, 17 units each with a population of zero (e.g. water bodies) are merged with neighboring units. Even though census blocks are more finely granular, the use of census tracts is justified since fewer than 5% of the census tracts are split between multiple districts as per Wisconsin’s current (2016) district plan. The current congressional district plans have a maximum population deviation of 4.8% from the average, and the same value is set as the population deviation threshold $\tau$. Voter information is compiled as the average of the 2012 and 2016 presidential election results (The Guardian, 2018). This information is available at the precinct level, which is aggregated to the county level, and then distributed to the census tracts proportionally based on their populations. The computations reported in this paper are performed using an 2.7 GHz Intel Core i5 machine with 8 GB memory. The computational times are reported using both clock time (seconds) and CPU time (ticks).

5.1 Coarsening

Starting with the adjacency graph at the finest level (census tracts), the matching-based coarsening procedure proceeds as in Section 4.2. The number of levels is set to four since it was observed that the optimization solver performs well for instance sizes less than 200 units, and four levels of near-perfect matchings would reduce the instance size to satisfy this criterion. Figure 3 depicts the adjacency graphs of the 4 levels of instances produced, with the coarsest layer graph consisting of 116 units.

Based on the level 4 problem instance described by the tuple $(G(V, E), p, p^D, p^R, K)$, three PDPs are formulated, each with their respective secondary fairness objective being efficiency gap, partisan
asymmetry and competitiveness. In the primary compactness objective, the distance $d_{ij}$ between every pair of units $i, j \in V$ is defined to be the number of edges in the shortest path between $i$ and $j$ in $G$. This distance metric has integer objective values and is hence preferred over a continuous distance metric since the exact method performs better when using an integer objective function. For each of the PDPs, the initial solution from the multistart local-search is first computed with the termination criteria being either 5 iterations or a time limit of 15 minutes. IBM’s CPLEX Optimizer 12.6 is used for solving the MIP formulations. The initial solution is passed into the root of the branch-and-cut tree, and a time limit of six hours is set for each run.

5.2 Equitable Districts

The PDP with objective set $\{ \phi_{comp}, \phi_{EG} \}$, formulated by (9),(2)-(8), (12)-(17) and (28), is solved. For each value of $\epsilon_{EG}$ iteratively reduced from 0.5, the MIP is solved either to optimality, or until the time limit is reached. Figure 4a depicts the Pareto-frontier from the solutions obtained, highlighting the trade-off between the compactness and efficiency gap objectives. The computational results for the six solutions found are presented in Section C.1 of the online appendix. The time taken by the solver increases as the value of $\epsilon_{EG}$ decreases, since it becomes increasingly difficult to find good feasible solutions that satisfy the $\epsilon_{EG}$-constraint, in addition to the other constraints. The number of nodes visited in the branch-and-cut (B&C) tree also increases as $\epsilon_{EG}$ is decreased, signifying the influence of the quality of a feasible solution on the extent of exploration in the tree. Among the six solutions forming the frontier, one is an approximate Pareto-optimal solution with an optimality gao of 2%. Note that for an $\epsilon_{EG}$ value less than 0.0089, neither the initial solution heuristic, nor the optimization solver could find a feasible solution within their time limits.

The district plan corresponding to an efficiency gap of 0.89% is depicted in Figure 4b. In comparison, the Supreme Court sets a threshold of 8% above which a district plan is deemed gerrymandered (Royden and Li, 2017). For this district plan, three of the eight districts have a margin less than 5%, party D wins two districts and party R wins three districts with a margin more than 5%. Considering just simple-majorities (with any margin), each party wins exactly four districts. In total, parties D and R waste 712,445 and 738,314 votes respectively. Figure 4c shows the vote-shares between the parties within each of the eight districts. It can be seen that districts 1 and 7 lean strongly D and districts 2 and 6 lean strongly R. With relatively less margin, districts 3 and 8 lean slightly D, while districts 4 and 5 lean slightly R. However, D wastes 0.89% of votes more than R. Additionally, note that this solution’s partisan asymmetry is 0.0423 and its competitiveness (maximum margin) is 19.21%.

5.3 Symmetric Districts

The PDP with objective set $\{ \phi_{comp}, \phi_{PA} \}$, formulated by (9),(2)-(8), is solved, and the feasibility of an incumbent solution (with respect to the $\epsilon_{PA}$ constraint) in the branch-and-bound tree is verified.
using the CPLEX Node Callback functionality. Figure 5a depicts the Pareto-frontier highlighting the relationship between partisan asymmetry and compactness, with the computational details presented in Section C.2 of the online appendix. Note that for an $\epsilon_{PA}$ value less than 0.00783, no feasible solution was found by the solver within the six hour time limit.

Figure 5b depicts a district plan with a partisan asymmetry value of 0.0078. One of the eight districts is 5%—competitive, while party D wins four districts, and party R wins three districts with their margins more than 5%. Figure 5c shows the corresponding vote-seat curve. D’s actual overall vote-share in Wisconsin is 0.51, and D’s corresponding seat-share in this district plan is 0.5. For every other vote-share, it can be seen that the corresponding seat-shares for D is close to the seat-shares for party R. Even though this district plan has the smallest partisan asymmetry value that is found within the time limit, note that the discrete nature of the inputs prohibits from attaining a completely symmetric outcome. Further, this district plan’s efficiency gap is 3.38% (D wastes more voters than R) and its competitiveness is 8.31%.
5.4 Competitive Districts

The PDP with objective set \{\phi_{comp}, \phi_{cmpttv}\}, formulated by (9),(2)-(8),(22)-(27), is solved. The Pareto-optimal solutions representing the trade-off between competitiveness and compactness, are depicted in Figure 6a, with the computational details presented in Section C.3 of the online appendix. As the \epsilon_{cmpttv} value is increased, it takes longer time to solve the problem. Note that among the 22 solutions presented, 19 are optimally compact, 3 are approximately optimal within a 3% optimality gap, and no solution with a margin less than 4% was found within the time limit. Figure 6b depicts the optimally compact district plan with all eight districts with their margins less than 4%, and 6c depicts the corresponding district vote-shares. Further, this district plan’s partisan asymmetry is 0.0291 and its efficiency gap is 20.96% (R wastes more voters than D).

6 Conclusions

Political districting is a problem of national interest with consequences to electoral representation. This paper addresses the issue of gerrymandering by providing a practical multi-objective framework that can be adopted by a mechanism that focuses on political fairness. The approach optimizes a set of fairness metrics that cater to the different facets of fairness. A case study for congressional
(a) Pareto-frontier highlighting the trade-off between the competitiveness and compactness objectives.

(b) Eight competitive districts, each within a margin of 4%.

(c) Bar chart depicting the vote-shares of each party in the plan with a maximum margin of 4%.

Figure 6: The results from solving the PDP with competitiveness and compactness objectives.

districting in Wisconsin highlights that solutions that are optimal with respect to one objective may be poor with respect to the other. In summary, this paper re-frames the redistricting process and serves as a starting point to investigate how political fairness can be integrated into computational approaches for district design. There are several aspects of this framework that future work can address. First, the algorithm relies on the use of historic population and voting data to draw districts for the future. Inculcating the potential variations in voting trends into the mathematical models such as using robust optimization could produce district maps that are more relevant to future needs. Second, the goal of the matching-based coarsening procedure is to coarsen so as to produce a homogeneous population distribution. Other edge weights that explicitly consider the fairness objectives could potentially improve the quality of solutions found, for example merging units with equal but opposing party support in order to produce districts that are competitive. Third, improving the efficiency of exactly solving the PDPs using methods that utilize the structure of the MIPs (such as using decomposition techniques) would yield more solutions within the time limits and hence achieving an exhaustive Pareto-frontier.
References


Appendices

A  Proof for Theorem 1

Theorem 1. For a general graph $G = (V, E)$, unit populations $\{p_i, p_i^A, p_i^B\}_{i \in V}$ and number of districts $K \geq 2$, the Equitable Districting Problem is NP-complete.

Proof. Given a district plan, since feasibility can be verified in polynomial time using MIP constraints (2)-(8) it is clear that Equitable-DP $\in$ NP. To show that equitable districting is NP-Hard, a polynomial-time reduction is shown from a variant of the number partitioning problem introduced in this paper, defined as follows. Given a set of tuples $\{(a_i, b_i)\}_{i \in V}$, the 2-dimensional number partitioning (2-DNP) problem finds a subset $S \subseteq V$ that minimizes $\sum_{i \in S}(|3a_i - b_i|) - \sum_{i \notin S}(|3b_i - a_i|)$. The classical number partitioning problem is a special case of 2-DNP when $a_i = b_i \forall i \in V$ (Garey et al., 1976). Since number partitioning is NP-Complete, so is 2-DNP.

Given an instance of 2-DNP, an instance of Equitable-DP is now constructed. Consider the contiguity-free setting of Equitable-DP (G is a complete graph), and $\tau$ is set to be large enough that the population balance constraints are always satisfied. For each unit $i \in V$, let $p_i = a_i + b_i$, $p_i^A = a_i$, and $p_i^B = b_i$. For any arbitrary number of districts $K$, let $z^* : V \rightarrow [K]$ be an optimal solution to Equitable-DP. Let $S^A(z^*) := \{i \in V : \sum_{j \in V : z^*_j = z^*_i} p_j^A \geq \sum_{j \in V : z^*_j = z^*_i} p_j^B\}$ be the subset of units that are assigned to districts in which party $A$ wins, and $S^B(z^*) := V \setminus S^A(z^*)$. Let $[K]^A := \{k \in [K] : \sum_{j \in V : z^*_j = k} p_j^A \geq \sum_{j \in V : z^*_j = k} p_j^B\}$ be the subset of districts in which party $A$ wins, and $[K]^B := [K] \setminus [K]^A$. Within $[K]^A$, the difference in wasted votes between the two parties ($A$ minus $B$) is given by $\sum_{k \in [K]^A} \left( \sum_{j \in V : z^*_j = k} \frac{p_j^A - p_j^B}{2} - \sum_{j \in V : z^*_j = k} p_j^B \right) = \sum_{k \in [K]^A} \left( \sum_{j \in V : z^*_j = k} \frac{p_j^A - 3p_j^B}{2} \right) = \sum_{j \in S^A(z^*)} \frac{p_j^A - 3p_j^B}{2}$. Similarly within $[K]^B$, the difference in wasted votes between the two parties ($A$ minus $B$) is given by $\left( - \sum_{j \in S^B(z^*)} \frac{p_j^B - 3p_j^A}{2} \right)$. The absolute difference in wasted votes across all districts is given by $\left| \sum_{j \in S^A(z^*)} \frac{p_j^A - 3p_j^B}{2} - \sum_{j \in S^B(z^*)} \frac{p_j^B - 3p_j^A}{2} \right| = \frac{1}{2} \left| \sum_{j \in S^A(z^*)} (3p_j^A - p_j^B) - \sum_{j \in S^B(z^*)} (3p_j^B - p_j^A) \right|$. Note that this quantity is a constant times the objective function of Equitable-DP, and $z^*$ minimizes Equitable-DP. Hence, $S^A(z^*)$ is an optimal solution to 2-DNP. Since the reduction from 2-DNP to Equitable-DP is constructed in polynomial time, Equitable-DP without contiguity and population balance constraints is NP-Complete.

B  Proof for Theorem 2

Theorem 2. For a general graph $G = (V, E)$, unit populations $\{p_i, p_i^A, p_i^B\}_{i \in V}$ and number of districts $K \geq 2$, the Aggregate Competitive Districting Problem (ACDP) is NP-Complete.

Proof. For a given district plan $z$, the feasibility of $z$ can be verified in polynomial time and hence Aggregate Competitive-DP is in NP. Consider an instance of a variant of the NP-hard multi-way partitioning (MWP) problem (Korf, 2010). For a given set of integers $V = \{a_1, a_2, \ldots, a_n\}$ and a positive integer $K < n$, MWP seeks to divide $V$ into $K$ partitions, $\{V_k\}_{k=1}^K$, such that $\max_{a_i \in [K]} \sum_{i \in V_k} a_i$ is minimized. Given an instance of MWP, consider the contiguity-free setting of ACDP ($G$ is a complete graph) and $\tau$ is set to be large enough that the population balance
constraints are always satisfied. The \((p_i, p_i^A, p_i^B)\) values for each unit \(i \in V\) is set to be \((a_i, a_i, 0)\). It is clear that the optimal solution to the ACDP for this instance optimally solves the MWP. Hence, ACDP without contiguity and population balance constraints is NP-complete.

\[\square\]

C Computational Results for Pareto-Frontiers

C.1 Efficiency Gap and Compactness

<table>
<thead>
<tr>
<th>(\phi_{EG})</th>
<th>(\phi_{comp})</th>
<th>Clock time (s)</th>
<th>CPU time (ticks)</th>
<th># B&amp;C nodes</th>
<th>Opt. gap ((UB-LB))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1442</td>
<td>254</td>
<td>1,704</td>
<td>727,467</td>
<td>2,483</td>
<td>-</td>
</tr>
<tr>
<td>0.0225</td>
<td>256</td>
<td>3,379</td>
<td>1,476,713</td>
<td>2,904</td>
<td>-</td>
</tr>
<tr>
<td>0.0201</td>
<td>257</td>
<td>8,345</td>
<td>3,673,589</td>
<td>18,728</td>
<td>-</td>
</tr>
<tr>
<td>0.0108</td>
<td>259</td>
<td>4,946</td>
<td>1,776,842</td>
<td>15,375</td>
<td>-</td>
</tr>
<tr>
<td>0.0105</td>
<td>261</td>
<td>7,460</td>
<td>2,862,877</td>
<td>31,533</td>
<td>-</td>
</tr>
<tr>
<td>0.0089</td>
<td>262</td>
<td>21,600</td>
<td>9,249,306</td>
<td>8,3752</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 1: Computational results from the \(\epsilon\)-constraint method to obtain the Pareto-frontier between the efficiency gap and compactness

C.2 Partisan Asymmetry and Compactness

<table>
<thead>
<tr>
<th>(\phi_{PA})</th>
<th>(\phi_{comp})</th>
<th>Clock time (s)</th>
<th>CPU time (ticks)</th>
<th># B&amp;C nodes</th>
<th>Opt. gap ((UB-LB))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0549</td>
<td>254</td>
<td>110</td>
<td>29,334</td>
<td>1,524</td>
<td>-</td>
</tr>
<tr>
<td>0.0481</td>
<td>255</td>
<td>146</td>
<td>50,641</td>
<td>1,980</td>
<td>-</td>
</tr>
<tr>
<td>0.0191</td>
<td>256</td>
<td>20,778</td>
<td>21,198,410</td>
<td>122,254</td>
<td>-</td>
</tr>
<tr>
<td>0.0088</td>
<td>295</td>
<td>21,600</td>
<td>17,665,126</td>
<td>102,364</td>
<td>0.13</td>
</tr>
<tr>
<td>0.0083</td>
<td>304</td>
<td>21,600</td>
<td>19,617,605</td>
<td>119,614</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 2: Computational results from the \(\epsilon\)-constraint method to obtain the Pareto-frontier between the partisan asymmetry and compactness objectives
C.3 Competitiveness and Compactness

<table>
<thead>
<tr>
<th>$\phi_{comp_{te}}$</th>
<th>$\phi_{comp}$</th>
<th>Clock time (s)</th>
<th>CPU time (ticks)</th>
<th># B&amp;C nodes</th>
<th>Opt. gap ($\frac{UB-LB}{LB}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.164</td>
<td>254</td>
<td>155</td>
<td>91,757</td>
<td>199</td>
<td>-</td>
</tr>
<tr>
<td>0.146</td>
<td>256</td>
<td>187</td>
<td>74,967</td>
<td>302</td>
<td>-</td>
</tr>
<tr>
<td>0.145</td>
<td>257</td>
<td>94</td>
<td>46,078</td>
<td>590</td>
<td>-</td>
</tr>
<tr>
<td>0.135</td>
<td>259</td>
<td>101</td>
<td>48,605</td>
<td>746</td>
<td>-</td>
</tr>
<tr>
<td>0.118</td>
<td>261</td>
<td>98</td>
<td>46,178</td>
<td>1,145</td>
<td>-</td>
</tr>
<tr>
<td>0.113</td>
<td>264</td>
<td>301</td>
<td>147,361</td>
<td>2,821</td>
<td>-</td>
</tr>
<tr>
<td>0.103</td>
<td>265</td>
<td>165</td>
<td>75,094</td>
<td>2,745</td>
<td>-</td>
</tr>
<tr>
<td>0.086</td>
<td>266</td>
<td>568</td>
<td>235,878</td>
<td>3,061</td>
<td>-</td>
</tr>
<tr>
<td>0.080</td>
<td>269</td>
<td>531</td>
<td>243,433</td>
<td>4,828</td>
<td>-</td>
</tr>
<tr>
<td>0.076</td>
<td>272</td>
<td>150</td>
<td>69,536</td>
<td>3,032</td>
<td>-</td>
</tr>
<tr>
<td>0.073</td>
<td>273</td>
<td>722</td>
<td>354,084</td>
<td>6,700</td>
<td>-</td>
</tr>
<tr>
<td>0.071</td>
<td>275</td>
<td>948</td>
<td>418,707</td>
<td>7,019</td>
<td>-</td>
</tr>
<tr>
<td>0.069</td>
<td>276</td>
<td>11,627</td>
<td>5,306,129</td>
<td>87,302</td>
<td>-</td>
</tr>
<tr>
<td>0.065</td>
<td>277</td>
<td>1,684</td>
<td>760,658</td>
<td>9,874</td>
<td>-</td>
</tr>
<tr>
<td>0.064</td>
<td>279</td>
<td>2,194</td>
<td>832,865</td>
<td>15,135</td>
<td>-</td>
</tr>
<tr>
<td>0.058</td>
<td>281</td>
<td>4,370</td>
<td>2,094,608</td>
<td>24,380</td>
<td>-</td>
</tr>
<tr>
<td>0.055</td>
<td>284</td>
<td>21,124</td>
<td>8,282,561</td>
<td>103,807</td>
<td>-</td>
</tr>
<tr>
<td>0.052</td>
<td>287</td>
<td>4,029</td>
<td>1,607,074</td>
<td>18,410</td>
<td>-</td>
</tr>
<tr>
<td>0.049</td>
<td>289</td>
<td>13,132</td>
<td>5,828,313</td>
<td>89,599</td>
<td>-</td>
</tr>
<tr>
<td>0.045</td>
<td>292</td>
<td>21,600</td>
<td>8,852,710</td>
<td>81,752</td>
<td>0.02</td>
</tr>
<tr>
<td>0.042</td>
<td>295</td>
<td>21,600</td>
<td>8,202,587</td>
<td>56,001</td>
<td>0.03</td>
</tr>
<tr>
<td>0.040</td>
<td>300</td>
<td>21,600</td>
<td>7,717,568</td>
<td>55,829</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 3: Computational results from the $\epsilon-$constraint method to obtain the Pareto-frontier between competitiveness and compactness objectives