Bounding the number and the diameter of optimal compact Black-majority districts

Samuel Kroger[1], Hamidreza Validi[2], Illya V. Hicks[1], and Tyler Perini[3]

[1]Department of Computational Applied Mathematics and Operations Research, Rice University
[2]Department of Industrial, Manufacturing & Systems Engineering, Texas Tech University
[3]Mathematics Department, United States Naval Academy

August 12, 2024

Abstract

Section 2 of the Voting Rights Act (VRA) prohibits voting practices that minimize or cancel out minority voting strength. While this section provides no clear framework for avoiding minority vote dilution and creating minority-majority districts, the Supreme Court proposed the *Gingles* test in the 1986 case *Thornberg v Gingles*. Mathematical optimization models are increasingly employed to analyze the first prong of the Gingles test: compactness and numerosity. This paper proposes mixed integer programming (MIP) formulations and techniques to explore the maximum number of Black-majority congressional districts for multiple states of the United States. Furthermore, we generalize the diameter-based compactness criterion of Garfinkel and Nemhauser (*Management Science*, 1970) and provide a framework for optimizers to capture compactness in constraints rather than the objective function. To alleviate the solving process, we propose fixing procedures and symmetry-breaking constraints. Our proposed MIP formulations provide (i) an upper bound on the number of Black-majority districts and (ii) lower bounds for the diameter of districts. We finally run a state-of-the-art districting package, GerryChain, to provide feasible bounds for the optimum number of Black-majority districts and their diameters. This provides the best existing optimality gaps that can be closed by both districters and optimizers in the future.

Keywords: political redistricting; diameter-bounded districts; Black-majority districts; mixed integer programming;

1 Introduction

The earliest legal debates surrounding Alabama's redistricting process are dated decades ago. In July 1957, the City of Tuskegee's borders were redrawn so that African-Americans were excluded from the legal borders of the city (See Figure 1). This led to the case *Gomillion v Lightfoot*, in which Charles Gomillion claimed that the new city boundaries were unconstitutional. This case and other related issues ignited the Selma to Montgomery protest in March of 1965. These marches from the Civil Rights Era resulted in the Voting Rights Act (VRA) (Duchin and Walch, 2021). Section 2 of the Voting Rights Act (VRA) prohibits voting practices that minimize or cancel out minority voting strength. While the section provides no clear framework for avoiding minority vote dilution and creating minority-opportunity districts, the Supreme Court proposed the *Gingles* test during the *Thornberg v Gingles* case in 1986. Motivated by the recent amicus curiae brief of Hirsch et al. (2022), which suggests mathematical optimization, not sampling methods, for finding remedial maps to enforce VRA, we propose multiple mixed integer programming (MIP) formulations to explore the possibility of creating Black-majority districts respecting the first prong of Gingles: compactness and numerosity.

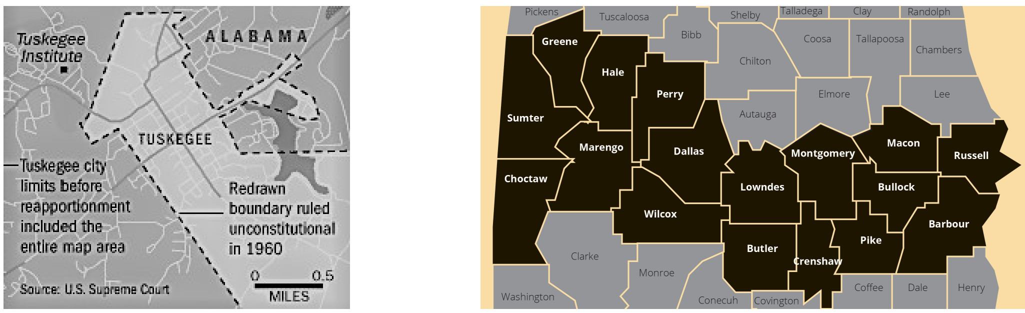

Figure 1: (Left): The redrawn map of Tuskegee in Macon county of Alabama shows the exclusion of Tuskegee Institute (one of the Historically Black Colleges and Universities) from the new boundaries; (Right): the traditional black belt of Alabama (The Alabama Bicentennial Commission, 2024).

In 2020, 28% of Alabama's population was African-American, despite this, the congressional districting plan included only one Black-majority district out of Alabama's seven congressional districts. In November 2021, multiple parties argued that the 2020 districting plan suppressed Black votes and did not comply with the VRA. On January 24th, 2022, a three-judge court unanimously ruled in favor of the plaintiff and found the 2020 districting plan unconstitutional as it violates the VRA. The case was then appealed to the Supreme Court. On June 8th, 2023, the Supreme Court, in a 5-4 decision, agreed the districting plan violated the VRA and required the state of Alabama to form an additional Black-majority district. Ignoring this ruling, Alabama proposed another districting plan with one Black-majority district, which was immediately rejected by a federal three-judge panel. The federal court has decided to appoint an independent expert to draw the new districts instead of letting the state legislature make another attempt to form a two-district plan. This is only a snapshot of the legal proceedings from 2020 to 2023.

Political redistricting is the decennial process of redrawing boundaries of US political districts in which the following essential criteria are respected: (i) population balance, (ii) contiguity, and (iii) compactness. Furthermore, the Gingles test states that a minority-majority district can be created if at least the following conditions (called "Gingles prongs") are satisfied: (i) compactness and numerosity (e.g., if the minority group makes up more than 50% of the citizen voting-age population), (ii) political cohesion, and (iii) sufficiency of White majority votes to defeat the preferred candidate of minorities (Hebert et al., 2010). This paper explores MIP approaches to capture the first prong of Gingles (i.e., compactness and numerosity.) While Lublin et al. (2020) specify that a district can be considered a minority-majority district with only 40 to 50 percent of total votes, we consider a threshold of 50% for numerosity purposes following *Bartlett v Strickland* case in which the Court held that only reaching the 50% threshold can qualify a minority population to form a minority-majority district (Hebert et al., 2010).

Since 1965, operations researchers have started developing mathematical optimization models to capture three basic criteria of redistricting: (i) population balance, (ii) contiguity, and (iii) compactness. Compactness is usually considered in the objective function of optimization formulations of the districting problems. However, Garfinkel and Nemhauser (1970) incorporate compactness as a constraint for the first time in their redistricting MIP formulation. They claim that compactness is not generally as important as other redistricting criteria to be incorporated in the objective function. Instead, they minimize the maximum population deviation from the mean population in their optimization problem. This is consistent with Duchin and Walch (2021) that believe "compactness is over-emphasized as a cure for gerrymandering." Furthermore, Gurnee and Shmoys (2021) emphasize that "fairness is a far more important consideration, and therefore we want to encourage practitioners to treat compactness as a constraint rather than an objective."

Garfinkel and Nemhauser (1970) incorporate the following distance-based measure for creating compact districts as a constraint in their MIP formulation with exponentially many variables: two land parcels cannot belong to the same district if their distance is more than a threshold, say s. We revisit the distance-based compactness criterion of Garfinkel and Nemhauser (1970) and propose a cut-based MIP reformulation (with exponentially many constraints) of their formulation (with exponentially many variables). We also propose procedures and algorithms for finding all the possible values of s for states with Black-majority opportunity districts. Similarly, Mehrotra et al. (1998) employs radius as the compactness measure in their heuristic branch-and-price framework. Specifically, they eliminate the districting plans for which the distance of a land parcel from the center of its district is greater than three.

To assess the appropriateness of the distance-based compactness measure, we compare it with other measures based on five compactness criteria of Young (1988):

1. Data availability and simplicity;
2. Threshold: there is no specific numerical threshold for determining a compact districting plan;
3. Shape: the shape of land parcels should not impact a compactness measure;
4. Size: the population of a district should not impact a compactness measure;
5. Boundaries: a compactness measure should consider the shape of a state's boundaries.

Table 1 summarizes the five metrics of Young (1988) for (i) the population-distance measures that are widely employed by operations researchers (e.g., moment of inertia (Hess et al., 1965; Swamy et al., 2019; Validi et al., 2022) and total population flow (Zhang et al., 2024b)), (ii) cut edges (Becker and Solomon, 2021; Validi and Buchanan, 2022; Shahmizad and Buchanan, 2023; Ludden et al., 2023), (iii) Polsby-Popper (Belotti et al., 2023), and (iv) diameter/radius of a district (Garfinkel and Nemhauser, 1970; Mehrotra et al., 1998). Table 1 shows that the diameter/radius measure satisfies at least four criteria of compactness proposed by Young (1988).

As we consider compactness as a set of constraints, we can optimize an objective function rather than compactness. This paper explores the maximum number of Black-majority districts in a state with respect to different feasible values of s. We also propose multiple fixing procedures and a set of symmetry-breaking constraints to alleviate the running time of the proposed MIP formulation. We finally break the proposed optimization MIP formulation into multiple feasibility MIP formulations to increase the lower bounds for the s values. We have no claim on the appropriateness of the generated maps in this paper. Furthermore, we focus on only the Black population as a

Table 1: Comparisons between compactness measures employed in the operations research literature based on the criteria of Young (1988).

Measures	Criteria				
	Data & Simplicitity	Threshold	Shape	Size	Boundaries
Population-distance	✓	✓	✓	×	×
Number of cut edges	✓	✓	✓	✓	×
Polsby-Popper	×	✓	✓	✓	✓
Diameter/radius	✓	✓	✓	✓	×

minority group in this paper. Our models can be easily employed for other minority groups. Our contributions are summarized as follows:

1. We propose a new cut-based MIP formulation that captures connectivity and compactness in constraints and lets districters optimize other districting measures in the objective function;
2. We provide valid upper bounds and lower bounds for the optimum number of Black-majority districts and the optimum diameter of districts, respectively;
3. We propose a set of variable fixing procedures and a set of symmetry-breaking constraints to alleviate the solving process of the proposed MIP;
4. We improve the proposed lower bound of the diameter by running a set of feasibility variants of the proposed MIP formulation; and
5. We run a state-of-the-art districter, i.e., GerryChain, to generate feasible maps that, along with the bounds in item 4, provide the best existing optimality gaps for the number of Black-majority districts and their diameters.

Outline. Section 2 provides a background on the political redistricting problem and optimization methods employed for solving different variants of it. Section 3 explains the notation, data, and computational setup on which we run our computational experiments. Section 4 proposes a new cut-based MIP formulation that maximizes the number of Black-majority districts in objective function while capturing connectivity and compactness in constraints. Section 5 provides procedures for finding an upper bound on the optimum number of Black-majority districts and a lower bound on the optimum diameter of a district in a state. Section 6 proposes computational enhancements for the MIP formulation in Section 4 and employs the formulation to improve the proposed dimeter's lower bound in Section 5. Section 7 reports feasible maps by running a state-of-the-art non-exact redistricting solver (i.e., GerryChain) and provides feasible lower bounds and upper bounds for the number of Black-majority districts and the diameter of districts, respectively. We conclude the paper in Section 8.

2 Background

Mathematical models and MIP technology have been employed in designing political districts since the 1960s, beginning the era of computational redistricting. Weaver and Hess (1963) provide an iterative framework to heuristically generate solutions for the problem, which respects the basic redistricting criteria (i.e., population balance, contiguity, and compactness). Nagel (1964) proposes a procedure for solving the bipartisan redistricting problem; they clarify that their procedure differs considerably from that of Weaver and Hess (1963) as they simultaneously consider bipartisan gerrymandering, population balance, and compactness. Their heuristic procedure also guarantees contiguity, and the user may further specify specific land parcels to keep intact. Hess et al. (1965) develop a MIP formulation for the nonpartisan districting problem in which population balance and compactness are captured simultaneously; however, contiguity is imposed in a post-processing step. Their formulation—coded in FORTRAN IV, solved on an IBM 7040 machine—provided more compact districts for Delaware (Senate and House of Representatives) than an enacted plan.

Garfinkel and Nemhauser (1970) consider the pairwise distance of land parcels as a compactness measure (i.e., "distance compactness") in an implicit enumeration framework that minimizes the maximum population deviation. Instances with at most 40 counties were solvable in under 10 minutes on an IBM 7094 machine, but an instance with 55 counties was intractable at the time. Similarly, Mehrotra et al. (1998) employ the idea of bounding the radius of districts to three to generate compact maps in a heuristic-based column generation framework. They provide a case study for South Carolina, consisting of 46 counties assigned to six congressional districts, where optimization required less than 5 minutes.

The redistricting optimization models are usually defined on the contiguity graph (Ricca et al., 2013) in which the vertex set corresponds to land parcels (e.g., counties and tracts), and its edge set represents the adjacency of vertices. To capture the diameter of districts, one can think of partitioning the contiguity graph of a state into low-diameter clusters, known as the s-club partitioning problem (Deogun et al., 1997). The diameter of a graph is defined as the maximum length of the shortest paths between any pair of vertices in the graph. Figure 2 illustrates a graph partitioned into subgraphs with diameters of at most two and three. In the minimum s-club partitioning problem, one seeks to find the minimum number of parts such that the diameters of their induced subgraphs are at most s. The problem is NP-hard due to the clustering problem being hard. Yezerska et al. (2019) propose a combinatorial branch-and-bound approach for solving the minimum s-club partitioning problem. Furthermore, Gschwind et al. (2021) develop a branch-and-price framework for solving the problem. Recently, Zhang et al. (2024a) propose new MIP formulations and techniques that solve the 2-club and 3-club partitioning problems on most of the benchmark instances of Yezerska et al. (2019) and Gschwind et al. (2021) to optimality.

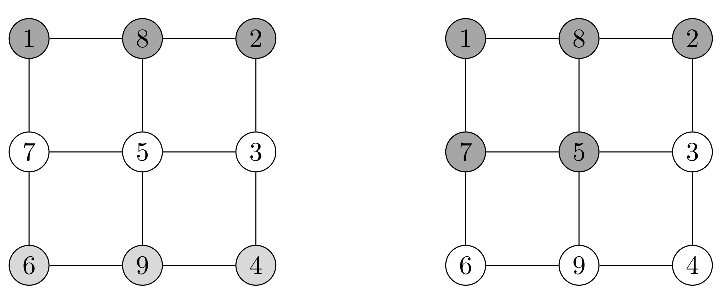

Figure 2: (Left) an optimal 2-club partitioning with size 3; (Right) an optimal 3-club partitioning with size 2.

While compactness is a crucial remedy for racial gerrymandering, identifying potential minority-majority districts is paramount for the formation of plans that comply with Section 2 of the VRA. So, one might consider both compactness and minority-majority districts as objective functions of a MIP formulation. Arredondo et al. (2021) propose a two-phase MIP framework for solving the problem with distinct objectives: (i) phase one constructs the minority districts via a MIP that considers the upper bound of the population balance in the objective function, and (ii) phase two completes the districting plan through a MIP with a nonlinear objective function that optimizes compactness. Recently, Belotti et al. (2023) propose mathematical optimization formulations that find compact minority-majority districts. They optimize compactness (with respect to the Polsby-Popper metric) in the objective function and capture the numerosity of minority groups in the constraints. Furthermore, Cannon et al. (2023) propose a modified Markov chain approach for maximizing the number of minority-majority districts. They test their approach for the Louisiana House of Representatives.

There are other important redistricting metrics that are studied from the lens of operations research in the literature: the number of county splits (Birge, 1983; Shahmizad and Buchanan, 2023; Buchanan et al., 2024), population deviation (Garfinkel and Nemhauser, 1970), efficiency gap, partisan asymmetry, and competitiveness (Swamy et al., 2023; Dobbs et al., 2024). We note that these measures can be easily incorporated into the objective function of our proposed diameter-based compactness setting. Furthermore, there are multiple non-exact methods and approaches for finding feasible solutions to the redistricting problems: Markov chain Monte Carlo (MCMC) (DeFord et al., 2020, 2021; Dobbs et al., 2023), local search algorithms (Ricca and Simeone, 2008) and metaheuristics (Bozkaya et al., 2003; Tomczyk and Kadziński, 2024).

3 Notation, Data and Computational Setup

In this paper, we define contiguity graph $G = (V, E)$ with vertex set V that represents the set of land parcels (e.g., county and tracts) and edge set E that represents the set of adjacent pairs of land parcels. We also define n and m as the sizes of vertex set V and edge set E, respectively. Furthermore, k represents the number of districts in a state, which is calculated by the apportionment after each census. For every land parcel $v \in V$, p_v represents the total population of the land parcel. We also use p_v^{VAP} and p_v^{BVAP} as the total voting-age population and the total Black voting-age population for every land parcel $v \in V$, respectively. We also define $f \in (0, 1)$ as the fraction threshold for forming a Black-majority district. Throughout the paper, we set $[k] := \{1, \dots, k\}$ and $\binom{V}{2}$ as the set of all pairs of vertices. For every vertex subset $S \subseteq V$, $G[S]$ denotes the subgraph induced by S. For every pair of vertices $\{u, v\} \in \binom{V}{2}$, we define $\text{dist}_G(u, v)$ as the length of a shortest path between u and v in contiguity graph G. Then, the diameter of contiguity graph G is defined as $\text{diam}(G) := \max\limits_{\{u,v\} \in \binom{V}{2}} \{\text{dist}_G(u, v)\}$.

We run our computational experiments on 17 interesting county-level and tract-level instances that are summarized in Table 2. Appendix A provides the excluded instances and the reasoning behind them. We employ the 2020 U.S. Census data that is processed and shared with us by Daryl DeFord. The data and code are available at our GitHub repository: `https://github.com/samuel-kroger/Bounding-the-number-and-the-diameter-of-optimal-compact-Black-majority-districts`. All computational results are generated by a machine running Red Hat Enterprise Linux Workstation x64 version 7.6 with an Intel(R) Core(TM) i7-9800X CPU (3.8Ghz, 19.25MB, 165W) using 1 core with 32GB RAM. All codes are written in Python, and we employ

Gurobi 11.0.2 and GerryChain 0.3.1 to solve the MIP models and find feasible redistricting plans. We also employ GerryChain's short bursts module to create feasible districting maps with the largest number of Black-majority districts. The version that we used is available at `https://github.com/AustinLBuchanan/short_bursts`.

Table 2: Black-majority redistricting instances at county and tract levels. The last three columns show the diameter of the contiguity graph, the current percentage of the Black population, and the number of Black-majority districts in the 2020 enacted plans, respectively.

state	parcel	k	n	m	state diam.	% Black pop.	# Black distr.
MS	county	4	82	202	13	35.25	1
MS	tract	4	878	2,378	32	35.25	1
LA	tract	6	1,388	3,861	32	30.07	1
SC	tract	7	1,323	3,677	32	24.20	1
AL	tract	7	1,437	4,014	34	25.06	1
MD	tract	8	1,475	3,993	42	29.16	2
MO	tract	8	1,654	4,488	38	10.85	0
TN	tract	9	1,701	4,691	53	15.18	1
VA	tract	11	2,198	6,064	45	18.41	0
NJ	tract	12	2,181	6,061	45	12.86	0
MI	tract	13	3,017	7,989	41	13.03	2
NC	tract	14	2,672	7,422	52	20.10	0
GA	tract	14	2,796	7,762	43	30.27	4
OH	tract	15	3,168	8,747	44	11.79	1
IL	tract	17	3,265	8,728	52	13.74	2
PA	tract	17	3,446	9,641	59	10.34	1
TX	tract	38	6,896	18,554	49	12.02	0

4 A Cut-based MIP Formulation

This section provides a MIP formulation for maximizing the number of Black-majority districts and partitioning the land parcels into connected and compact clusters such that the population balance (i.e., "one person, one vote" principle) is respected. Due to the successful performance of the labeling formulation in the districting context and its small size (e.g., see Validi and Buchanan (2022)), we employ a labeling assignment model to formulate the Black-majority districting problem. For every district $j \in \{1, \dots, k\}$, binary decision variable z_j (w_j) is one if district j is a Black-majority (non-Black-majority) district. For every district $j \in \{1, \dots, k\}$ and every land parcel $v \in V$, binary decision variable x_{vj} (y_{vj}) is one if land parcel v is assigned to Black-majority (non-Black-majority) district j.

By maximizing the following objective function, we seek to explore the maximum possible number of Black-majority districts. We emphasize that maximizing the number of Black-majority districts is not necessarily an appropriate objective in practice; however, the main goal of this paper is *exploring* the maximum number of Black-majority districts with respect to basic criteria of political redistricting.

$$\max \sum_{j=1}^{k} z_j. \tag{1a}$$

The following constraints partition the land parcels into k non-empty districts.

$$\sum_{j=1}^{k}(x_{vj} + y_{vj}) = 1 \qquad \forall v \in V \tag{1b}$$

$$z_j + w_j = 1 \qquad \forall j \in [k] \tag{1c}$$

$$z_j \leq \sum_{v \in V} x_{vj} \qquad \forall j \in [k] \tag{1d}$$

$$w_j \leq \sum_{v \in V} y_{vj} \qquad \forall j \in [k] \tag{1e}$$

$$x, y \in \{0,1\}^{n \times k}; z, w \in \{0,1\}^k. \tag{1f}$$

Constraints (1b) imply that each land parcel must belong to exactly one district. Constraints (1c) imply that every district is either a Black-majority or a non-Black-majority district. Constraints (1d) (constraints (1e)) imply that if a district is selected as a Black-majority (non-Black-majority) district, then it must contain at least one land parcel. Let $\mathcal{P}_0$ be the set of nonnegative and continuous points that satisfy constraints (1b)-(1e). Then, the following lemma shows that $\mathcal{P}_0$ is an integral polytope.

Lemma 1. *Let* $\mathcal{P}_0 := \{(x,y,z,w) \in \mathbb{R}_+^{(nk)^2 \times k^2} \mid (x,y,z,w)$ *satisfies constraints* (1b)-(1e)$\}$. *Then,* $\mathcal{P}_0$ *is an integral polytope.*

Proof. The proof follows by the fact that the coefficient matrix of constraints (1b)-(1e) is totally unimodular. In particular, (i) the matrix contains elements that only belong to the set $\{-1, 0, 1\}$; (ii) each column of the matrix has at most two non-zero elements; and (iii) the rows of the matrix can be partitioned into two sets such that the non-zero elements of every vector belong to different parts. The result follows from Corollary 2.8 of Wolsey and Nemhauser (1999). □

Remark 1. *The integrality property in Lemma 1 is preserved even if there are more than two types of districts (e.g., Asian-, Hispanic-, or Native-American-majority districts).*

The labeling model (1b)-(1e) suffers from symmetry. The following symmetry-breaking constraints are usually added to labeling formulations to alleviate the symmetry issue.

$$z_k \leq z_{k-1} \leq \cdots \leq z_2 \leq z_1. \tag{1g}$$

Constraints (1g) imply that the $(i+1)$-th Black-majority (non-Black-majority) district cannot be formed if the i-th Black-majority (non-Black-majority) district is not created, where $i \in [k-1]$. Remark 2 presents a set of implied symmetry constraints. In Section 6.2, we present further symmetry-breaking constraints.

Remark 2. *Constraints* $w_1 \leq \cdots \leq w_{k-1} \leq w_k$ *are implied by constraints* (1c) *and* (1g).

Proof. Let $(\hat{x}, \hat{y}, \hat{w}, \hat{z})$ be a point that satisfies constraints (1b)-(1g). By constraints (1c) and (1g), we have

$$1 - \hat{w}_k \leq 1 - \hat{w}_{k-1} \leq \cdots \leq 1 - \hat{w}_2 \leq 1 - \hat{w}_1.$$

So, we have $\hat{w}_1 \leq \cdots \leq \hat{w}_{k-1} \leq \hat{w}_k$. □

The following lemma shows that adding the symmetry-breaking constraints (1g) to constraints (1b)-(1e) maintains the integrality property.

Lemma 2. *Let* $\mathcal{P}_1 := \{(x, y, z, w) \in \mathbb{R}_+^{(nk)^2 \times k^2} \mid (x, y, z, w)$ *satisfies constraints* (1b)-(1e) *and* (1g)$\}$. *Then,* $\mathcal{P}_1$ *is an integral polytope.*

Proof. By the contradiction, suppose there is a fractional extreme point $\hat{q} = (\hat{x}, \hat{y}, \hat{z}, \hat{w})$ of $\mathcal{P}_1$. However, $\hat{q}$ can be written as a convex combination of a subset of the extreme points of the convex hull $\mathcal{P}_0$ that respect the orderings in (1g). This contradicts the definition of $\hat{q}$ as an extreme point. □

Population balance requirements are federally mandated for all states. According to the Apportionment Clause of Article I of Section 2, a state's districts must have a roughly equal population. While the law does not specify an exact population deviation, one percent is typically utilized in the computational redistricting literature (Altman and McDonald, 2018). For every land parcel $v \in V$, we recall that p_v denotes its population. Then, we define the following lower and upper bounds for the population of each district with respect to the one percent deviation.

$$L := \left\lceil \frac{\sum_{v \in V} p_v}{k} \times 0.995 \right\rceil \quad \text{and} \quad U := \left\lfloor \frac{\sum_{v \in V} p_v}{k} \times 1.005 \right\rfloor.$$

The following constraints capture the population lower-bounds for the Black-majority and non-Black-majority districts.

$$Lz_j \leq \sum_{v \in V} p_v x_{vj} \qquad \forall j \in [k] \tag{1h}$$

$$Lw_j \leq \sum_{v \in V} p_v y_{vj} \qquad \forall j \in [k]. \tag{1i}$$

Lemma 3. *The population lower-bound inequalities* (1h)-(1i) *are implied by inequalities* (1d)-(1e) *if* $p_v/L \geq 1$ *for every land parcel* $v \in V$.

Proof. Let $(\hat{x}, \hat{y}, \hat{z}, \hat{w}) \in \mathbb{R}_+^{(nk)^2 \times k^2}$ be a nonnegative point that satisfies constraints (1b)-(1e) and the population lower-bound inequalities. For every district $j \in [k]$, we have

$$\hat{z}_j \leq \sum_{v \in V} \hat{x}_{vj} \leq \sum_{v \in V} (p_v/L)\hat{x}_{vj},$$

$$\hat{w}_j \leq \sum_{v \in V} \hat{y}_{vj} \leq \sum_{v \in V} (p_v/L)\hat{y}_{vj}.$$

Here, the first inequality holds by constraints (1d) and constraints (1e). The second inequality holds by the assumption. □

Theorem 1. *Let*

$$\mathcal{R} := \{(x, y, z, w) \in \mathbb{R}_+^{(nk)^2 \times k^2} \mid (x, y, z, w) \text{ satisfies constraints (1b)-(1g) and (1h)-(1i)}\}.$$

Then, $\mathcal{R}$ is an integral polytope if $p_v/L \geq 1$ for every land parcel $v \in V$.

Proof. The proof follows by Lemmata 1, 2 and 3. □

Constraints (1j)-(1k) impose the population upper-bounds for the Black-majority redistricting problem.

$$\sum_{v \in V} p_v x_{vj} \leq U z_j \qquad \forall j \in [k] \tag{1j}$$

$$\sum_{v \in V} p_v y_{vj} \leq U w_j \qquad \forall j \in [k]. \tag{1k}$$

Unfortunately, we lose the integrality of the formulation after adding the population upper-bound inequalities. This is consistent with the hardness of the redistricting problem due to the population balance constraints (Validi and Buchanan, 2022).

For every land parcel $v \in V$, we recall that p_v^{VAP} and p_v^{BVAP} denote the voting-age population and Black voting-age population, respectively. We also recall that $f \in (0, 1)$ is the fraction threshold for forming a Black-opportunity district, varying between 0.40 to 0.50 (Lublin et al., 2020). In this paper, we fix the value of f to 0.5 as we seek to find Black-majority (not Black-opportunity) districts. Constraints (1l)-(1m) capture the "numerosity" criterion of the Gingles prongs for Black-majority districts. While Lawless and Günlük (2024) employ big-M to capture the numerosity of minorities in their MIP formulation (see their constraints (3)), we note that employing both x and y variables in the MIP formulation (1) enables us to eliminate the necessity of using big-M to capture the following numerosity constraints.

$$\sum_{v \in V} p_v^{\text{BVAP}} x_{vj} \geq f \sum_{v \in V} p_v^{\text{VAP}} x_{vj} \qquad \forall j \in [k] \tag{1l}$$

$$\sum_{v \in V} (p_v^{\text{VAP}} - p_v^{\text{BVAP}}) y_{vj} \geq (1 - f) \sum_{v \in V} p_v^{\text{VAP}} y_{vj} \qquad \forall j \in [k]. \tag{1m}$$

Contiguity and compactness are state-based criteria required for congressional redistricting by 23 and 18 states, respectively. In operations research, contiguity is usually imposed by a set of either flow variables and constraints (Shirabe, 2005, 2009) or cutting planes (Oehrlein and Haunert, 2017). In this paper, we employ specific cutting plains that simultaneously capture contiguity and compactness. The cutting planes are called length-bounded separator cuts. We provide a definition of length-bounded a, b-separators as follows.

Definition 1 (Length-s a, b-separator (Salemi and Buchanan, 2020))**.** *A vertex subset $C \subseteq V \setminus \{a, b\}$ in a graph $G = (V, E)$ is called a* length-s a, b-separator *if the distance between vertices a and b in the graph $G - C$ is greater than s.*

Intuitively, an a, b-separator is a set of vertices whose removal eliminates all paths from a to b, whereas a length-s a, b-separator eliminates paths of length $1, 2, \ldots, s$. Figure 3 provides an illustration for a, b-separators and length-s a, b-separators.

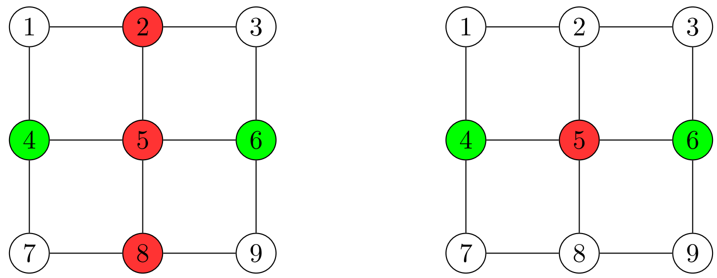

Figure 3: (Left): $\{2, 5, 8\}$ is a 4,6-separator; (Right): $\{5\}$ is a length-3 4,6-separator.

Given s as the maximum allowable pairwise distance in a district's contiguity graph, we provide the following connectivity and compactness constraints. Constraints (1n) (constraints (1o)) imply that if land parcels a and b are assigned to a Black-majority (non-Black-majority) district, then at least one land parcel from every length-s a,b-separator must also belong to the Black-majority (non-Black-majority) district. Similar to Salemi and Buchanan (2020), we use (a, b, C) as a shorthand to denote distinct vertex pair $\{a, b\} \in \binom{V}{2}$ and all length-s a,b-separators $C \subseteq V$.

$$x_{aj} + x_{bj} \leq z_j + \sum_{c \in C} x_{cj} \qquad \forall (a, b, C),\ \forall j \in [k] \tag{1n}$$

$$y_{aj} + y_{bj} \leq w_j + \sum_{c \in C} y_{cj} \qquad \forall (a, b, C),\ \forall j \in [k]. \tag{1o}$$

The complexity of integer separation for constraints (1n)**-**(1o)**.** Since there are exponentially many constraints of types (1n) and (1o), they need to be added on-the-fly. Algorithm 1 provides a polytime integer separation procedure for generating these inequalities.

Algorithm 1 IntegerSeparation$(G, s, k, (x^*, y^*, z^*, w^*))$

for $j \in [k]$ with $z_j^* = 1$ or $w_j^* = 1$ **do**
  define $V_j := \{i \in V \mid x_{ij}^* = 1\}$ and $G_j := G[V_j]$
  let $b \in V_j$ be a vertex in the smallest component of G_j
  solve the single source shortest path problem in G_j with source vertex b
  for each component G' of G_j that does not contain vertex b **do**
    let $a \in V(G')$ with maximum $\text{dist}_G(a, b)$
    let C be the minimal a,b-separator by Fischetti et al. (2017) (in $O(m)$)
    minimalize C (in $O(mn)$)
    add cut $x_{aj} + x_{bj} \leq z_j + \sum_{c \in C} x_{cj}$
    add cut $y_{aj} + y_{bj} \leq w_j + \sum_{c \in C} y_{cj}$
  let G_b be the component that contains vertex b
  let $\{u, v\}$ be a vertex pair with the largest $\text{dist}_{G_b}(u, v)$ and $\text{dist}_{G_b}(u, v) > s$
  let C be a minimal u,v-separator in graph G_b by Salemi and Buchanan (2020) (in $O(mn)$)
  add cut $x_{uj} + x_{vj} \leq z_j + \sum_{c \in C} x_{cj}$
  add cut $y_{uj} + y_{vj} \leq w_j + \sum_{c \in C} y_{cj}$

The minimal set C obtained by line 7 is minimalized in line 8. By "minimalizing set C" in line 8, we mean that we convert the minimal a,b-separator, obtained in line 7, to a length-s a,b-separator. The following proposition shows the time complexity of Algorithm 1.

Proposition 1. *Algorithm 1 runs in $O(mn)$.*

Proof. The time complexity of Algorithm 1 is determined by the complexities of lines 8 and 13 with a total time complexity of $O(mn)$. We note that the total time complexity of line 13 is $O(m)$. $\square$

During our computational experiments, we found out that adding the separator cuts for any pair of vertices $\{u,v\}$ with $\text{dist}_{G_b}(u,v) > s$ improves our computational performance. Hence, we add the separator cuts for any pair of vertices with $\text{dist}_{G_b}(u,v) > s$ in our code, which increases the time complexity to $O(mn^3)$.

Conflict constraints. Inequalities (1n)-(1o) can be added upfront when a pair of vertices $\{a,b\} \in \binom{V}{2}$ cannot belong to the same district; i.e., $\text{dist}_G(a,b) > s$. We define power graph $G^s := (V, E^s)$ with $E^s := \{\{u,v\} \in \binom{V}{2} : \text{dist}_G(u,v) \leq s\}$. Let $\bar{G}^s$ be the complement of G^s with $\mathcal{C}$ be the set of its maximal cliques. Then, the following conflict constraints can be added upfront.

$$\sum_{q \in Q} x_{qj} \leq z_j \qquad \forall \text{ maximal clique } Q \in \mathcal{C},\ \forall j \in [k] \tag{1p}$$

$$\sum_{q \in Q} y_{qj} \leq w_j \qquad \forall \text{ maximal clique } Q \in \mathcal{C},\ \forall j \in [k]. \tag{1q}$$

The following proposition shows that inequalities (1p) and (1q) are at least as strong as traditional strengthening inequalities in the redistricting literature.

Proposition 2. *For every land parcel $v \in V$ and every district $j \in [k]$, constraints $x_{vj} \leq z_j$ and $y_{vj} \leq w_j$ are implied by constraints* (1p) *and* (1q)*, respectively.*

Proof. Let $(\hat{x}, \hat{y}, \hat{z}, \hat{w}) \in \mathbb{R}_+^{(nk)^2 \times k^2}$ be a nonnegative point that satisfies constraints (1p)-(1q). For every land parcel $v \in V$, there exists a maximal clique Q_v that contains vertex v. So, for every land parcel $v \in V$ and every district $j \in [k]$, we have

$$\hat{x}_{vj} \leq \sum_{q \in Q_v} \hat{x}_{qj} \leq \hat{z}_j$$

$$\hat{y}_{vj} \leq \sum_{q \in Q_v} \hat{y}_{qj} \leq \hat{w}_j.$$

The first inequalities hold as $v \in Q_v$. The second inequalities hold by constraints (1p) and (1q). $\square$

5 Bounds on the number of Black-majority districts and s

This section proposes an upper bound on the number of Black-majority districts and a lower bound on the s values discussed in Section 4.

5.1 An upper bound on the number of Black-majority districts

The MIP model (1) allows every district to be a Black-majority district; however, one can provide an upper bound on the number of Black-majority districts and specify a range for the Black-majority districts. This section proposes a MIP formulation that finds an upper bound on the number of Black-majority districts (i.e., k_b). This upper bounding procedure decreases the number of x and z variables by $n(k-k_b)$ and $k-k_b$, respectively. In other words, the procedure fixes $(1-k_b/k)\times 100$ percent of x and z variables to zero. As we have at most k_b Black-majority districts, we will have at least $k-k_b$ non-Black-majority districts that result in $k-k_b$ one fixing of w variables (i.e., $(1-k_b/k)\times 100$ percent of w variables are fixed to one). To find k_b, we solve a truncated version of the MIP model (1) by removing the connectivity and compactness constraints from the MIP formulation (1). So, the proposed MIP formulation is valid for the Black-majority redistricting problem regardless of the compactness criterion.

$$\max \sum_{j=1}^{k} z_j \tag{2a}$$

$$Lz_j \le \sum_{v\in V} p_v x_{vj} \le U z_j \qquad \forall j \in [k] \tag{2b}$$

$$\sum_{v\in V} p_v^{\text{BVAP}} x_{vj} \ge f \sum_{v\in V} p_v^{\text{VAP}} x_{vj} \qquad \forall j \in [k] \tag{2c}$$

$$x_{vj} \le z_j \qquad \forall v \in V,\ \forall j \in [k] \tag{2d}$$

$$z_k \le z_{k-1} \le \cdots \le z_2 \le z_1 \le 1 \tag{2e}$$

$$x \in \{0,1\}^{n\times k}. \tag{2f}$$

Here, objective function (2a) maximizes the number of Black-majority districts. Constraints (2b) impose the population balance constraints. Constraints (2c) impose the numerosity condition of the Gingles prongs. Constraints (2d) imply that if a land parcel is assigned to a Black-majority district, then the district must be selected as a Black-majority district. Constraints (2e) are imposed for symmetry-breaking purposes. We note that the integrality of z variables is implied by constraints (2d)-(2f).

Our computational results for solving the MIP model (2) in a time limit of 3,600 seconds are summarized in Table 3. This table reports an upper bound on the number of Black-majority districts (k_b), the solving times (in seconds) of the model, and the zero fixing percentage of x and z variables resulting from solving the MIP model. Two observations from Table 3 are of interest: (i) at least half of the x and z variables are fixed to zero at both county and tract levels after solving the MIP model (2), and (ii) all the county level instances and most of the tract level instances are solved in less than a minute.

5.2 A valid lower bound for s

We recall that compactness constraints (1n)-(1o) of the MIP model (1) depend on the value of s; i.e., the diameter of a district's induced subgraph in a contiguity graph G. This section discusses valid values for the lower bound of s that will provide a starting point for our MIP-based lower bounding procedure in Section 6. To calculate a valid lower bound of s (i.e., ℓ_s), we first define

Table 3: Computational results for solving the MIP model (2) within a time limit of 7,200 seconds. The last two columns show the percentages of zero fixing for x and z variables and one fixing for w variables.

state	parcel	k	n	m	obj. (k_b)	time	x & z fixed (%)	w 1-fix (%)
MS	county	4	82	202	1.00	0.04	75.00	75.00
MS	tract	4	878	2,378	2.00	0.98	50.00	50.00
LA	tract	6	1,388	3,861	3.00	3.53	50.00	50.00
SC	tract	7	1,323	3,677	2.00	3.09	71.43	71.43
AL	tract	7	1,437	4,014	2.00	8.24	71.43	71.43
MD	tract	8	1,475	3,993	4.00	7.15	50.00	50.00
MO	tract	8	1,654	4,488	1.00	3.74	87.50	87.50
TN	tract	9	1,701	4,691	2.00	7.29	77.78	77.78
VA	tract	11	2,198	6,064	2.00	8.56	81.82	81.82
NJ	tract	12	2,181	6,061	1.00	5.45	91.67	91.67
MI	tract	13	3,017	7,989	2.00	32.50	84.62	84.62
NC	tract	14	2,672	7,422	3.00	20.20	78.57	78.57
GA	tract	14	2,796	7,762	7.00	138.12	50.00	50.00
OH	tract	15	3,168	8,747	2.00	94.76	86.67	86.67
IL	tract	17	3,265	8,728	3.00	204.64	82.35	82.35
PA	tract	17	3,446	9,641	2.00	55.61	88.24	88.24
TX	tract	38	6,896	18,554	3.00	3,829.72	92.11	92.11

power graph $G^s = (V, E^s)$ with

$$E^s := \left\{ \{u, v\} \in \binom{V}{2} : \text{dist}_G(u, v) \leq s \right\}, \tag{3}$$

where $\binom{V}{2}$ and $\text{dist}_G(u, v)$ are the set of all pairs of vertices and the distance between vertices u and v in contiguity graph G, respectively. Furthermore, let $\alpha(G^s)$ be the independence number of the power graph G^s. We note that if $\alpha(G^s) > k$, then it is impossible to partition the vertices of the contiguity graph G into k districts. Figure 4 illustrates a districting instance in which $\alpha(G^s) > k$.

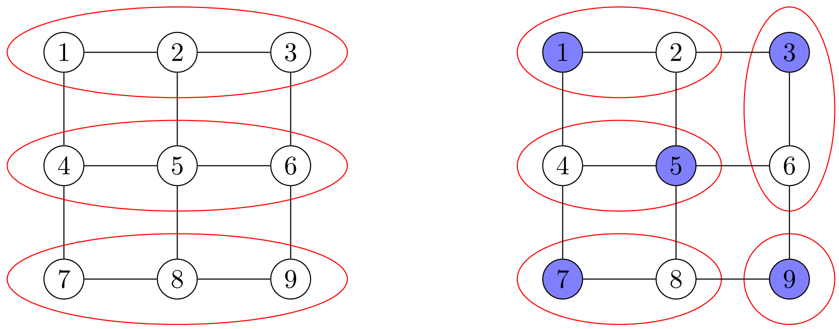

Figure 4: An illustration for the s lower bounding idea on a 3×3 contiguity graph with $p_v = 1$ for every vertex $v \in V$, $L = 1$, $U = 3$, and $k = 3$: (Left) a feasible redistricting plan with $s = 2$; (Right) an infeasible redistricting plan with $s = 1$ that fails to respect the number of districts ($k = 3$) because $\alpha(G^s) = \alpha(G) = 5 > 3 = k$.

Proposition 3 proves that the MIP fromulation (1) is infeasible if $\alpha(G^s) > k$.

Proposition 3. *The* MIP *formulation* (1) *with integer values of k and s is infeasible if $\alpha(G^s) > k$.*

Proof. Let $I \subseteq V$ be a maximum independent set of the power graph G^s with size $\alpha(G^s) > k$. Inequalites (1n)-(1o) of the MIP model (1) can be written as follows for the vertex set I because the length-s separator set is empty for any pair of vertices in I.

$$x_{aj} + x_{bj} \le z_j \qquad \forall \{a,b\} \in \binom{I}{2},\ \forall j \in [k]$$

$$y_{aj} + y_{bj} \le w_j \qquad \forall \{a,b\} \in \binom{I}{2},\ \forall j \in [k].$$

These constraints, along with assignment constraints (1b), imply that exactly k vertices of set I must be assigned to k parts (districts). This means that $|I| - k$ vertices cannot be assigned to any district (partition). So, the MIP model (1) is infeasible by the Pigeonhole principle. □

Using Proposition 3, we propose Algorithm 2 to provide a valid lower bound for s. Algorithm 2 determines ℓ_s by computing the independence number for power graphs of G, iteratively. If the independence number of the power graph is greater than k, then MIP model (1) is infeasible by Proposition 3. Algorithm 2 employs a binary search to find a lower bound on the value of s. In line 2 of the algorithm, we start the algorithm with a feasible lower bound of diam(G) for which $\alpha(G^{\text{diam}(G)}) = 1 < 2 \le k$. Furthermore, we solve the maximum independent set problem in line 6. Although finding a maximum independent set is an NP-hard problem (Garey and Johnson, 1979), our computational results show the efficacy of the Gurobi solver in solving the problem for all interesting instances. Table 4 summarizes the computational results for Algorithm 2. The lower bound of s (i.e., ℓ_s), the number of iterations till finding ℓ_s, and the running time of the algorithm are reported in this table. Interestingly, the proposed algorithm finds ℓ_s in less than 10 minutes for all instances at the county and tract level, except for TX.

Finally, Table 5 wraps up this section by summarizing the computational results provided in Sections 5.1 and 5.2. We employ these results in the subsequent sections of the paper.

Algorithm 2 LowerBounding-s (G, k)

$s_\ell \leftarrow 1$
$s_u \leftarrow \text{diam}(G)$
$\underline{s} \leftarrow 1$
while $s_u \neq s_\ell$ **do**
    $\underline{s} \leftarrow \lfloor (s_\ell + s_u)/2 \rfloor$
    create power graph $G^{\underline{s}}$ and calculate $\alpha(G^{\underline{s}})$
    if $\alpha(G^{\underline{s}}) > k$ **then**
        $s_\ell \leftarrow \underline{s} + 1$
    else
        $s_u \leftarrow \underline{s}$
return $\underline{s}$

Table 4: Computational results for the lower bounding Algorithm 2 for finding ℓ_s in interesting instances of the problem: "iter" shows the number of iterations and "time" illustrates the time spent for running the algorithm.

state	land parcel	k	n	m	ℓ_s	iter	time (s)
MS	county	4	82	202	6	3	0.14
MS	tract	4	878	2,378	15	5	25.40
LA	tract	6	1,388	3,861	14	5	69.42
SC	tract	7	1,323	3,677	13	5	59.78
AL	tract	7	1,437	4,014	14	5	72.76
MD	tract	8	1,475	3,993	14	5	82.20
MO	tract	8	1,654	4,488	14	5	91.08
TN	tract	9	1,701	4,691	14	6	90.44
VA	tract	11	2,198	6,064	13	6	152.75
NJ	tract	12	2,181	6,061	12	6	165.62
MI	tract	13	3,017	7,989	14	5	413.54
NC	tract	14	2,672	7,422	13	6	237.69
GA	tract	14	2,796	7,762	14	5	314.85
OH	tract	15	3,168	8,747	13	5	338.88
IL	tract	17	3,265	8,728	13	6	478.99
PA	tract	17	3,446	9,641	13	6	383.96
TX	tract	38	6,896	18,554	12	6	3,045.12

6 Improving Diameter's Lower Bound

This section proposes solving the feasibility variants of the MIP formulation (1) to increase the lower bound ℓ_s for every feasible number of Black-majority districts. To accelerate the solving process of the feasibility variants of MIP formulation (1), we first propose variable fixing procedures and a set of symmetry-breaking constraints. It turns out that even our proposed variable fixing procedures may increase the lower bound for some instances at the tract level.

Table 5: A summary of results for k_b and ℓ_s of all interesting instances of the problem.

state	parcel	k	n	m	k_b	ℓ_s
MS	county	4	82	202	1	6
MS	tract	4	878	2,378	2	15
LA	tract	6	1,388	3,861	3	14
SC	tract	7	1,323	3,677	2	13
AL	tract	7	1,437	4,014	2	14
MD	tract	8	1,475	3,993	4	14
MO	tract	8	1,654	4,488	1	14
TN	tract	9	1,701	4,691	2	14
VA	tract	11	2,198	6,064	2	13
NJ	tract	12	2,181	6,061	1	12
MI	tract	13	3,017	7,989	2	14
NC	tract	14	2,672	7,422	3	13
GA	tract	14	2,796	7,762	7	14
OH	tract	15	3,168	8,747	2	13
IL	tract	17	3,265	8,728	3	13
PA	tract	17	3,446	9,641	2	13
TX	tract	38	6,896	18,554	3	12

6.1 Variable fixing procedures

In Section 5.1, we introduced a MIP formulation that fixes at least 50% of x, z, and w variables. This section proposes more variable fixing procedures for x to reduce the size of the MIP formulation (1). These procedures detect land parcels that *cannot* belong to a Black-majority district. For illustration purposes, Figure 5 shows Tishomingo County (at the right top corner of the 2020 map of Mississippi) with a small BVAP percentage of 2.42%. As this county is surrounded by counties with small BVAP percentages, it is intuitively unlikely that one can develop a Black-majority district that includes Tishomingo County. Motivated by this observation, we propose a series of MIPs in three phases (Phase 1, Phase 2, and Phase 3) to determine whether a land parcel may be assigned to a feasible Black-majority district.

These MIP formulations differ by the inclusion of contiguity and compactness constraints and are listed as follows:

1. The MIP model of Phase 1 imposes (i) population balance and (ii) the numerosity criterion of the Gingles prongs;

2. The MIP model of Phase 2 imposes (i) population balance, (ii) the numerosity criterion of the Gingles prongs, and (iii) contiguity; and

3. The MIP model of Phase 3 imposes (i) population balance, (ii) the numerosity criterion of the Gingles prongs, (iii) contiguity, and (iv) a subset of compactness constraints.

Despite the differences in the MIP formulations of these phases, they all follow a similar fixing concept: if land parcel v cannot join *any* compact Black-majority district, then $x_{vj} \leq 0$ is a valid inequality for any district $j \in [k_b]$.

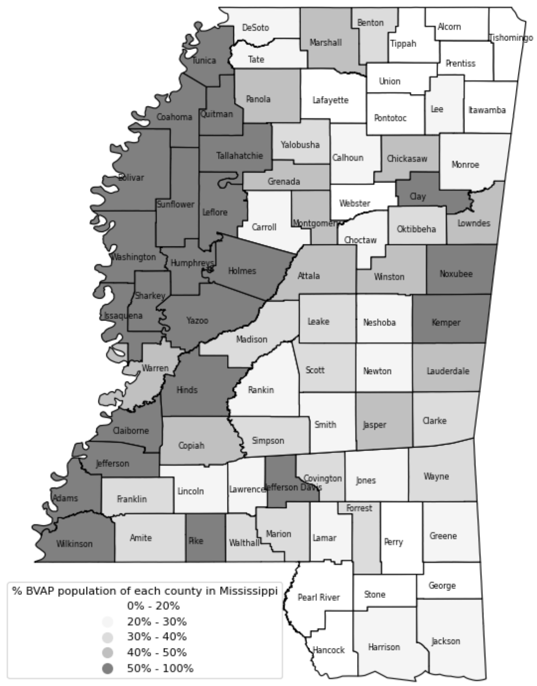

Figure 5: Mississippi's Black population percentages at the county level.

6.1.1 The MIP model of Phase 1

For every land parcel $v \in V$, the following MIP formulation checks if v can belong to a Black-majority district with respect to (i) population balance and (ii) the numerosity criterion of the Gingles prongs. We define $V_v := \{u \in V : \text{dist}_G(u, v) \leq s\}$ and $G_v := G[V_v]$ for every land parcel $v \in V$. We recall that $G[V_v]$ is the subgraph induced by V_v, and $\text{dist}_G(u, v)$ denotes the distance between vertices u and v in the contiguity graph G. For every land parcel $u \in V_v$, binary decision variable t_u is one if u belongs to the same Black-majority district that contains land parcel v. The MIP formulation has no objective function as our main concern is its feasibility.[1]

$$\text{(Phase 1)} \quad L \leq \sum_{u \in V_v} p_u t_u \leq U \tag{4a}$$

$$\sum_{u \in V_v} p_u^{\text{BVAP}} t_u \geq f \sum_{u \in V_v} p_u^{\text{VAP}} t_u \tag{4b}$$

$$t_v = 1 \tag{4c}$$

$$t_u \in \{0, 1\} \qquad \forall u \in V_v \setminus \{v\}. \tag{4d}$$

Constraint (4a) ensures that the Black-majority district, which contains land parcel v, respects population balance. Constraint (4b) enforces that the Black-majority district has a sufficient BVAP.

[1] Our experiments with the feasibility problem (i.e., MIP formulation (4) with no objective function) yielded better computational results rather than including an objective function (e.g., max $\sum_{u \in V_v} t_u$) to the MIP formulation).

Constraint (4c) guarantees that the Black-majority district includes land parcel v.

6.1.2 The MIP model of Phase 2

To fix more variables, we incorporate contiguity constraints into MIP formulation (4). Let F_0^1 be the set of vertices that cannot belong to a Black-majority district based on Phase 1 (i.e., their corresponding Black-majority assignment decision variables are fixed to zero after solving MIP formulation (4)). To impose contiguity, we employ a set of flow-based decision variables and constraints introduced by Shirabe (2005, 2009). This requires the bi-directed variant of the contiguity graph G, where $D = (V, A)$ is the bi-directed variant of G with $A := \{(u, v) \cup (v, u) : \{u, v\} \in E\}$. For every vertex $v \in V$, let $\delta^-(v)$ and $\delta^+(v)$ be the set of incoming and outgoing arcs of vertex v, respectively. For every directed edge $(i, j) \in A$, we define g_{ij} as the nonnegative flow on arc (i, j). Furthermore, we define $M_v := |V_v| - |V_v \cap F_0^1|$ for every $v \in V$. For every vertex $v \in V \setminus F_0^1$, the MIP of Phase 2 is provided as follows.

$$
\begin{aligned}
&\text{(Phase 2)} \quad && \text{constraints (4a)} - \text{(4d)} && && \text{(5a)}\\
& && t_u = 0 && \forall u \in V_v \cap F_0^1 && \text{(5b)}\\
& && g(\delta^-(u)) - g(\delta^+(u)) = t_u && \forall u \in V_v \setminus \{v\} && \text{(5c)}\\
& && g(\delta^-(u)) \leq (M_v - 1) t_u && \forall u \in V_v \setminus \{v\} && \text{(5d)}\\
& && g(\delta^-(v)) = 0 && && \text{(5e)}\\
& && g_{ij} \geq 0 && \forall (i, j) \in A(G[V_v]). && \text{(5f)}
\end{aligned}
$$

Phase 2 includes constraints (4a)-(4d) from Phase 1. Constraints (5b) do not allow the model to choose land parcels that are fixed in Phase 1. Constraints (5c) ensure that the net incoming flow to a land parcel is one if it is selected in the district and zero otherwise. Constraints (5d) imply that a land parcel cannot have incoming flow if it is not selected in the district. Constraint (5e) ensures that land parcel v gets no incoming flow.

6.1.3 The MIP model of Phase 3

In this phase, we include compactness-seeking constraints to fix more decisions. Let F_0^2 be the set of vertices that cannot belong to a compact (with respect to s value) Black-majority district based on the MIP formulation of Phase 2. We recall power graph $G^s := (V, E^s)$ with $E^s := \{\{u, v\} \in \binom{V}{2} : \text{dist}_G(u, v) \leq s\}$ from conflict constraints in Section 4. Let $\bar{G}^s$ be the complement of G^s with $\mathcal{C}$ be the set of the maximal cliques of $\bar{G}^s$. For every vertex $v \in V \setminus (F_0^1 \cup F_0^2)$, the following MIP formulation checks if v can belong to a compact Black-majority district with respect to the value of s.

$$
\begin{aligned}
&\text{(Phase 3)} \quad && \text{constraints (5a)} - \text{(5f)} && && \text{(6a)}\\
& && t_u = 0 && \forall u \in V_v \cap F_0^2 && \text{(6b)}\\
& && \sum_{q \in Q} t_q \leq 1 && \forall \text{ maximal clique } Q \in \mathcal{C} && \text{(6c)}\\
& && t_u \in \{0, 1\} && \forall u \in V_v. && \text{(6d)}
\end{aligned}
$$

Figure 6 shows the results of applying the three phases for fixing variables in Mississippi at the tract level with $s = 15$. Comparing Figure 6 to Figure 5, observe that for land parcels with low-density BVAP, many Black-majority assignment variables may be safely fixed to zero. Table 6

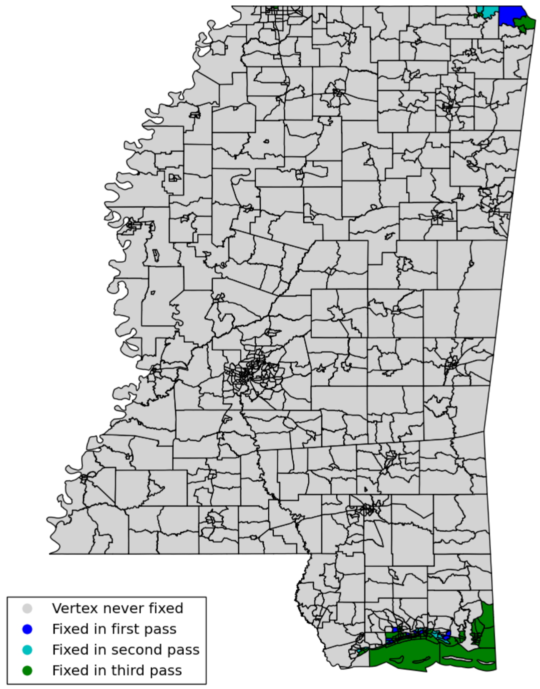

Figure 6: A tract-level map of MS showing parcels that are fixed using formulations (4), (5), and (6) with $s = 15$.

provides computational results of the three-phase fixing procedure for MS at the county level. For MS at the county instance, we start running the fixing procedure from $s = \ell_s$ and continue running the experiments for larger s values till no fixing is obtained. For MS at the county level, we start running the fixing procedure from $s = 6$. One can observe that the fixing percentage decreases at both county and tract levels as the s value increases. At the tract level, an interesting observation is that the fixing procedure may increase the lower bound of s (i.e., ℓ_s) for some states. For example, Table 7 shows that the value of ℓ_s, which was originally 14 by running Algorithm 2 for MO at the tract level, is increased to 28 after applying the three-phase fixing procedure. We observe a similar increase of ℓ_s for MI (from 14 to 15), NC (from 14 to 16), OH (14 to 18), and TX (from 12 to 16) at the tract level. Due to space limitations, the rest of the tract-level results for the three-phase fixing procedure are provided in Appendix B. We impose a 10-second time limit for every MIP process at each iteration of the three-phase fixing procedure.

6.2 Symmetry breaking constraints

The MIPs of the political districting problems usually suffer from symmetry as they are specific cases of the partitioning problem. In Section 4, we propose a set of symmetry-breaking constraints for breaking the symmetry between Black-majority districts (i.e., $z_k \leq \cdots \leq z_1$) and discuss an implied set of symmetry-breaking constraints for non-Black-majority districts (i.e., $w_1 \leq \cdots \leq w_k$).

Table 6: Results for the three-phase fixing procedure of x variables at the county level for MS. Here, ${\ell_s}^{\text{base}}$ and ${\ell_s}^{\text{fix}}$ represent the lower bounds obtained by Algorithm 2 in Section 5 and the three-phase fixing procedure, respectively.

state	k	n	m	s	# fixed	% fixed	time	${\ell_s}^{\text{base}}$	${\ell_s}^{\text{fix}}$
MS	4	82	202	6	12	14.63	0.17	6	6
				7	7	14.63	0.20		
				8	2	8.54	0.22		
				9	1	1.22	0.23		
				10	1	1.22	0.30		
				11	1	1.22	0.49		
				12	1	1.22	0.44		
				13	1	1.22	0.39		

Table 7: Results for the three-phase fixing procedure of x variables for Missouri at the tract level. Here, ${\ell_s}^{\text{base}}$ and ${\ell_s}^{\text{fix}}$ represent the lower bounds obtained by Algorithm 2 in Section 5 and the three-phase fixing procedure, respectively.

state	k	n	m	s	# fixed	% fixed	time	${\ell_s}^{\text{base}}$	${\ell_s}^{\text{fix}}$
MO	8	1,654	4,488	14	1,654	100.00	0.00	14	28
				15	1,654	100.00	0.00		
				16	1,654	100.00	0.00		
				17	1,654	100.00	0.00		
				18	1,654	100.00	0.00		
				19	1,654	100.00	11.48		
				20	1,654	100.00	227.16		
				21	1,654	100.00	369.77		
				22	1,654	100.00	641.00		
				23	1,654	100.00	836.26		
				24	1,654	100.00	1,303.61		
				25	1,654	100.00	1,605.74		
				26	1,654	100.00	2,003.66		
				27	1,654	100.00	2,234.26		
				28	138	8.34	6,392.45		
				29	71	4.29	3,801.05		
				30	18	1.09	3,024.75		
				31	1	0.06	2,671.90		

This section proposes a new set of symmetry-breaking constraints based on the following facts given a specific $s \in [\ell_s, u_s]$: (i) $z_{k_b+1} = \cdots = z_k = 0$; (ii) $w_{k_b+1} = \cdots = w_k = 1$; and (iii) no vertex pair of an independent set of a power graph G^s can belong to a similar district (see Figure 7 for an illustration). Proposition 4 proposes a set of super valid symmetry-breaking constraints for the MIP formulation (4).

Proposition 4 (Symmetry-breaking Constraints)**.** *Let $s \in [\ell_s, u_s]$ and I as an independent set of*

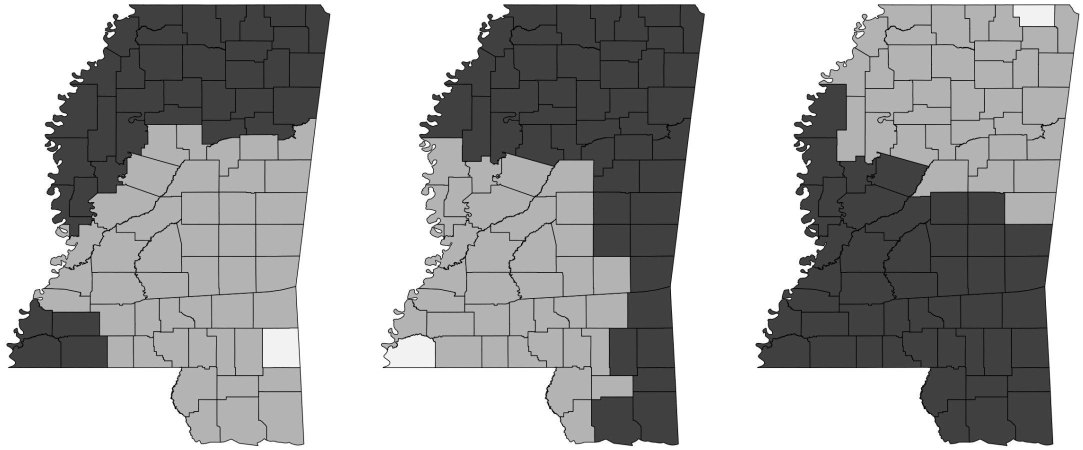

Figure 7: A maximum independent set (white land parcels) for the power graph of MS with $s = 6$. The gray land parcels are reachable from the white parcel, and the black parcels are not reachable from the white land parcel.

G^s and define $q := \min\{|I|, k - k_b\}$. Then, constraints

$$y_{v_i,k-i+1} + \sum_{j=1}^{\min\{i,k_b\}} x_{v_i,j} = 1 \qquad \forall i \in \{1, 2, \dots, q\} \tag{7}$$

are super valid for formulation (1).

Proof. By the contradiction. Suppose that no optimal solution of the MIP formulation (1) satisfies constraints (7). Let (x^*, y^*, z^*, w^*) be an optimal solution of formulation (1). Then, (x^*, y^*, z^*, w^*) does not satisfy at least one constraint (say for $i^* \in \{1, 2, \dots, q\}$) of type (7). Then, we have either of the following cases:

1. $y^*_{v_{i^*},k-i^*+1} = 1$ and $\sum_{j=1}^{\min\{i^*,k_b\}} x^*_{v_{i^*},j} = 1$. Then, (x^*, y^*, z^*, w^*) does not satisfy constraints (1b). This is a contradiction.

2. $y^*_{v_{i^*},k-i^*+1} = 0$ and $\sum_{j=1}^{\min\{i^*,k_b\}} x_{v_{i^*},j} = 0$. Then, we have either of the following cases:
   - v_{i^*} belongs to a non-Black majority district $\ell \neq k - i^* + 1$ based on the optimal solution (x^*, y^*, z^*, w^*). Then, we can switch the districts to obtain another optimal solution $(\hat{x}, \hat{y}, \hat{z}, \hat{w})$. This is a contradiction.
   - v_{i^*} belongs to a Black majority district $\ell \notin \{1, \dots, \min\{i^*, k_b\}\}$ based on the optimal solution (x^*, y^*, z^*, w^*). We note that $\ell \leq k_b$ as $z^*_j = 0$ for every $j > k_b$. So, $\ell \in [i^* + 1, k_b]$ and this happens only if $i^* = \min\{i^*, k_b\} < k_b$. By the symmetry-breaking constraints (1g), there is a district $q \in\in \{1, \dots, i^*\}$ with $z^*_q = 1$. Then, we can switch the labels of districts ℓ and q and get a new optimal solution $(\hat{x}, \hat{y}, \hat{z}, \hat{w})$. This is a contradiction.

□

We finally note that our proposed symmetry-breaking constraints decrease the MIP solution time from 11,880.99 to 9.81 seconds to obtain the optimum objective value of one for MS at the county level with $s = 6$.

6.3 Improving bounds on s by feasibility MIPs

This section proposes a procedure in which we solve a series of feasibility variants of the MIP formulation (1) for specific values of k_b and s to increase the lower bounds obtained in the previous sections. In better words, we start solving the feasibility MIP from $k_b = 1$ and $s = \ell_s$ and continue till a feasible solution is obtained or the 1-hour time limit is reached. For MS at the county level, the MIP formulation finds a feasible solution for $k_b = 1$ and $s = 6$ in the first iteration, which proves the optimality of 6 in less than 5 seconds. Furthermore, the MIP feasibility approach increases the lower bound of s for the following tract-level instances:

- LA from 14 to 16 in less than 90 seconds when $k_b = 3$;
- SC from 13 to 16 in less than 100 seconds when $k_b = 2$;
- TN from 14 to 33 in less than 15 seconds when $k_b = 2$;
- VA from 13 to 22 in less than 400 seconds when $k_b = 2$;
- MI from 15 to 17 in less than 90 seconds when $k_b = 2$;
- NC from 16 to 17, 16 to 18, and 16 to 25 in less than 3,000 seconds when $k_b = 1$, $k_b = 2$, and $k_b = 3$, respectively;
- GA from 14 to 16 in less than 2,000 seconds when $k_b = 7$;
- OH from 18 to 19 and 18 to 26 in less than 500 seconds when $k_b = 1$ and $k_b = 2$, respectively;
- TX from 16 to 17 in less than 200 seconds when $k_b \in \{1, 2, 3\}$.

We note that due to the large sizes of tract-level instances for IL, PA, and TX, the feasibility variant of MIP formulation (1) is run only for the Black-majority districts with x and z variables. Table 8 summarizes the s lower bounds that are obtained by Algorithm 2 (base), the three-phase fixing procedures (fix), and the feasibility MIP (feas). Due to space limitations, the details of our computational experiments are provided on our GitHub repository.

7 Optimality Gaps

This section provides feasible districting plans with Black-majority districts that are obtained by GerryChain, a state-of-the-art districter that can maximize the number of Black-majority districts employing short bursts. In the short burst approach, one (i) runs an unbiased random walk for a small number of steps and then (ii) restarts the random walk from the best plan encountered in the last burst (Cannon et al., 2023). We employ the feasible solutions obtained by 10,000 iterations of GerryChain to assess a best existing optimality gap for s. Due to space limitations, a detailed report of our experiments with the GerryChain short bursts module is available at our GitHub repository.

Table 8: Lower bounds of s (i.e., ℓ_s) that are obtained by Algorithm 2 (base), the three-phase fixing procedures (fix), and the feasibility MIP (feas). Asterisks denote the lower bounds that are obtained by the restricted feasibility MIP with only x and z variables.

state	parcel	k	n	m	ℓ_s^{base}	ℓ_s^{fix}	k_b	ℓ_s^{feas}
MS	county	4	82	202	6	6	1	6
MS	tract	4	878	2,378	15	15	1	15
							2	15
LA	tract	6	1,388	3,861	14	14	1	14
							2	14
							3	16
SC	tract	7	1,323	3,677	13	13	1	13
							2	16
AL	tract	7	1,437	4,014	14	14	1	15
							2	15
MD	tract	8	1,475	3,993	14	14	1	14
							2	14
							3	14
							4	14
MO	tract	8	1,654	4,488	14	28	1	28
TN	tract	9	1,701	4,691	14	14	1	14
							2	33
VA	tract	11	2,198	6,064	13	13	1	13
							2	22
NJ	tract	12	2,181	6,061	12	12	1	12
MI	tract	13	3,017	7,989	14	15	1	15
							2	17
NC	tract	14	2,672	7,422	13	16	1	17
							2	18
							3	25
GA	tract	14	2,796	7,762	14	14	1	14
							2	14
							3	14
							4	14
							5	14
							6	14
							7	16
OH	tract	15	3,168	8,747	13	18	1	19
							2	26
IL	tract	17	3,265	8,728	13	13	1	13*
							2	13*
							3	13*
PA	tract	17	3,446	9,641	13	13	1	13*
							2	13*
TX	tract	38	6,896	18,554	12	16	1	17*
							2	17*
							3	17*

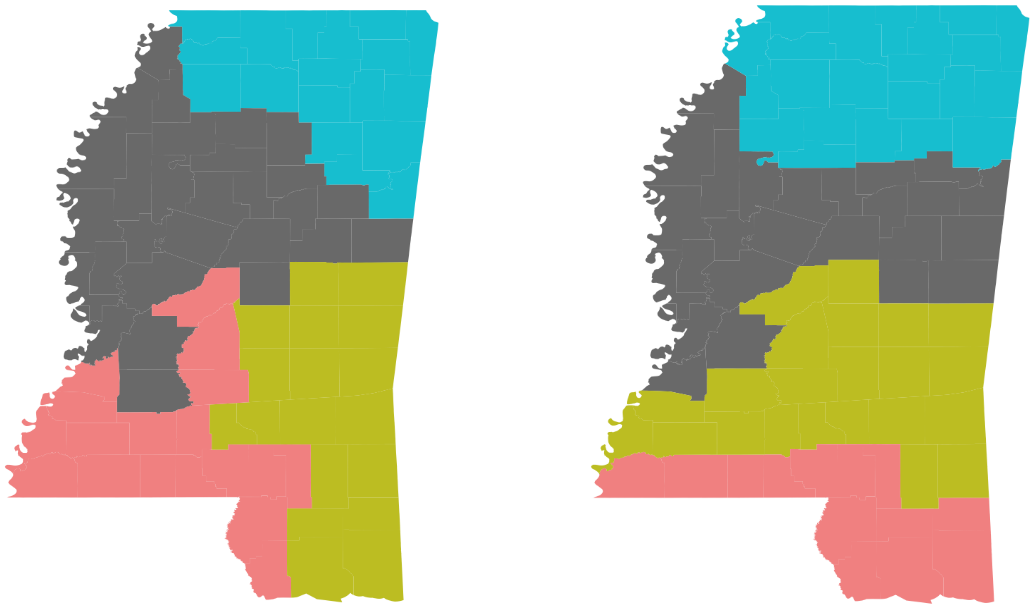

Figure 8: MS maps at the county level with one Black-majority district (in gray): (Left) an optimal map with $s = 6$ obtained by MIP formulation (1); (Right) a feasible map with $s = 7$ obtained by short bursts of GerryChain.

Table 9 illustrates the optimality gaps for the number of Black-majority districts (i.e., k_b) and the diameter of districts (i.e., s). For MS at the county level, our proposed MIP formulation provides an optimal map with $k_b = 1$ and $s = 6$; however, GerryChain provides a feasible map with $k_b = 1$ and $s = 7$. Figure 8 illustrates the county-level maps that are obtained by the proposed MIP and GerryChain. At the tract level, we obtain optimality gaps between 6.25 (for MS) and 44.00 (for GA) percentages. Furthermore, it is not surprising that u_s value increases as the value of k_b increases. The most considerable jump is the increase of u_s from 21 to 25 for GA at the tract level. Furthermore, we observe that the short bursts module of GerryChain returns no feasible map with Black-majority districts for LA, MO, VA, NJ, NC, OH, and TX at the tract level. These observations are encouraging for optimizers to close the gap by providing either more quality maps or more efficient bounds.

Table 9: Lower (ℓ_s) and upper (u_s) bounds for s obtained by our proposed procedures and 10,000 iterations of GerryChain's short bursts, respectively. The asterisk denotes the optimality of s.

state	parcel	k	n	m	k_b	ℓ_s	u_s	s-gap (%)
MS	county	4	82	202	1	6*	7	14.29
MS	tract	4	878	2,378	1	15	16	6.25
					2	15	-	-
LA	tract	6	1,388	3,861	1	14	-	-
					2	14	-	-
					3	16	-	-
SC	tract	7	1,323	3,677	1	13	20	35.00
					2	16	-	-
AL	tract	7	1,437	4,014	1	15	24	37.50
					2	15	-	-
MD	tract	8	1,475	3,993	1	14	22	36.36
					2	14	22	36.36
					3	14	23	39.13
					4	14	-	-
MO	tract	8	1,654	4,488	1	28	-	-
TN	tract	9	1,701	4,691	1	14	17	17.64
					2	33	-	-
VA	tract	11	2,198	6,064	1	13	-	-
					2	22	-	-
NJ	tract	12	2,181	6,061	1	12	-	-
MI	tract	13	3,017	7,989	1	15	21	28.57
					2	17	-	-
NC	tract	14	2,672	7,422	1	17	-	-
					2	18	-	-
					3	25	-	-
GA	tract	14	2,796	7,762	1	14	20	30.00
					2	14	20	30.00
					3	14	21	33.33
					4	14	25	44.00
					5	14	-	-
					6	14	-	-
					7	16	-	-
OH	tract	15	3,168	8,747	1	17	-	-
					2	26	-	-
IL	tract	17	3,265	8,728	1	13	18	27.78
					2	13	18	27.78
					3	13	-	-
PA	tract	17	3,446	9,641	1	13	19	31.58
					2	13	-	-
TX	tract	38	6,896	18,554	1	17	-	-
					2	17	-	-
					3	17	-	-

8 Conclusion

This paper proposes a redistricting MIP formulation that captures a compactness measure (i.e., the diameter of districts) in constraints and provides opportunities for optimizers to optimize other redistricting measures in the objective function of their MIP formulations. We consider maximizing the number of Black-majority districts in the objective function of our proposed MIP formulation to explore the possibility of creating the maximum number of reasonably compact Black-majority districts. Furthermore, we propose fixing procedures and symmetry-breaking constraints to speed up the computational performance of our MIP formulation. Our computational experiments show that the proposed MIP formulation can find a districting map for MS at the county level with one Black-majority district and a diameter of 6 (both optimum) in a matter of seconds, while a state-of-the-art districting optimization package (i.e., GerryChain) finds only one Black-majority districts and a diameter of 7 after 10,000 iterations. At the tract level, our proposed MIP formulation provides bounds on the number of Black-majority districts and the diameters of their corresponding maps. The optimality gap between our bounds and the feasible maps found by GerryChain suggests a huge opportunity for exact and non-exact optimizers to close the gaps either by producing quality feasible maps or increasing infeasibility bounds.

Acknowledgement

This work is based upon work supported by Rice University's Building Research on Inequality and Diversity to Grow Equity (BRIDGE) seed grant and the National Science Foundation (NSF) grant DMS-2318790 titled *AMPS: Novel Combinatorial Optimization Techniques for Smartgrids and Power Networks*. Also, we would like to thank Austin Buchanan for the early discussions and constructive comments.

References

Micah Altman and Michael McDonald. Redistricting by formula: An Ohio reform experiment. *American Politics Research*, 46(1):103–131, 2018.

Verónica Arredondo, Miguel Martínez-Panero, Teresa Peña, and Federica Ricca. Mathematical political districting taking care of minority groups. *Annals of Operations Research*, 305(1-2): 375–402, 2021.

Amariah Becker and Justin Solomon. Redistricting algorithms. *Political Geometry*, 2021.

Pietro Belotti, Austin Buchanan, and Soraya Ezazipour. Political districting to optimize the polsby-popper compactness score. *Available on Optimization-Online*, 2023.

John R Birge. Redistricting to maximize the preservation of political boundaries. *Social Science Research*, 12(3):205–214, 1983.

Burcin Bozkaya, Erhan Erkut, and Gilbert Laporte. A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, 144(1):12–26, 2003.

Austin Buchanan, Soraya Ezazipour, and Maral Shahmizad. A widespread belief about county splits in political districting plans is wrong. *Available at SSRN 4670830*, 2024.

Sarah Cannon, Ari Goldbloom-Helzner, Varun Gupta, JN Matthews, and Bhushan Suwal. Voting rights, Markov chains, and optimization by short bursts. *Methodology and Computing in Applied Probability*, 25(1):36, 2023.

Daryl DeFord, Moon Duchin, and Justin Solomon. A computational approach to measuring vote elasticity and competitiveness. *Statistics and Public Policy*, 7(1):69–86, 2020.

Daryl DeFord, Moon Duchin, and Justin Solomon. Recombination: A family of markov chains for redistricting. *Harvard Data Science Review*, 3(1):3, 2021.

Jitender S Deogun, Dieter Kratsch, and George Steiner. An approximation algorithm for clustering graphs with dominating diametral path. *Information Processing Letters*, 61(3):121–127, 1997.

Kiera W. Dobbs, Douglas M. King, and Sheldon H. Jacobson. Redistricting optimization with recombination: A local search case study. *Computers & Operations Research*, 160:106369, 2023. ISSN 0305-0548. doi: https://doi.org/10.1016/j.cor.2023.106369. URL `https://www.sciencedirect.com/science/article/pii/S0305054823002332`.

Kiera W. Dobbs, Douglas M. King, Ian G. Ludden, and Sheldon H. Jacobson. Facilitating compromise in redistricting with transfer distance midpoints. *INFORMS Journal on Optimization*, 2024. doi: 10.1287/ijoo.2023.0029. URL `https://doi.org/10.1287/ijoo.2023.0029`.

Moon Duchin and Olivia Walch. Political geometry, 2021.

Matteo Fischetti, Markus Leitner, Ivana Ljubić, Martin Luipersbeck, Michele Monaci, Max Resch, Domenico Salvagnin, and Markus Sinnl. Thinning out Steiner trees: a node-based model for uniform edge costs. *Mathematical Programming Computation*, 9(2):203–229, 2017.

Michael R Garey and David S Johnson. *Computers and intractability*, volume 174. Freeman San Francisco, 1979.

Robert S Garfinkel and George L Nemhauser. Optimal political districting by implicit enumeration techniques. *Management Science*, 16(8):B–495, 1970.

Timo Gschwind, Stefan Irnich, Fabio Furini, and Roberto Wolfler Calvo. A branch-and-price framework for decomposing graphs into relaxed cliques. *INFORMS Journal on Computing*, 33(3): 1070–1090, 2021.

Wes Gurnee and David B Shmoys. Fairmandering: A column generation heuristic for fairness-optimized political districting. In *SIAM Conference on Applied and Computational Discrete Algorithms (ACDA21)*, pages 88–99. SIAM, 2021.

J Gerald Hebert, Martina E Vandenberg, and Paul Smith. The realist's guide to redistricting: Avoiding the legal pitfalls, 2010.

SW Hess, JB Weaver, HJ Siegfeldt, JN Whelan, and PA Zitlau. Nonpartisan political redistricting by computer. *Operations Research*, 13(6):998–1006, 1965.

Sam Hirsch, Jessica Amunson, and Mary Marshall. Brief of computational redistricting experts as Amici curiae in support of appellees and respondents, 2022.

Connor Lawless and Oktay Günlük. Fair minimum representation clustering. In *International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 20–37. Springer, 2024.

David Lublin, Lisa Handley, Thomas L Brunell, and Bernard Grofman. Minority success in non-majority minority districts: Finding the "sweet spot". *Journal of Race, Ethnicity, and Politics*, 5(2):275–298, 2020.

Ian G Ludden, Rahul Swamy, Douglas M King, and Sheldon H Jacobson. A bisection protocol for political redistricting. *INFORMS Journal on Optimization*, 5(3):233–255, 2023.

Anuj Mehrotra, Ellis L Johnson, and George L Nemhauser. An optimization based heuristic for political districting. *Management Science*, 44(8):1100–1114, 1998.

Stuart S Nagel. Simplified bipartisan computer redistricting. *Stan. L. Rev.*, 17:863, 1964.

Johannes Oehrlein and Jan-Henrik Haunert. A cutting-plane method for contiguity-constrained spatial aggregation. *Journal of Spatial Information Science*, 2017(15):89–120, 2017.

Federica Ricca and Bruno Simeone. Local search algorithms for political districting. *European Journal of Operational Research*, 189(3):1409–1426, 2008.

Federica Ricca, Andrea Scozzari, and Bruno Simeone. Political districting: from classical models to recent approaches. *Annals of Operations Research*, 204(1):271–299, 2013.

Hosseinali Salemi and Austin Buchanan. Parsimonious formulations for low-diameter clusters. *Mathematical Programming Computation*, 12(3):493–528, 2020.

Maral Shahmizad and Austin Buchanan. Political districting to minimize county splits. *Available on Optimization-Online*, 2023.

Takeshi Shirabe. A model of contiguity for spatial unit allocation. *Geographical Analysis*, 37(1): 2–16, 2005.

Takeshi Shirabe. Districting modeling with exact contiguity constraints. *Environment and Planning B: Planning and Design*, 36(6):1053–1066, 2009.

Rahul Swamy, Douglas M. King, and Sheldon H. Jacobson. A case for transparency in the design of political districts, 2019. Working paper.

Rahul Swamy, Douglas M. King, and Sheldon H. Jacobson. Multiobjective optimization for politically fair districting: A scalable multilevel approach. *Operations Research*, 71(2):536–562, 2023. doi: 10.1287/opre.2022.2311. URL `https://doi.org/10.1287/opre.2022.2311`.

The Alabama Bicentennial Commission. The Alabama Black belt, 2024. `https://alafricanamerican.com/beyond-the-book-honoree-archives/beyond-the-book-the-alabama-black-belt/`.

Michał K Tomczyk and Miłosz Kadziński. Evolutionary algorithms for solving single-and multiple-objective political redistricting problems: The case study of Poland. *Applied Soft Computing*, 152:111258, 2024.

Hamidreza Validi and Austin Buchanan. Political districting to minimize cut edges. *Mathematical Programming Computation*, 14(4):623–672, 2022.

Hamidreza Validi, Austin Buchanan, and Eugene Lykhovyd. Imposing contiguity constraints in political districting models. *Operations Research*, 70(2):867–892, 2022.

James B Weaver and Sidney W Hess. A procedure for nonpartisan districting: Development of computer techniques. *Yale Law Journal*, 73:288–308, 1963.

Laurence A Wolsey and George L Nemhauser. *Integer and combinatorial optimization*, volume 55. John Wiley & Sons, 1999.

Oleksandra Yezerska, Foad Mahdavi Pajouh, Alexander Veremyev, and Sergiy Butenko. Exact algorithms for the minimum s-club partitioning problem. *Annals of Operations Research*, 276 (1-2):267–291, 2019.

H Peyton Young. Measuring the compactness of legislative districts. *Legislative Studies Quarterly*, 13(1):105–115, 1988.

Jack Zhang, Lucas Silveira, Hamidreza Validi, Austin Buchanan, and Illya V. Hicks. Integer programming formulations for partitioning a graph into low-diameter clusters, 2024a. Working paper.

Jack Zhang, Hamidreza Validi, Austin Buchanan, and Illya V. Hicks. Linear-size formulations for connected planar graph partitioning and political districting. *Optimization Letters*, 18(1):19–31, 2024b. doi: 10.1007/s11590-023-02070-0. URL `https://doi.org/10.1007/s11590-023-02070-0`.

Appendices

Appendix A. Excluded instances

Table 10: States excluded from this paper with the reason for their omission.

Land parcel	Reason for exclusion	States
County/tract	One-district state	AK, DE, ND, SD, VT, WY
County	Not connected	HI
	Containing a county v with $p_v > U$	AZ, CA, CO, CT, FL, GA, IA, IL, IN, KY, MA, MD, MI, MN, MO, NC, NJ, NY, OH, OK, OR, PA, RI, TN, TX, UT, VA, WA, WI
	No Black-majority district by MIP (2)	AL, AR, ID, KS, LA, ME, MT, NE, NH, NM, SC, WV
Tract	Not connected	CA, FL, HI, NY, RI
	No Black-majority district by MIP (2)	AR, AZ, CO, CT, IA, ID, IL, IN, KS, KY, MA, ME, MN, MT, NH, NM, OK, OR, UT, WA, WV

Appendix B. Three-phase fixing results for tract-level instances

Table 11: Results for the three-phase fixing procedure of x variables for Mississippi at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
MS	4	878	2,378	15	100	11.39	56.49	15	15
				16	45	5.13	59.60		
				17	10	1.14	42.82		
				18	0	0.00	42.49		

Table 12: Results for the three-phase fixing procedure of x variables for Louisiana at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
LA	6	1,388	3,861	14	122	8.79	465.78	14	14
				15	59	4.25	340.96		
				16	11	0.79	355.90		
				17	1	0.07	316.43		
				18	0	0.00	296.42		

Table 13: Results for the three-phase fixing procedure of x variables for South Carolina at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
SC	7	1,323	3,677	13	456	34.47	470.14	13	13
				14	303	22.90	298.56		
				15	245	18.52	329.55		
				16	175	13.23	337.84		
				17	132	9.98	386.45		
				18	100	7.56	376.78		
				19	69	5.22	501.28		
				20	26	1.97	468.49		
				21	4	0.30	488.87		
				22	0	0.00	488.87		

Table 14: Results for the three-phase fixing procedure of x variables for Alabama at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
AL	7	1,437	4,014	14	638	44.40	695.32	14	14
				15	340	23.66	679.53		
				16	207	14.41	547.71		
				17	89	6.19	523.89		
				18	23	1.60	442.13		
				19	1	0.07	442.89		
				20	0	0.00	436.91		

Table 15: Results for the three-phase fixing procedure of x variables for Maryland at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
MD	8	1,475	3,993	14	198	13.42	314.87	14	14
				15	130	8.81	314.87		
				16	67	4.54	304.66		
				17	56	3.80	357.15		
				18	48	3.25	378.67		
				19	34	2.31	381.99		
				20	32	2.17	466.88		
				21	31	2.10	480.43		
				22	30	2.03	487.39		
				23	28	1.90	587.68		
				24	21	1.42	614.75		
				25	17	1.15	635.49		
				26	16	1.08	585.62		
				27	12	0.81	889.74		
				28	8	0.54	1,212.77		
				29	5	0.34	1,524.20		
				30	1	0.07	1,420.53		
				31	0	0.00	1,364.92		

Table 16: Results for the three-phase fixing procedure of x variables for Tennessee at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
TN	9	1,701	4,691	14	1,343	78.95	38.73	14	14
				15	1,286	75.60	36.14		
				16	1,270	74.66	29.78		
				17	1,250	73.49	66.77		
				18	1,228	72.19	59.95		
				19	1,209	71.08	124.78		
				20	1,180	69.37	169.03		
				21	1,136	66.78	197.34		
				22	1,059	62.26	242.74		
				23	965	56.73	321.37		
				24	888	52.20	409.91		
				25	796	46.80	488.25		
				26	712	41.86	578.85		
				27	652	38.33	504.57		
				28	616	36.21	594.05		
				29	577	33.92	683.96		
				30	542	31.86	787.15		
				31	512	30.10	787.97		
				32	472	27.75	952.98		
				33	414	24.34	1,024.36		
				34	371	21.81	1,189.47		
				35	334	19.64	1,275.33		
				36	304	17.87	1,319.14		
				37	265	15.58	1,504.31		
				38	221	12.99	1,895.11		
				39	163	9.58	2,125.67		
				40	120	7.05	2,420.92		
				41	95	5.58	2,246.89		
				42	75	4.41	2,711.68		
				43	57	3.35	2,474.11		
				44	42	2.47	2,896.58		
				45	22	1.28	3,529.58		
				46	7	0.41	3,973.20		
				47	2	0.12	3,801.23		
				48	0	0.00	4,305.52		

Table 17: Results for the three-phase fixing procedure of x variables for Virginia at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
VA	11	2,198	6,064	13	1,365	62.10	284.89	13	13
				14	1,283	58.37	279.60		
				15	1,162	52.87	459.50		
				16	1,081	49.18	594.35		
				17	955	43.45	805.35		
				18	872	39.67	919.64		
				19	791	35.99	1,171.79		
				20	674	30.66	1,480.09		
				21	555	25.25	1,885.22		
				22	483	21.97	2,030.59		
				23	427	19.43	2,162.42		
				24	363	16.52	2,578.60		
				25	303	13.79	2,585.15		
				26	237	10.78	3,071.22		
				27	190	8.64	3,222.14		
				28	146	6.64	3,254.63		
				29	102	4.64	3,755.10		
				30	68	3.09	3,821.19		
				31	38	1.73	4,005.55		
				32	1	0.05	4,482.87		
				33	0	0.00	4,672.11		

Table 18: Results for the three-phase fixing procedure of x variables for New Jersey at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
NJ	12	2,181	6,061	12	1,853	84.96	138.54	12	12
				13	1,691	77.53	222.13		
				14	1,537	70.47	317.48		
				15	1,380	63.27	430.47		
				16	1,271	58.28	532.78		
				17	1,158	53.09	762.78		
				18	1,054	48.33	1,000.93		
				19	949	43.51	1,305.67		
				20	880	40.35	1,401.94		
				21	830	38.06	1,566.73		
				22	797	36.54	1,812.97		
				23	763	34.98	2,151.16		
				24	660	30.26	2,902.20		
				25	522	23.93	3,546.68		
				26	336	15.41	3,874.18		
				27	277	12.70	3,876.91		
				28	211	9.67	3,945.32		
				29	138	6.33	4,064.90		
				30	100	4.59	4,633.14		
				31	73	3.35	4,628.49		
				32	46	2.11	4,935.96		
				33	24	1.10	4,956.61		
				34	5	0.23	4,818.38		
				35	0	0.00	5,086.64		

Table 19: Results for the three-phase fixing procedure of x variables for Michigan at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
MI	13	3,017	7,989	14	3,017	100.00	443.20	14	15
				15	2,639	87.47	761.21		
				16	2,106	69.80	1,283.50		
				17	1,909	63.27	1,593.33		
				18	1,716	56.88	2,096.85		
				19	1,592	52.77	3,113.22		
				20	1,494	49.52	3,478.95		
				21	1,388	46.01	4,466.98		
				22	1,270	42.09	5,272.72		
				23	1,143	37.89	5,892.18		
				24	994	32.95	7,277.84		
				25	836	27.71	7,194.58		
				26	720	23.86	7,377.45		
				27	581	19.26	8,291.18		
				28	379	12.56	9,536.25		
				29	258	8.55	10,297.87		
				30	150	4.97	10,413.86		
				31	54	1.79	10,477.02		
				32	6	0.20	9,115.45		
				33	0	0.00	8,400.27		

Table 20: Results for the three-phase fixing procedure of x variables for North Calorlina at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
				13	2,672	100.00	55.07		
				14	2,672	100.00	444.16		
				15	2,672	100.00	1,290.53		
				16	2,342	87.65	2,046.44		
				17	1,830	68.49	4,071.87		
				18	1,037	38.81	8,966.52		
				19	762	28.52	8,732.58		
				20	630	23.58	7,292.84		
				21	545	20.40	5,768.35		
				22	439	16.43	5,100.45		
NC	14	2,672	7,422	23	348	13.02	5,380.70	14	16
				24	249	9.32	5,243.85		
				25	188	7.04	5,552.53		
				26	158	5.91	5,853.85		
				27	115	4.30	5,845.93		
				28	83	3.11	5,888.18		
				29	53	1.98	5,864.96		
				30	32	1.20	6,197.21		
				31	23	0.86	6,257.99		
				32	12	0.45	6,239.49		
				33	5	0.19	6,737.23		
				34	0	0.00	6,978.24		

Table 21: Results for the three-phase fixing procedure of x variables for Georgia at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
				14	769	27.50	680.66		
				15	583	20.85	801.69		
				16	421	15.06	1,117.22		
				17	293	10.48	1,635.13		
GA	14	2,796	7,762	18	186	6.65	1,794.78	14	14
				19	101	3.61	2,086.20		
				20	43	1.54	2,411.56		
				21	11	0.39	2,346.62		
				22	2	0.07	2,391.54		
				22	0	0.00	2,676.54		

Table 22: Results for the three-phase fixing procedure of x variables for Ohio at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
				14	3,168	100.00	168.61		
				15	3,168	100.00	518.61		
				16	3,168	100.00	809.12		
				17	3,168	100.00	1,517.15		
				18	2,226	70.27	4,401.44		
				19	1,964	61.99	7,742.47		
				20	1,652	52.15	12,396.60		
OH	15	3,168	8,747	21	1,404	44.32	19,691.57	14	18
				22	1,258	39.71	23,179.93		
				23	1,105	34.88	23,877.97		
				24	1030	32.51	24,014.47		
				25	878	27.71	22,612.05		
				26	23	0.73	26,729.44		
				27	6	0.19	20,309.69		
				28	0	0.00	18,562.97		

Table 23: Results for the three-phase fixing procedure of x variables for Illinois at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
				13	2,168	66.40	684.12		
				14	2,020	61.87	1,040.60		
				15	1,850	56.66	1,468.91		
				16	1,612	49.37	2,218.98		
				17	1,436	43.98	3,613.57		
				18	1,298	39.75	5,533.23		
				19	1,111	34.03	8,136.93		
				20	985	30.17	8,144.64		
				21	861	26.37	8,164.84		
				22	700	21.44	8,563.19		
IL	17	3,265	8,728	23	578	17.70	8,186.91	13	13
				24	457	14.00	7,360.04		
				25	387	11.85	6,503.76		
				26	322	9.86	7,158.64		
				27	254	7.78	7,392.11		
				28	212	6.49	7,726.03		
				29	78	2.39	8,272.64		
				30	20	0.61	8,027.02		
				31	7	0.21	8,815.54		
				32	2	0.06	9,213.13		
				33	0	0.00	9,291.57		

Table 24: Results for the three-phase fixing procedure of x variables for Pennsylvania at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
PA	17	3,446	9,641	13	2,718	78.87	143.49	13	13
				14	2,626	76.20	217.91		
				15	2,547	73.91	239.56		
				16	2,486	72.14	291.27		
				17	2,426	70.40	423.18		
				18	2,375	68.92	418.02		
				19	2,328	67.56	521.03		
				20	2,267	65.79	696.30		
				21	2,189	63.52	983.01		
				22	2,109	61.20	1,012.36		
				23	2,035	59.05	1,176.42		
				24	1,964	56.99	1,481.35		
				25	1,893	54.93	1,959.91		
				26	1,827	53.02	2,143.54		
				27	1,747	50.70	2,779.22		
				28	1,634	47.42	2,720.94		
				29	1,481	42.98	3,287.12		
				30	1,376	39.93	3,524.22		
				31	1,279	37.12	4,118.11		
				32	1,197	34.74	5,121.10		
				33	1,138	33.02	5,446.57		
				34	1,084	31.46	5,507.93		
				35	1,037	30.09	5,914.00		
				36	999	28.99	5,716.39		
				37	960	27.86	6,874.42		
				38	917	26.61	7,244.41		
				39	872	25.30	8,366.82		
				40	823	23.88	9,227.90		
				41	760	22.05	9,544.13		
				42	687	19.94	10,715.47		
				43	598	17.35	12,717.48		
				44	502	14.57	13,374.11		
				45	415	12.04	14,510.65		
				46	339	9.84	15,324.53		
				47	277	8.04	16,008.34		
				48	212	6.15	16,928.79		
				49	147	4.27	19,051.57		
				50	58	1.68	24,463.49		
				51	13	0.38	28,949.15		
				52	0	0.00	32,020.69		

Table 25: Results for the three-phase fixing procedure of x variables for Texas at the tract level.

state	k	n	m	s	# fixed	% fixed	time	ℓ_s^{base}	ℓ_s^{fix}
TX	38	6,896	18,554	12	6,896	100.00	142.14	12	16
				13	6,896	100.00	120.82		
				14	6,896	100.00	386.70		
				15	6,896	100.00	1,644.88		
				16	6,848	99.30	5,534.52		
				17	5,089	73.80	25,584.68		
				18	3,306	47.94	62,748.45		
				19	2,603	37.75	86,569.55		
				20	1,868	27.09	124,357.55		
				21	1,514	21.95	146,101.26		
				22	1,212	17.58	162,068.40		
				23	939	13.62	155,364.26		
				24	716	10.38	153,727.57		
				25	518	7.51	158,371.49		
				26	328	4.76	164,372.57		
				27	189	2.74	175,875.28		
				28	86	1.25	180,701.38		
				29	37	0.54	188,155.75		
				30	14	0.20	191,318.99		
				31	3	0.04	192,414.65		