

# Drawing Competitive Districts in Redistricting

**Gabriel Chuang** ✉

Computer Science, Columbia University, New York, NY, USA

**Oussama Hanguir** ✉

Lyft, Inc., New York, NY, USA

Columbia University, New York, NY, USA

**Clifford Stein** ✉

Industrial Engineering and Operations Research, Columbia University, New York, NY, USA

---

## Abstract

In the process of redistricting, one important metric is the number of *competitive districts*, that is, districts where both parties have a reasonable chance of winning a majority of votes. Competitive districts are important for achieving proportionality, responsiveness, and other desirable qualities; some states even directly list competitiveness in their legally-codified districting requirements. In this work, we discuss the problem of drawing plans with at least a fixed number of competitive districts. In addition to the standard, “vote-band” measure of competitiveness (i.e., how close was the last election?), we propose a measure that explicitly considers “swing voters” - the segment of the population that may choose to vote either way, or not vote at all, in a given election. We present two main, contrasting results. First, from a computational complexity perspective, we show that the task of drawing plans with competitive districts is NP-hard, even on very natural instances where the districting task itself is easy (e.g., small rectangular grids of population-balanced cells). Second, however, we show that a simple hill-climbing procedure can in practice find districtings on real states in which *all* the districts are competitive. We present the results of the latter on the precinct-level graphs of the U.S. states of North Carolina and Arizona, and discuss trade-offs between competitiveness and other desirable qualities.

**2012 ACM Subject Classification** Theory of computation → Design and analysis of algorithms

**Keywords and phrases** Redistricting, Computational Complexity, Algorithms

**Digital Object Identifier** 10.4230/LIPIcs.FORC.2024.7

**Related Version** *Full Version:* <https://arxiv.org/abs/2404.10964>

**Funding** *Gabriel Chuang:* NSF Graduate Research Fellowship.

*Clifford Stein:* Research supported in part by NSF grant CCF-2218677 and ONR grant ONR-13533312, and by the Wai T. Chang Chair in Industrial Engineering and Operations Research.

## 1 Introduction

In the United States, *redistricting* is the task of geographically dividing a state into a fixed number of regions called *districts*, each of which elects one representative to a legislative body (such as the U.S. House of Representatives or a state legislature). The process is prone to various types of manipulation, collectively known as *gerrymandering*, in which parties draw districting maps that are optimized for particular outcomes. For example, a party may wish to maximize seats in which their preferred voters constitute a majority, protect their party’s incumbents, or force opposition-party incumbents to run against each other. This process often results in many districts that are uncompetitive, i.e., districts in which one party’s voters constitute such a large majority that voters are denied any meaningful choice and the winning party is effectively pre-determined.



© Gabriel Chuang, Oussama Hanguir, and Clifford Stein;  
licensed under Creative Commons License CC-BY 4.0

5th Symposium on Foundations of Responsible Computing (FORC 2024).

Editor: Guy N. Rothblum; Article No. 7; pp. 7:1–7:22

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

To combat gerrymandering, many quantitative and qualitative measures capturing various normative criteria have been proposed. These include notions of proportionality [18], responsiveness [26], partisan symmetry [30], typicality [15], stability under perturbation [17], and many others (e.g., [9], [10]; see [31] for a comparison of several partisan-based measures).

In this work, we focus on *competitive districts*: those where one expects elections to be close and highly contested, i.e., where the outcomes of future elections are not pre-determined by the geography of the district. There are three reasons we consider competitiveness to be of particular importance:

1. Several jurisdictions in the United States explicitly require competitiveness as a quality that their districting plans must satisfy. Colorado, for example, has a requirement that plans must “maximize the number of politically competitive districts,” where competitive is defined as “having a reasonable potential for the party affiliation of the district’s representative to change at least once between federal decennial censuses” [14]. See [15] for an overview.
2. Having competitive districts is key for the responsiveness of the plan. Responsiveness is a measure of how much a given change in popular vote translates into a change in the proportion of seats held by a particular party. For example, the 14-district Congressional map enacted in North Carolina in 2023 is highly non-responsive: the state could swing 5 points more Democratic and result in zero more Democratic-held seats, or 17 points more Republican and result in only one Republican pickup. This lack of responsiveness is a direct consequence of there being only one competitive seat of the fourteen.
3. Competitive elections are generally seen as promoting various positive civic qualities, such as voter engagement, high turnout, close attention to local issues, and others.

The number of competitive districts in the country has been shrinking rapidly. In 2020, only 45 of 435 districts have a Cook Partisan Voting Index between R+3 and D+3, down from 107 in 1999 [32]; Wasserman estimates that around half of the lost swing seats are due to changes to district boundaries, rather than “true” changes in electoral behavior (e.g. changing voter preferences, geographic polarization, etc.).

## 1.1 Contributions

Motivated by the desirable qualities of competitive districts, we study a version of the districting problem where the goal is to draw maps with at least some fixed number of competitive districts. In addition to the intuitive notion of competitiveness, where a district is considered competitive if recent elections have been decided by close margins (e.g., by 5% or less), we consider a characterization that relies on separately counting “swing voters” – that is, voters who have a reasonable chance of voting for either party, a formulation that aligns with the Colorado requirement of having districts with a “reasonable chance of being won by either party”. We show that the problem of maximizing swing districts is NP-hard for both our characterization and the standard model, even on instances where the underlying districting problem is polynomial-time solvable: that is, the hardness comes from the competitive-districts requirement, rather than the inherent hardness of balanced graph partitioning. Despite this, we show that a simple hill-climbing procedure can achieve a very high number of competitive districts without significantly sacrificing compactness, equal population, and other desirable qualities. We demonstrate the results on data from North Carolina and Arizona, showing that it is possible to make every district competitive (although we do not necessarily advocate for doing so). We also complement the NP-completeness results by giving restrictions that make the problem of maximizing swing districts more tractable.

## 2 Related Work

### 2.1 Competitiveness of districting plans

Among the plethora of proposed evaluation metrics for districting plans, Deford et al. [15] take a comprehensive look at various criteria that aim to operationalize competitiveness, including “evenness” (how close is the vote share to 50%?), “typicality” (how close is the vote share to the national or statewide average?), and “vote-band” metrics (does the vote fall within a fixed percentage of 50%, or the statewide average?). They observe that “there is no guarantee that it is even possible to construct a plan with a large number of ... districts [that fall within a given vote-band] while adhering to reasonable compactness and boundary preservation norms,” an idea we will expand on in this work. We will adopt the vote-band metric, because its binary nature allows us to easily express the problem of drawing competitive districts as a decision problem. In addition, we will propose another metric based on swing voters.

In the same work, Deford et al. [15] also conduct an extensive ensemble analysis of how many competitive districts arise in “typical” districtings, and present two hill-climbing algorithms for optimizing directly for competitive districts. We conduct similar experiments for both vote-band and swing-voter metrics, but use a randomized weighted scheme that incorporates compactness directly via the isoperimetric score.

Other works that explore the competitiveness of enacted and proposed plans largely use margins in recent elections as their measurement for competitiveness, including [14], which investigates plans for Colorado, and [20], which studies whether independent commissions tend to draw more competitive plans.

### 2.2 Computational complexity of redistricting

There has also been work exploring hardness of the districting problem in a computational complexity sense, including work on auditing for local deviating groups [24], on minimizing the margin of victory in non-geographically-bounded districts [29], and on describing classes of graphs for which various redistricting tasks are NP-hard [21]. Notably, even the balanced graph partition problem (i.e., drawing population-balanced districts, without any other restrictions) is NP-hard, even for planar graphs [5]. Given this, [25] show that the problem of drawing districts where each party wins at least  $c$  seats is NP-hard *even on instances where valid (contiguous and population-balanced) maps can be found in polynomial time*, by showing a reduction from Var-Linked Planar 3-SAT. We will adopt this structure for our hardness results: we will show reductions that create instances where population-balanced districts can be drawn in polynomial time but drawing competitive districts (both vote-band competitive and swing) is hard.

## 3 Preliminaries and Problem Formulation

### 3.1 Voting setting

Our setting consists of a set of voters distributed over a given geographic area (such as a U.S. state), represented by a (typically planar) graph  $G$  with  $n$  cells (nodes). The cells are some fixed geographic units (such as counties, precincts, etc.) and edges represent adjacency<sup>1</sup>.

---

<sup>1</sup> Depending on the jurisdiction, legal requirements may mandate either rook adjacency or queen adjacency.

## 7:4 Drawing Competitive Districts in Redistricting

We assume that there are only two parties in consideration, Party **A** and Party **B**; every voter may be either a partisan voter or a swing voter. Specifically, each cell  $c_i : i \in \{1, \dots, n\}$  has the following four quantities, all non-negative integers:

- $Pop_i \geq 0$  indicating the total population of the cell, which we abbreviate  $Pop_i$ ;
- $a_i$ , the number of voters that vote for Party **A**;
- $b_i$ , the number of voters that vote for Party **B**;
- optionally,  $s_i$ , the number of swing voters (who may vote either way).

### 3.2 Districtings and Competitive Districts

For a fixed  $d \in \{2, \dots, n-1\}$ , a  $d$ -*districting* is a partition of the cells of  $G$  into  $d$  disjoint subgraphs  $D_1, \dots, D_d$ , which we call *districts*. For any district  $D_j$ , we have:

$$Party_A(D_j) = A_j = \sum_{c_i \in D_j} a_i, \quad Party_B(D_j) = B_j = \sum_{c_i \in D_j} b_i, \quad Swing(D_j) = S_j = \sum_{c_i \in D_j} s_i.$$

Since every voter is either A, B, or swing, we have  $Pop(D_j) = Pop_j = A_j + B_j + S_j$ .

In general, a  $d$ -districting of a graph  $G$  is  $\varepsilon$ -valid if it satisfies the following constraints:

1. *Contiguity*, i.e., the subgraph induced by  $D_j$  must be connected for each  $j = 1, \dots, d$ . Thus, each district is contiguous, which is typically required by law.
2.  $\varepsilon$ -*Population-balance*, i.e.  $(1 - \varepsilon) \left( \frac{Pop(G)}{d} \right) \leq Pop(D_j) \leq (1 + \varepsilon) \left( \frac{Pop(G)}{d} \right)$  for each  $j = 1, \dots, d$ , which ensures that each district has approximately the same population, which is required under the ‘‘One Person, One Vote’’ rule [28]. We only consider  $\varepsilon < \frac{1}{6}$ , as  $\varepsilon \geq \frac{1}{6}$  allows one district to have *double* the population of another district. Different jurisdictions may require different values of  $\varepsilon$ .

#### 3.2.1 Competitiveness

We consider two notions that aim to capture the competitiveness of a given district.

First, we consider the intuitive notion that a district is competitive if the most recent election was decided by a very close margin - for example, 51% of votes cast for Party A and 49% cast for Party B. Deford et al. [15] call this ‘‘vote-band’’ competitiveness, named for the ‘‘band’’ of outcomes (for example, 45-55%) that the vote share should fall into in order for the election to be considered competitive<sup>2</sup>. This notion does *not* depend on counting swing voters separately, so, in this context,  $s_i = 0$  for all cells  $c_i$ .

► **Definition 1.** A district  $D_j$  is  $\delta$ -Vote-Band Competitive ( $\delta$ -VBC) iff  $\frac{A_j}{Pop_j}, \frac{B_j}{Pop_j} \in \left[ \frac{1}{2} - \delta, \frac{1}{2} + \delta \right]$ .

An alternative notion is that a district is competitive if the outcome depends on how the swing voters vote. In particular, elections can be both close and uncompetitive. For example, a district comprised of 40% Party A voters, 40% Party B voters, and 20% swing voters (who may vote either way) is likely to be highly competitive, whereas a district with 52% Party A voters, 47% Party B voters, and 1% swing voters is less likely to be competitive.

► **Definition 2.** A district  $D_j$  is Swing if  $S_j \geq |A_j - B_j|$ .

<sup>2</sup> Deford et al. consider both centering the vote band around 50% and centering it around the statewide or nationwide average. For simplicity, we only discuss the former.

This formulation more directly captures the notion that districts are competitive if there is a reasonable chance for either party to win. However, evaluation of this metric depends on having a reasonable estimate of the number of such voters; we discuss using statistical methods of Ecological Inference for this task in Section 5.1.2. In particular, while polling data is valuable for estimating voter transitions, we need this information at a precinct-by-precinct level to do redistricting analysis, and polls are mostly conducted at a national or statewide level.

This formulation also allows us to capture the fact that, due to partisan polarization, elections in the US are often determined by turnout, rather than vote-switching [19]. For example, we can count “unreliable partisan voters” (those who will either vote for Party A or stay home) as half of a “true” swing voter (those who may vote for Party A or Party B), as their decision will affect the eventual vote margin in their precinct, district, or state by half as much.

### 3.3 Maximizing Competitive Districts

► **Definition 3** ( $\delta$ -VBC-Max). *Given a graph  $G$  with  $n$  cells  $c_1, \dots, c_n$ ,  $d \in [2, n - 1]$ ,  $\varepsilon > 0$ , and  $\delta > 0$ , the  $\delta$ -VBC District Maximization Problem is to compute a  $\varepsilon$ -valid  $d$ -districting that maximizes the number of  $\delta$ -VBC districts.*

► **Definition 4** (Swing-Max). *Given a graph  $G$  with  $n$  cells  $c_1, \dots, c_n$ ,  $d \in [2, n - 1]$  and  $\varepsilon > 0$ , the Swing District Maximization Problem is to compute a  $\varepsilon$ -valid  $d$ -districting that maximizes the number of swing districts.*

## 4 Hardness results

Our main results in this section are to show that the  $\delta$ -VBC District Maximization Problem and the Swing District Maximization Problem are both NP-hard. On its own, deciding the existence of any  $\varepsilon$ -valid  $d$ -districting for a graph  $G$  is NP-hard, regardless of competitive districts. Therefore, we show that both  $\delta$ -VBC-Max and Swing-Max are hard, even on instances where:

- the underlying graph  $G$  is a grid,
- the number of districts  $d$  is 2, and
- there exists a polynomial-time computable  $\varepsilon$ -valid  $d$ -districting.

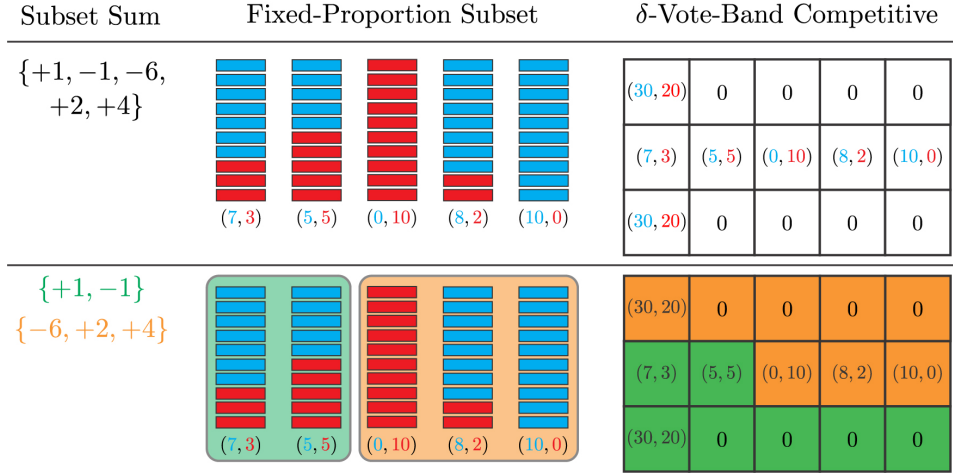
This setup is similar that of Kueng et al. [25], who show that drawing districtings where both parties have at least  $c$  seats is hard even on instances where balanced-population districting is easy.

► **Theorem 1** (Vote-Band Hardness). *For all positive  $\varepsilon < \frac{1}{6}$ ,  $\delta < \frac{1}{2}$ , the  $\delta$ -VBC  $\varepsilon$ -valid District Maximization Problem ( $\delta$ -VBC-Max( $\varepsilon$ )) is NP-hard, even on instances where  $G$  is a grid,  $d = 2$ , and  $\varepsilon$ -valid districtings (that are not  $\delta$ -VBC-maximal) are poly-time computable.*

► **Theorem 2** (Swing Hardness). *For all positive  $\varepsilon < \frac{1}{6}$ , the Swing  $\varepsilon$ -valid District Maximization Problem (Swing-Max( $\varepsilon$ )) is NP-hard, even on instances where  $G$  is a grid,  $d = 2$ , and  $\varepsilon$ -valid districtings (that are not Swing-maximal) are poly-time computable.<sup>3</sup>*

<sup>3</sup> This constraint of  $\varepsilon < \frac{1}{6}$  is extremely lenient: in practice, most districtings have population balance under 1%, and  $\varepsilon = \frac{1}{6}$  would allow one district to have *double* the population of another - which is certainly illegal.

## 7:6 Drawing Competitive Districts in Redistricting



■ **Figure 1** The chain of reductions from Subset Sum to FPSP to  $\delta$ -VBC-Max, for  $\delta = 0.1, \varepsilon = \frac{1}{6}$ . Given the Subset Sum instance  $T$  (left), we construct an instance of FPSP (center) where bin  $i$  has  $6 + t_i$  type-A units and  $4 - t_i$  type-B units. This induces a districting instance (right), where the central row has Party A and Party B voters corresponding to type-A and type-B units. The green and orange districts (where each district has population in  $(\frac{1}{2} \pm \varepsilon)pop_{total}$  and exactly 60% Party A voters) correspond to a FPSP partition (where each partition has exactly 60% type-A units), which in turn corresponds to a solution to the original Subset Sum instance.

We will prove both of these via a reduction from Subset Sum [22]. Specifically, we will reduce through an intermediate problem, the Fixed-Proportion Subset Problem.

### 4.1 The Fixed-Proportion Subset Problem

► **Definition 5** (FPSP( $\delta$ )). For any fixed positive  $\delta < \frac{1}{2}$ , the Fixed-Proportion Subset Problem (FPSP) takes a list of tuples  $S = [(a_1, b_1) \cdots (a_n, b_n)]$ , where:

$$\sum_{i=1}^n a_i = \left(\frac{1}{2} + \delta\right) \sum_{i=1}^n (a_i + b_i) \quad \text{and} \quad a_1 + b_1 = a_2 + b_2 = \cdots = a_n + b_n$$

The task is to find nonempty proper subsets  $S_1, S_2 \subsetneq [n]$  that partition  $[n]$ , such that

$$\sum_{i \in S_1} a_i = \left(\frac{1}{2} + \delta\right) \sum_{i \in S_1} (a_i + b_i) \quad \text{and} \quad \sum_{i \in S_2} a_i = \left(\frac{1}{2} + \delta\right) \sum_{i \in S_2} (a_i + b_i).^4$$

Intuitively, the Fixed-Proportion Subset Problem asks to partition a population of units with  $(\frac{1}{2} + \delta)$  type- $a$  units and  $(\frac{1}{2} - \delta)$  type- $b$  units, which are grouped into equal-size bins, into two subsets, where each subset is also exactly  $(\frac{1}{2} + \delta)$  type- $a$  and  $(\frac{1}{2} - \delta)$  type- $b$ .

► **Theorem 3** (FPSP Hardness). The Fixed-Proportion Subset Problem is NP-hard for all  $\delta \in [0, \frac{1}{2})$ .

**Proof.** Fix any  $\delta \in [0, \frac{1}{2})$ . Given an instance of Subset Sum  $T = [t_1, \cdots, t_{n-1}]$  with desired sum 0, let  $T' = [t_1, \cdots, t_{n-1}, t_n]$ , where  $t_n = -\sum_{i=1}^{n-1} t_i$ , so that  $\sum T' = 0$ .

<sup>4</sup> The existence of any  $S_1$  satisfying the first condition is sufficient to have  $S_2 = [n] \setminus S_1$  satisfy the second condition, but we explicitly write both for clarity.

We construct an instance of FPSP( $\delta$ ) as follows. Let  $c = -\frac{\min t_i}{\frac{1}{2} + \delta}$ . Our instance of FPSP will be

$$S = [(a_1, b_1) \cdots (a_{n+1}, b_{n+1})] \text{ where } a_i = t_i + \left(\frac{1}{2} + \delta\right) c \text{ and } b_i = c - a_i .$$

Each “bin” will have  $c$  units total, satisfying the second condition of FPSP. We can verify the first condition holds, using the fact that  $\sum t_i = 0$ :

$$\sum_{i=1}^n a_i = \sum_{i=1}^n \left( t_i + \left(\frac{1}{2} + \delta\right) c \right) = \sum_{i=1}^n t_i + \left(\frac{1}{2} + \delta\right) \sum_{i=1}^n c = 0 + \left(\frac{1}{2} + \delta\right) \sum_{i=1}^n (a_i + b_i) .$$

Now, suppose that  $S_1, S_2 \subsetneq [n]$  is a solution to FPSP( $\delta$ ) for this instance  $S$ . Without loss of generality, suppose  $n \notin S_1$ . We claim that  $\{t_i : i \in S_1\}$  is a solution to the original Subset Sum instance  $T$ :

$$\sum_{i \in S_1} a_i = \left(\frac{1}{2} + \delta\right) \sum_{i \in S_1} (a_i + b_i) \implies \sum_{i \in S_1} \left( t_i + \left(\frac{1}{2} + \delta\right) c \right) = \left(\frac{1}{2} + \delta\right) \sum_{i \in S_1} c \implies \sum_{i \in S_1} t_i = 0$$

◀

## 4.2 Reduction from FPSP to Competitive Districts

We prove Theorem 1 via a reduction from FPSP( $\delta$ ) to  $\delta$ -VBC-Max.

**Proof.** Fix  $\varepsilon$  and  $\delta$ . Let  $S = [(a_1, b_1) \cdots (a_n, b_n)]$  be an instance of FPSP( $\delta$ ), and let  $Z = \sum_{i=1}^n (a_i + b_i)$ . Our redistricting instance (“state”) for  $\delta$ -VBC-Max( $\varepsilon$ ) will be a  $3 \times n$  grid of cells  $C_{i,j}$ , where:

- The top left and bottom left cells  $C \in \{C_{1,1}, C_{3,1}\}$  will each have

$$\text{pop}(C) = \frac{1}{2}P, \quad \text{Party}_A(C) = \left(\frac{1}{2} + \delta\right) \frac{1}{2}P, \quad \text{Party}_B(C) = \left(\frac{1}{2} - \delta\right) \frac{1}{2}P ,$$

where  $P = \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} Z$ .

- All other cells in the first and third row have zero population<sup>5</sup>:  $\text{pop}(C_{1,\cdot}) = \text{pop}(C_{3,\cdot}) = 0$ .
- For each cell  $C_{2,i}$  in the second row (for  $i = 1, 2, \dots, n$ ),

$$\text{pop}(C_{2,i}) = (a_i + b_i) \quad \text{Party}_A(C_{2,i}) = a_i \quad \text{Party}_B(C_{2,i}) = b_i .$$

The total population of the state is  $\text{pop}_{total} = Z + P = \frac{\frac{1}{2} - \varepsilon + \frac{1}{2} + \varepsilon}{\left(\frac{1}{2} - \varepsilon\right)} Z = \frac{1}{\frac{1}{2} - \varepsilon} Z$ .

**Observation 0.** Since all cells  $C_{2,1 \dots n}$  have the same population, it is trivial to draw an  $\varepsilon$ -valid districting: one can assign the first row and half of the second row to  $D_1$ , and the rest to  $D_2$ . That is, one can district this instance in polynomial time, if competitiveness is not considered.

<sup>5</sup> One can easily modify the reduction to have all cells have nonzero population by multiplying all other cells by  $2(n+2)+1$  and having the first- and third-row cells have  $\text{pop}(C) = 2, \text{Party}_A(C) = 1, \text{Party}_B(C) = 1$ .

## 7:8 Drawing Competitive Districts in Redistricting

**Observation 1.** The two left “corner” cells  $C_{1,1}$  and  $C_{3,1}$  must be assigned to different districts. If they were assigned the same district  $D_i$  (along with at least one center-row cell  $c_j$ , required for connectivity), that district’s population would exceed  $(\frac{1}{2} + \varepsilon) \text{pop}_{total}$ :

$$\text{pop}(C_{1,1}) + \text{pop}(C_{3,1}) + \text{pop}(C_{2,j}) = P + (a_j + b_j) > P = \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} Z = \left(\frac{1}{2} + \varepsilon\right) \text{pop}_{total}.$$

**Observation 2.** Each district must contain at least one of the center-row cells  $C_{2,i}$ ; a “corner” cell alone is under the population bounds.

Besides these two cells, the structure of the state allows all other second-row cells to be assigned to either district while respecting contiguity. For example, one can assign all cells in the first row to district 1 and all cells in the third row to district 2.

**Observation 3.**  $\frac{1}{2} + \delta$  of the overall population is Party A voters:

$$\text{total Party}_A = \sum_{i=1}^n a_i + \left(\frac{1}{2} + \delta\right) \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} Z = \left(\frac{1}{2} + \delta\right) Z + \left(\frac{1}{2} + \delta\right) \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} Z = \left(\frac{1}{2} + \delta\right) \text{pop}_{total},$$

where the last equality holds because  $\text{pop}_{total} = \frac{1}{\frac{1}{2} - \varepsilon} Z$ .

Let  $\{D_1, D_2\}$  be a  $\varepsilon$ -valid  $\delta$ -Vote-Band-Competitive districting on this state. We claim that  $S_1 = \{i : C_{2,i} \in D_1\}, S_2 = \{i : C_{2,i} \in D_2\}$  is a valid solution to the FPSP( $\varepsilon$ ) instance.

For  $S_1, S_2$  to be a solution to FPSP( $\varepsilon$ ), we must show (a) both are nonempty (which follows from Observation 2 above), and (b)  $\sum_{i \in S_1} a_i = (\frac{1}{2} + \delta) \sum_{i \in S_1} (a_i + b_i)$ .

Since both districts’ margins fall in the  $\frac{1}{2} \pm \delta$  vote band, each one must have *exactly*  $\frac{1}{2} + \delta$  Party A voters; if (for example)  $D_1$  had  $< \frac{1}{2} + \delta$  Party A voters, then  $D_2$  would end up with  $> \frac{1}{2} + \delta$  Party A voters, falling outside the vote band. So,

$$\frac{1}{2} + \delta = \frac{\text{Party}_A(D_1)}{\text{Party}_A(D_1) + \text{Party}_B(D_1)} = \frac{(\frac{1}{2} + \delta) \frac{1}{2} P + \sum_{i \in S_1} a_i}{\frac{1}{2} P + \sum_{i \in S_1} (a_i + b_i)}.$$

Solving this equation for  $\sum_{i \in S_1} a_i$ , yields  $\sum_{i \in S_1} a_i = (\frac{1}{2} + \delta) \sum_{i \in S_1} (a_i + b_i)$  as required. Thus,  $S_1, S_2$  are a valid solution to the FPSP instance. ◀

The chain of reductions, using the Subset Sum instance  $T = \{1, -1, -6, 2, 4\}$  as an example, is shown in Fig. 1.

The proof for Swing district maximization is nearly identical. We present it in Appendix A.1.

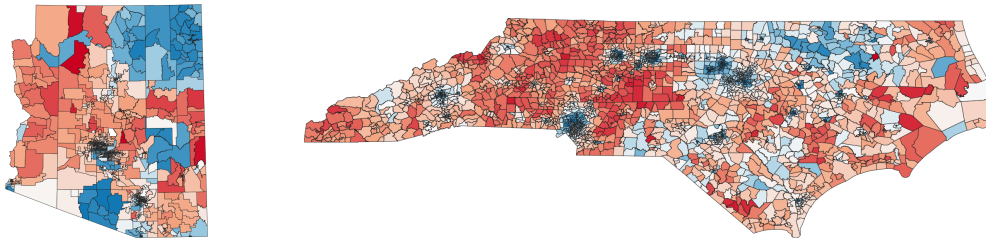
► **Corollary 4** (Hardness for more than two districts). *For any  $k, d$  with  $2 \leq k \leq d$ , it is NP-hard to generate an  $\varepsilon$ -valid  $d$ -districting plan where at least  $k$  districts are competitive (for either  $\delta$ -VBC or SWING).*

The proof for Corollary 4 is given in Appendix A.2.

### 4.3 Algorithms for Special Cases

In Appendix B, we present some special cases of graphs which admit polynomial time algorithms. Although these cases are somewhat artificial if expressed as standard redistricting instances, they may be of interest to those interested in districting-flavored problems on constrained structures, such as road networks or low-population grids.





■ **Figure 2** Precinct-level voting data for Arizona and North Carolina. Nodes are colored according to the margin in the 2020 presidential election.

#### 4.4 Discussion of Complexity Results

First, we re-emphasize that we have shown that the hardness is a result of the competitive district maximization, not the intrinsic hardness of the balanced partition problem, because the instances we construct can be partitioned to population balance  $\varepsilon = 0$  trivially.

There are reasons to see this reduction as somewhat more “natural” than previous hardness results, which often construct instances with strange features (for example, [25] constructs instances with that have long, tendril-like geography with many holes; [29] deals with a setting without geographic contiguity requirements; [24] creates instances where half of the nodes are connected to only one other node (i.e., there are many “donut” precincts, where one precinct entirely surrounds another)). In contrast, we construct instances where the underlying graph is a rectangular grid; many states (especially in the Midwest) have precinct graphs that have this approximate structure.

On the other hand, our reduction relies heavily on the hard cutoff at the edge of a desired “vote band” – that is, the hardness comes from distinguishing, in a binary way, between a district that is 59.99% Party A and one that is 60.01% Party A. In practice, the difference is likely to be negligible. Therefore, one may hope that we can, in practice, draw maps that significantly increase the number of competitive districts. We explain our approach to doing so in the subsequent section.

### 5 Algorithms, Heuristics, and Experiment Setup

Given the hardness of drawing competitive districting plans (as measured both in vote-band and swing-voter flavors), even on fairly constrained and natural varieties of graphs, one may wonder whether it is tractable to draw such districtings on real-world graphs. In this section, we argue that the answer is a definitive *yes*. In particular, we find that very simple hill-climbing procedure can yield reasonable districting plans that are highly competitive.

We propose a heuristic hill-climbing procedure based on making local moves called “single node flips”, and run several experiments on the U.S. states of North Carolina and Arizona: one set seeking to maximize  $\delta$ -Vote-Band-Competitive districts for  $\delta = 0.1, 0.05$  (corresponding to the thresholds used by the Center for Voting and Democracy for “landslides” (margin difference above 20%) and “competitive” (margin under 10%) [4]), and one seeking to maximize swing districts. In all cases, we fixed the allowable population deviation at  $\varepsilon = 5\%$  and required all districts to be contiguous.

#### 5.1 Data

We used precinct-level shapefiles and election data of the U.S. states of Arizona and North Carolina, two medium-sized states that have been highly competitive in recent elections. The

## 7:10 Drawing Competitive Districts in Redistricting

geographic data was collected and processed by the Metric Geometry and Gerrymandering Group [2], and the election data was obtained from the Redistricting Data Hub [3]. Specifically, the data includes the area, perimeter, and population of each precinct; voting history for the last several elections, and adjacency information. North Carolina has 2,650 precincts and fourteen Congressional districts. Arizona has 1489 precincts and nine Congressional districts. We chose these two states due to their extremely close statewide margins in recent elections, and their moderate size (for reasonable computational burden). They are shown in Figure 2.

### 5.1.1 Previous Election Votes for Vote-Band Competitiveness

Evaluating the vote-band competitiveness of a given districting plan depends on the votes cast in a past election. For these experiments, we used the results of the 2020 Presidential election, shown in Fig. 2.

### 5.1.2 Ecological Inference for Estimating Swing Voters

Evaluating whether a given district is swing or not depends on having estimates for the number of reliable voters of each party, and the number of swing voters. The task of estimating voter transitions from election data is a well-studied problem in Ecological Inference (EI). Given only top-level voting information (i.e., the number of votes cast for each candidate in two subsequent elections), the task of finding the number of voters who switched their vote from one election to the next (or who voted in one election and not the other) is highly underdetermined. As a result, EI techniques have no worst-case guarantees; however, they have been shown to perform well in practice. For an overview, see [23].

The task is made easier by the fact that we have the marginals (i.e., total votes cast) for each precinct in the state. `nsiphom` is a multi-iteration Linear Programming technique that takes advantage of this fact, using statewide homogeneity assumptions, developed by Pavia et al. [27]. It is available as an R package, which we used to estimate the inner cell values of the  $3 \times 3$  tables of the form shown in Fig. 3 (one table per precinct). For example, 71.30 is the estimate for the number of voters who voted for the Democratic candidate in 2012 but did not vote in 2016.

		2016 votes			Total
		Democratic	Republican	Nonvote/Other	
2012 votes	Democratic	407.45	9.24	71.30	488
	Republican	3.55	1583.69	73.76	1660
	Nonvote/Other	0.00	272.07	2713.93	2986
Total		410	1864	2859	

■ **Figure 3** The EI table for votes cast in Alamance County, North Carolina, Precinct 01. The inner cell values are estimated using the R package `nsiphom`, using the known marginal values. In this case, the estimated number of swing voters is  $3.55 + 9.24 + \frac{1}{2}(71.30 + 73.76 + 0.0 + 272.07) = 221.36$ .

We use the election results of the 2012 and 2016 elections to estimate swing voter counts and the results of the 2020 election to evaluate the final redistricting maps in order to keep estimation and evaluation metrics separate. For each precinct, we compute the final estimate for “swing voters” by summing the off-diagonal elements of the EI table, with the entries in a “Nonvote/Other” row or column halved. Intuitively, compared to a voter who switches their vote from Party A to Party B, a voter who merely goes from abstaining to voting (or from voting to abstaining) only changes the top-level margin by half as much.

## 5.2 Heuristic-based optimization procedure

In order to find districtings that maximize the number of competitive districts, we use a simple randomized greedy hill-climbing procedure based on repeatedly making “single node flips” (also called spin flips) that incrementally improve the plan when measured against some objective(s). These methods are extensively used in ensemble-based analyses of redistricting; see, for example, [13, 12].<sup>6</sup>

A *single-node-flip* is a step that flips a node that is on the boundary of two districts from one district to another, subject to some constraints. Specifically, a node  $u \in D_j$  is eligible to be flipped to district  $D_i$  (where  $i \neq j$ ) if (a)  $u$  is adjacent to some  $v \in D_i$ , (b)  $D_j \setminus \{u\}$  remains connected and within population bounds; and (c)  $D_i \cup \{u\}$  is within population bounds.

In each step, we randomly choose a single-node-flip  $(u, D_1, D_2)$  from the set of all valid flips, proportional to a score function  $J_i$ , which quantifies the “desirability” of the resulting districts  $(D_1 \setminus \{u\}, D_2 \cup \{u\})$ , compared to the original districts  $(D_1, D_2)$ , for particular objectives of interest:

$$J(u, D_1, D_2) = \exp \left( \sum_{\text{objectives } i} -w_i \left( J_i(D_1 \setminus \{u\}) + J_i(D_2 \cup \{u\}) - (J_i(D_1) + J_i(D_2)) \right) \right).$$

Here, low values for  $J_i$  are preferred; this weight function weights a flip more heavily the more it decreases the  $J_i$  values.<sup>7</sup> In this work, our weight includes two score functions: compactness and competitiveness (via swing districts or  $\delta$ -VBC districts).

We employ the widely-used measure of compactness, the *isoperimetric ratio* [8], defined as the perimeter squared divided by area. The more compact a district, the lower its isoperimetric ratio, and we define  $J_{iso}(D) = \frac{\text{perimeter}(D)^2}{\text{area}(D)}$ .

The vote-band competitiveness term in the score function prioritizes districts that are close to even (i.e., 50-50 margin), and further prioritizes districts where the margin is in the range  $\frac{1}{2} \pm \delta$ :

$$J_{VBC}(D) = \begin{cases} \left( \frac{\text{PartyA}(D)}{\text{Pop}(D)} - \frac{1}{2} \right)^2 & \text{if } \frac{\text{PartyA}(D)}{\text{Pop}(D)} \notin \frac{1}{2} \pm \delta, \\ \frac{1}{16} \left( \frac{\text{PartyA}(D)}{\text{Pop}(D)} - \frac{1}{2} \right)^2 & \text{if } \frac{\text{PartyA}(D)}{\text{Pop}(D)} \in \frac{1}{2} \pm \delta. \end{cases}$$

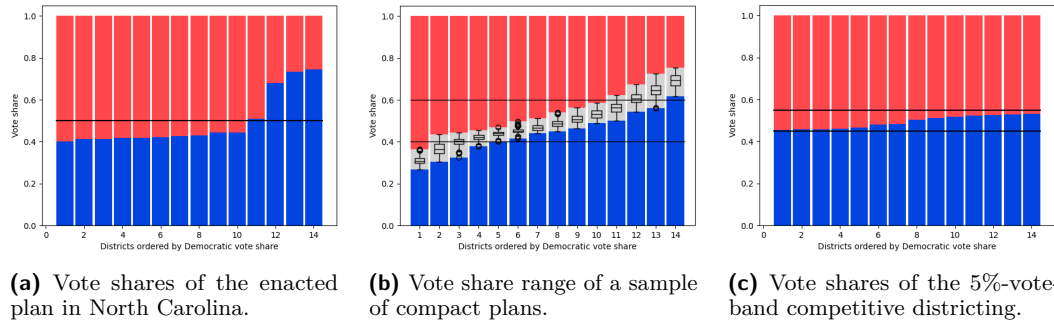
The swing term in the score function prioritizes having districts where the number of Party A voters plus half the swing voters are close to half of the total population, and further prioritizes districts where neither party’s reliable voters comprise more than half of the overall population:

$$J_{sw}(D) = \begin{cases} \left( \frac{1}{2} - \left( \frac{1}{2} \frac{\text{Swing}(D)}{\text{Pop}(D)} + \frac{\text{PartyA}(D)}{\text{Pop}(D)} \right) \right)^2 & \text{if } \frac{\text{PartyA}(D)}{\text{Pop}(D)} > \frac{1}{2} \text{ or } \frac{\text{PartyB}(D)}{\text{Pop}(D)} > \frac{1}{2}, \\ \left( \frac{1}{2} - \left( \frac{1}{2} \frac{\text{Swing}(D)}{\text{Pop}(D)} + \frac{\text{PartyA}(D)}{\text{Pop}(D)} \right) \right)^2 \cdot 0.8^2 & \text{if } \frac{\text{PartyA}(D)}{\text{Pop}(D)} < \frac{1}{2} \text{ and } \frac{\text{PartyB}(D)}{\text{Pop}(D)} < \frac{1}{2}. \end{cases}$$

<sup>6</sup> However, we are *not* doing an ensemble analysis; we will not attempt to sample from a measure or use the Metropolis-Hastings algorithm, as these works do. Rather, we are simply investigating the degree to which a direct optimization can achieve competitive districtings.

<sup>7</sup> This type of score function can be considered to be a “tempered choice” according to the measure  $\pi(D) = \exp \left( \sum_i -w_i J(D_i) \right)$ . This form of measure is commonly used in ensemble methods, e.g., in [7, 6].

## 7:12 Drawing Competitive Districts in Redistricting



■ **Figure 4** Explicitly considering competitiveness allows us to draw plans with significantly fewer safe districts than both the enacted plan and a sample of compact plans.

### 5.3 Run Parameters

We ran the hill-climbing procedure for 36,000 steps, restarting from a random initial state every 3,000 steps and taking the most compact plan with the maximal number of competitive districts. We used weights of  $w_{iso} = 3, w_{VBC} = 10^5$ , for VBC runs and  $w_{iso} = 3, w_{sw} = 10^5$  for the Swing runs. We implemented the procedure in Python using the `gerrychain` package [1]. With unoptimized, single-threaded Python code, the procedure takes about three hours to run on Intel Xeon Gold 6226 2.9 Ghz machines.

As a basis for comparison, we also ran the hill-climbing procedure for 40,000 steps, restarting every 200 steps, with  $w_{iso} = 3, w_{VBC} = 0, w_{sw} = 0$ , i.e., prioritizing only compactness. We logged the most-compact plan found in each 200-step interval. We will refer to this set of plans as the *compact sample* below.

## 6 Experimental Results

We find that the heuristics are extremely effective for constructing districting plans with a significant proportion of swing districts. In fact, we obtain districtings on North Carolina and Arizona where *every single* district is competitive (for 10%-vote-band, 5%-vote-band, and swing metrics).

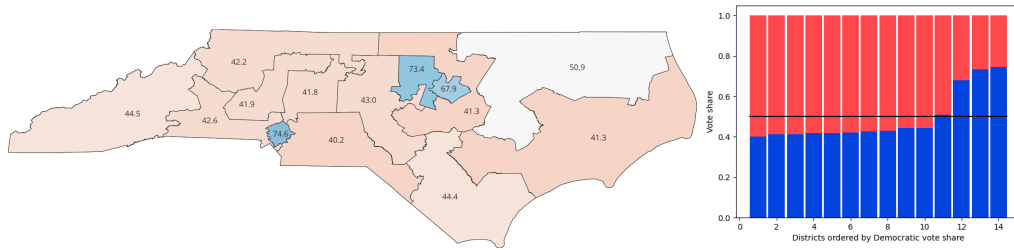
In Fig. 5, we display the most compact plan for each of  $\delta = 0.1, 0.05$ , and Swing, alongside the current enacted plan, for North Carolina. We also display the vote share of each district: that is, the percentage of Democratic and Republican (and Swing, for the last plot) voters in each district, based on votes cast in the 2020 presidential election. In Fig. 4, we compare the vote share distribution of the 5%-VBC plan to that of the enacted plan and of the compact sample.

We display the same results for Arizona in Fig. 6.

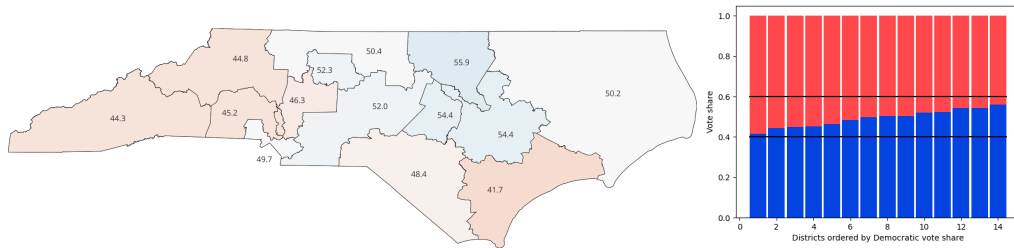
Notably, the VBC plans exhibit significantly improved responsiveness: the median district is extremely competitive (49.9% D-voting) and a small statewide swing in vote share would correspond to a larger change in number of seats won.

## 7 Discussion

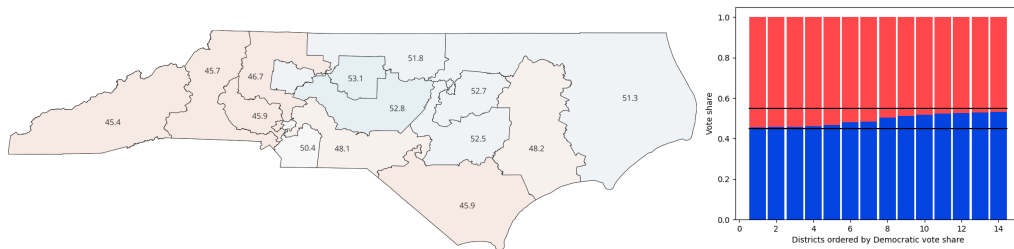
We explicitly do *not* present these plans as examples of *ideal* districting plans. Rather, we present them in contrast to the hardness results presented above, and to explore the consequences of fully prioritizing competitiveness as an objective. Although both swing district and vote-band-competitive district maximization is NP-hard (even if only seeking to make two out of  $d$  districts competitive), we explicitly show that it is tractable in practice,



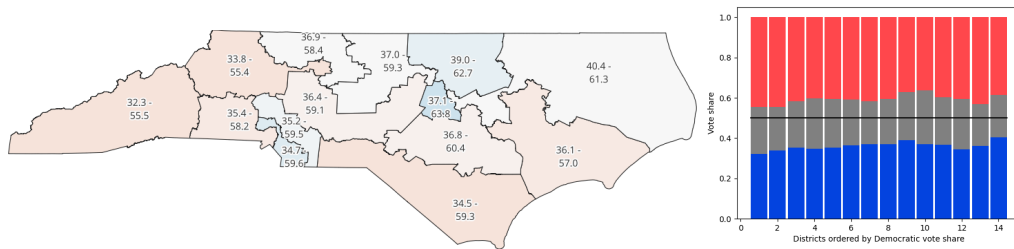
(a) The currently-enacted plan. Despite winning just over half of the two-party vote, Trump would have carried ten of the fourteen Congressional districts.



(b) In this plan, all districts are  $\delta$ -Vote-Band competitive for  $\delta = 10\%$ . That is, all districts have a Biden vote share between 40% and 60%. Note that Biden and Trump would have each carried seven of the fourteen districts: a result that is significantly more reflective of the fact that they won nearly the same number of votes.



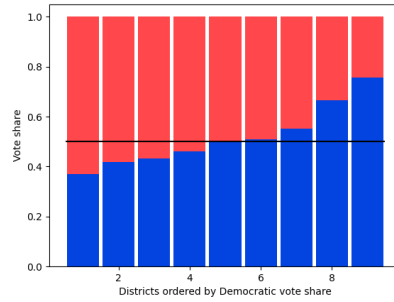
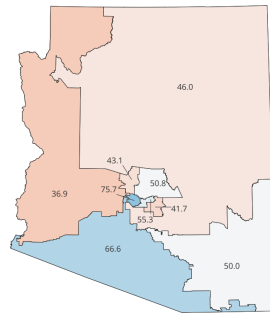
(c) In this plan, all districts are  $\delta$ -Vote-Band competitive for  $\delta = 5\%$ : all districts have a Biden vote share between 45% and 55%. Again, Biden and Trump would have each carried seven districts. However, the districts are noticeably less compact, with significantly contorted boundary shapes; the heavily Democratic areas of Charlotte and Raleigh-Durham are visibly “cracked” among many districts.



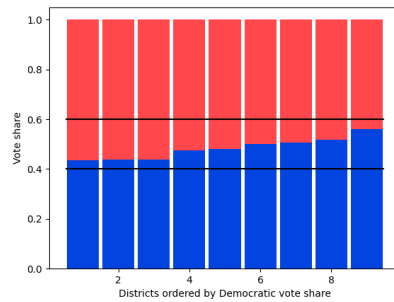
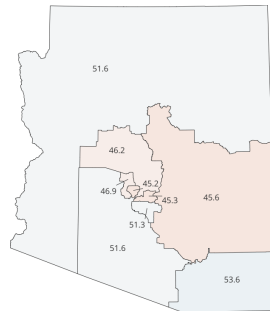
(d) In this plan, all districts are swing. The range of outcomes (ranging from all swing voters voting for Trump to all swing voters voting for Biden) is shown for each district.

■ **Figure 5** Our simple hill-climbing procedure successfully finds plans for North Carolina where all fourteen districts are competitive.

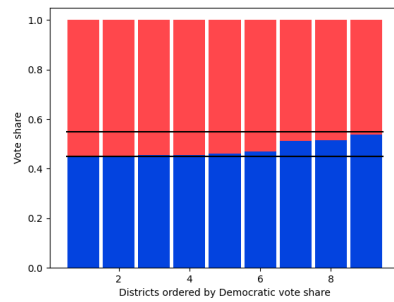
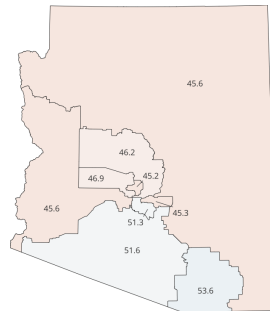
7:14 Drawing Competitive Districts in Redistricting



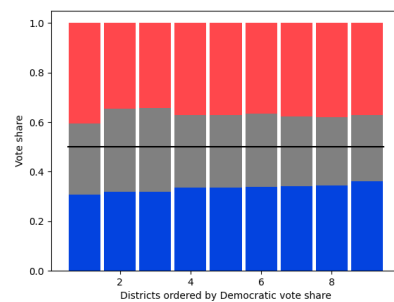
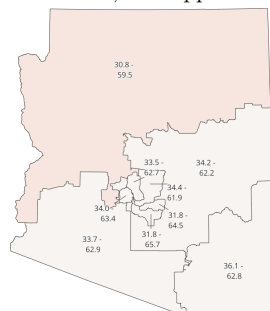
(a) The current enacted plan.



(b) A 0.1-Vote-Band-Competitive districting: all districts have a Biden vote share between 40% and 60%.



(c) A 0.05-Vote-Band-Competitive districting: all districts have a Biden vote share between 45% and 55%. Unlike in North Carolina, this appears to be achievable without significant loss of compactness.



(d) In this plan, all districts are swing. The range of outcomes (ranging from all swing voters voting for Trump to all swing voters voting for Biden) is shown for each district.

■ **Figure 6** Our hill-climbing procedure successfully finds plans for Arizona where all nine districts are competitive. Unlike in North Carolina, there is no significant sacrifice in compactness to achieve a  $\delta = 5\%$ -vote-band-competitive districting.

even achieving plans that make all  $d$  districts competitive. Thus, the point is to dissuade the reader from drawing the conclusion that “because drawing swing districts is NP-hard, attempting to do so is a lost cause”, or that “policymakers are exempt from drawing competitive districts because of computational intractability.” Instead, drawing competitive districts is very tractable in practice.

Notably, whereas the enacted plan “packs” the heavily-Democratic areas of Charlotte and Raleigh-Durham into as few districts as possible, maximizing competitive districts entails “cracking” those areas up into multiple districts. We observe this among all plans that achieved 14 competitive districts; in some sense, this is likely unavoidable given the political geography of the state (in which Democratic voters are heavily concentrated in small geographic areas).

Indeed, the *ideal* number of competitive districts is almost certainly not the *maximal* number. We leave the question of exactly *how many* districts should be competitive as a normative question (for discussion of the relationship between proportionality and competitive districts, see [18]); we simply present the result that it is tractable to achieve anywhere from 0 to  $d$  competitive districts on some real-world graphs.

In particular, while having *every* district in a state be competitive makes the makeup of the state’s Congressional delegation highly responsive (that is, small changes in vote share can result in large changes in topline number-of-seats-won), taking this to an extreme (for example, by enacting the plan in Fig. 5(c)) can result in a sharp decline in the proportionality of the results. For example, under the plan in Fig. 5(c), if one party won 55% of the votes statewide, they would sweep *all* of the congressional districts - certainly sending an unrepresentative delegation to represent the state.

We do not expect all states to admit fully-competitive districtings; indeed, states with remarkably homogeneously distributed electorates (such as Massachusetts) have been observed to be impossible to draw competitive districts on [16]. On the other extreme, some states may have voters that are so geographically polarized that drawing competitive districts may require splitting urban areas into unacceptably many districts. For example, Democratic voters in Pennsylvania and Illinois are so heavily concentrated in cities like Philadelphia and Chicago that maximizing competitive districts likely involves splitting them into over a dozen districts - likely an unacceptable result. We leave detailed investigation of these cases for future work.

## 8 Conclusion

We observe a very large gap between the theoretical intractability of drawing competitive districts (even on fairly natural instances) and the high performance of empirical heuristics on real instances. This is consistent with the literature on optimizing various metrics in redistricting, as well as with the fact that population-balanced districting itself is clearly achievable in reality while being complexity-wise infeasible in the worst case. We attribute the tractability on real instances to the fact that “close to optimal” is an acceptable substitute for “truly optimal” in the context of elections, where a large amount of variation and uncertainty is to be expected.

---

### References

- 1 GerryChain — GerryChain documentation. URL: <https://gerrychain.readthedocs.io/en/latest/>.
- 2 MGGG States. URL: <https://github.com/mggg-states>.

- 3 Redistricting Data Hub. URL: <https://redistrictingdatahub.org/>.
- 4 Dubious Democracy, 2003.
- 5 Konstantin Andreev and Harald Räcke. Balanced graph partitioning. In *Proceedings of the Sixteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '04, page 120–124, New York, NY, USA, 2004. Association for Computing Machinery. doi:10.1145/1007912.1007931.
- 6 Eric A. Autry, Daniel Carter, Gregory Herschlag, Zach Hunter, and Jonathan C. Mattingly. Multi-Scale Merge-Split Markov Chain Monte Carlo for Redistricting, August 2020. arXiv:2008.08054.
- 7 Eric A. Autry, Daniel Carter, Gregory J. Herschlag, Zach Hunter, and Jonathan C. Mattingly. Metropolized Multiscale Forest Recombination for Redistricting. *Multiscale Modeling & Simulation*, 19(4):1885–1914, January 2021. doi:10.1137/21M1406854.
- 8 Sachet Bangia, Christy Vaughn Graves, Gregory Herschlag, Han Sung Kang, Justin Luo, Jonathan C. Mattingly, and Robert Ravier. Redistricting: Drawing the Line, May 2017. arXiv:1704.03360.
- 9 Marion Campisi, Andrea Padilla, Thomas Ratliff, and Ellen Veomett. Declination as a Metric to Detect Partisan Gerrymandering. *Election Law Journal: Rules, Politics, and Policy*, 18(4):371–387, December 2019. doi:10.1089/e1j.2019.0562.
- 10 Marion Campisi, Thomas Ratliff, Stephanie Somersille, and Ellen Veomett. Geography and Election Outcome Metric: An Introduction. *Election Law Journal: Rules, Politics, and Policy*, 21(3):200–219, September 2022. doi:10.1089/e1j.2021.0054.
- 11 Tanima Chatterjee, Bhaskar DasGupta, Laura Palmieri, Zainab Al-Qurashi, and Anastasios Sidiropoulos. Alleviating partisan gerrymandering: Can math and computers help to eliminate wasted votes?, April 2018. doi:10.48550/arXiv.1804.10577.
- 12 Maria Chikina, Alan Frieze, and Jonathan C. Mattingly. Separating Effect From Significance in Markov Chain Tests. URL: <https://www.tandfonline.com/doi/full/10.1080/2330443X.2020.1806763>.
- 13 Maria Chikina, Alan Frieze, and Wesley Pegden. Assessing significance in a Markov chain without mixing | PNAS. *PNAS*, 2017.
- 14 Jeanne Clelland, Haley Colgate, Daryl DeFord, Beth Malmskog, and Flavia Sancier-Barbosa. Colorado in context: Congressional redistricting and competing fairness criteria in Colorado. *Journal of Computational Social Science*, 5(1):189–226, May 2022. doi:10.1007/s42001-021-00119-7.
- 15 Daryl DeFord, Moon Duchin, and Justin Solomon. A Computational Approach to Measuring Vote Elasticity and Competitiveness, September 2020. URL: <https://www.tandfonline.com/doi/full/10.1080/2330443X.2020.1777915>.
- 16 Moon Duchin, Taissa Gladkova, Eugene Henninger-Voss, Ben Klingensmith, Heather Newman, and Hannah Wheelen. Locating the Representational Baseline: Republicans in Massachusetts. *Election Law Journal: Rules, Politics, and Policy*, 2018.
- 17 Moon Duchin, Tom Needham, and Thomas Weighill. The (homological) persistence of gerrymandering, July 2020. doi:10.48550/arXiv.2007.02390.
- 18 Moon Duchin and Gabe Schoenbach. Redistricting for Proportionality. *The Forum*, 20(3-4):371–393, December 2022. doi:10.1515/for-2022-2064.
- 19 Thomas G. Hansford and Brad T. Gomez. Estimating the Electoral Effects of Voter Turnout. *American Political Science Review*, 104(2):268–288, May 2010. doi:10.1017/S0003055410000109.
- 20 John A. Henderson, Brian T. Hamel, and Aaron M. Goldzimer. Gerrymandering Incumbency: Does Nonpartisan Redistricting Increase Electoral Competition? *The Journal of Politics*, 80(3):1011–1016, July 2018. doi:10.1086/697120.
- 21 Takehiro Ito, Naoyuki Kamiyama, Yusuke Kobayashi, and Yoshio Okamoto. Algorithms for gerrymandering over graphs, May 2021. URL: <https://www.sciencedirect.com/science/article/pii/S0304397521001894>.



- 22 Jon Kleinberg and Eva Tardos. *Algorithm Design*. Pearson, 2006.
- 23 André Klima, Paul W. Thurner, Christoph Molnar, Thomas Schlesinger, and Helmut Küchenhoff. Estimation of voter transitions based on ecological inference: An empirical assessment of different approaches. *AStA Advances in Statistical Analysis*, 100(2):133–159, April 2016. doi:10.1007/s10182-015-0254-8.
- 24 Shao-Heng Ko, Erin Taylor, Pankaj Agarwal, and Kamesh Munagala. All Politics is Local: Redistricting via Local Fairness. *Advances in Neural Information Processing Systems*, 35:17443–17455, December 2022.
- 25 Richard Kueng, Dustin G. Mixon, and Soledad Villar. Fair redistricting is hard. *Theoretical Computer Science*, 791:28–35, October 2019. doi:10.1016/j.tcs.2019.04.004.
- 26 John F. Nagle. What Criteria Should Be Used for Redistricting Reform? *Election Law Journal: Rules, Politics, and Policy*, 18(1):63–77, March 2019. doi:10.1089/e1j.2018.0514.
- 27 Jose M. Pavia and Rafael Romero. Improving Estimates Accuracy of Voter Transitions. Two New Algorithms for Ecological Inference Based on Linear Programming. *Sociological Methods & Research*, page 00491241221092725, May 2022. doi:10.1177/00491241221092725.
- 28 J. Douglas Smith. *On Democracy's Doorstep: The Inside Story of How the Supreme Court Brought "One Person, One Vote" to the United States*. Farrar, Straus and Giroux, June 2014.
- 29 Ana-Andreea Stoica, Abhijnan Chakraborty, Palash Dey, and Krishna P. Gummadi. Minimizing Margin of Victory for Fair Political and Educational Districting, September 2019. doi:10.48550/arXiv.1909.05583.
- 30 Samuel S.-H. Wang. Three Tests for Practical Evaluation of Partisan Gerrymandering. *Stanford Law Review*, 68(6):1263–1322, 2016.
- 31 Gregory S. Warrington. A Comparison of Partisan-Gerrymandering Measures. *Election Law Journal: Rules, Politics, and Policy*, 18(3):262–281, September 2019. doi:10.1089/e1j.2018.0508.
- 32 David Wasserman. Realignment, More Than Redistricting, Has Decimated Swing House Seats, April 2023. URL: <https://www.cookpolitical.com/cook-pvi/realignment-more-redistricting-has-decimated-swing-house-seats>.

## A Further Proofs

### A.1 Hardness of Swing Competitiveness

In this section, we prove Theorem 2, which we restate here for convenience:

► **Theorem 2 (Swing Hardness).** *For all positive  $\varepsilon < \frac{1}{6}$ , the Swing  $\varepsilon$ -valid District Maximization Problem (Swing-Max( $\varepsilon$ )) is NP-hard, even on instances where  $G$  is a grid,  $d = 2$ , and  $\varepsilon$ -valid districtings (that are not Swing-maximal) are poly-time computable.<sup>8</sup>*

The proof is nearly identical to the proof of Theorem 1.

**Proof.** Fix  $\varepsilon$  and let  $\delta = 0$ . Let  $S = [(a_1, b_1) \cdots (a_n, b_n)]$  be an instance of FPSP(0), and let  $Z = \sum_{i=1}^n (a_i + b_i)$ . Our redistricting instance (“state”) for Swing-Max( $\varepsilon$ ) will be a  $3 \times (n + 2)$  grid of cells  $C_{i,j}$ , where:

- All cells in the first and third row have zero population:  $\text{pop}(C_{1,\cdot}) = \text{pop}(C_{3,\cdot}) = 0$ .
- For second-row cells  $C_{2,i}$  for  $i = 1, 2, \dots, n$ ,

$$\text{pop}(C_{2,i}) = (a_i + b_i) \quad \text{Party}_A(C_{2,i}) = a_i \quad \text{Swing}(C_{2,i}) = b_i$$

<sup>8</sup> This constraint of  $\varepsilon < \frac{1}{6}$  is extremely lenient: in practice, most districtings have population balance under 1%, and  $\varepsilon = \frac{1}{6}$  would allow one district to have *double* the population of another - which is certainly illegal.

## 7:18 Drawing Competitive Districts in Redistricting

- Finally, for second-row cells  $C_{2,j}$  for  $j = n + 1, n + 2$ ,

$$\text{pop}(C_{2,j}) = \frac{1}{2}P, \text{ where } P = \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} (1 + Z)$$

$$\text{Party}_A(C_{2,j}) = \frac{1}{4}P \quad \text{Swing}(C_{2,j}) = \frac{1}{4}P$$

The total population of the state is

$$\text{Pop}_{\text{total}} = 0 + Z + 2 \left( \frac{1}{2}P \right) = Z + P = \frac{\frac{1}{2} - \varepsilon + \frac{1}{2} + \varepsilon}{\left(\frac{1}{2} - \varepsilon\right)} Z + \frac{\frac{1}{2} + \varepsilon}{\left(\frac{1}{2} - \varepsilon\right)} = \frac{1}{\frac{1}{2} - \varepsilon} Z + \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon}$$

Let  $\{D_1, D_2\}$  be a  $\varepsilon$ -valid Swing districting on this state.

The following observations still hold, from the proof of Theorem 1:

**Observation 0.** One can district this instance in polynomial time.

**Observation 1.** Cells  $C_{2,n+1}$  and  $C_{2,n+2}$  must be assigned to different districts. Without loss of generality, let  $C_{2,n+1} \in D_1$ .

**Observation 2.** Each district must contain at least one of  $C_{2,i}$  for  $i \in [n]$ .

Claim:  $S_1 = \{i : C_{2,i} \in D_1\}, S_2 = \{i : C_{2,i} \in D_2\}$  is a valid solution to the FPSP( $\varepsilon$ ) instance.

For  $S_1, S_2$  to be a solution to FPSP( $\varepsilon$ ), we must show (a) both are nonempty (which follows from Observation 2 above), and (b)  $\sum_{i \in S_1} a_i = \left(\frac{1}{2} + \delta\right) \sum_{i \in S_1} (a_i + b_i)$ .

$\frac{1}{2}$  of the overall population is Party A voters:

$$\begin{aligned} \frac{\text{total Party}_A}{\text{total population}} &= \frac{\sum_{i=1}^n a_i + \left(\frac{1}{2}\right) \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} (1 + Z)}{\sum_{i=1}^n (a_i + b_i) + \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} (1 + Z)} \\ &= \frac{\left(\frac{1}{2}\right) \left(\sum_{i=1}^n (a_i + b_i) + \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} (1 + Z)\right)}{\sum_{i=1}^n (a_i + b_i) + \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} (1 + Z)} = \frac{1}{2} \end{aligned}$$

Since both districts' margins are exactly  $\frac{1}{2}$ , each one must have *exactly*  $\frac{1}{2}$  Party A voters. So,

$$\begin{aligned} \frac{1}{2} &= \frac{\text{Party}_A(D_1)}{\text{Party}_A(D_1) + \text{Party}_B(D_1)} \\ &= \frac{\left(\frac{1}{2}\right) \frac{1}{2}P + \sum_{i \in S_1} a_i}{\frac{1}{2}P + \sum_{i \in S_1} (a_i + b_i)} \\ \left(\frac{1}{2}\right) \left(\frac{1}{2}P + \sum_{i \in S_1} (a_i + b_i)\right) &= \left(\frac{1}{2}\right) \frac{1}{2}P + \sum_{i \in S_1} a_i \\ \left(\frac{1}{2}\right) \frac{1}{2}P + \left(\frac{1}{2}\right) \sum_{i \in S_1} (a_i + b_i) &= \left(\frac{1}{2}\right) \frac{1}{2}P + \sum_{i \in S_1} a_i \\ \sum_{i \in S_1} a_i &= \left(\frac{1}{2}\right) \sum_{i \in S_1} (a_i + b_i) \end{aligned}$$

as required. Thus,  $S_1, S_2$  are a valid solution to the FPSP instance. ◀

## A.2 Hardness for an arbitrary number of districts

In this section, we prove Corollary 4:

► **Corollary 4** (Hardness for more than two districts). *For any  $k, d$  with  $2 \leq k \leq d$ , it is NP-hard to generate an  $\varepsilon$ -valid  $d$ -districting plan where at least  $k$  districts are competitive (for either  $\delta$ -VBC or SWING).*

**Proof.** We reduce from an instance with  $d = 2$ . Fix  $d', k$  with  $2 \leq k \leq d'$ . We simply add  $d' - 2$  cells to the instance. Each cell will have the maximal allowable district population.  $k - 2$  of the districts will have half Party A voters and half Party B voters (which makes that singleton district competitive under either Swing or  $\delta$ -VBC definitions), and the remainder will be entirely Party A voters (i.e., uncompetitive).

Any valid districting with  $k$  competitive districts must have the added cells as singleton districts, of which  $k - 2$  will be competitive. The other two districts must be a competitive districting of the original instance. ◀

## B Polynomial Algorithms and Approximations

In this section, we consider special cases of the Swing District Maximization problem in which either the graph  $G$  has a special structure (line, bounded degree tree, etc), the population constraint is relaxed, or the districts satisfy an additional structure.

### B.1 Line

We start by considering the case where  $G$  is the line graph on  $n$  vertices  $c_1, \dots, c_n$ , such that there is an edge between  $c_i$  and  $c_{i+1}$  for  $1 \leq i \leq n - 1$ . Satisfying the connectivity constraint in this case is easy; simply make sure that every district has only consecutive cells  $c_j, c_{j+1}, \dots, c_i$ . A district is feasible if it's connected and satisfies the population constraints. We can find the optimal  $d$ -districting by solving a dynamic program. For fixed  $d$ ,  $i \in \{1, \dots, n\}$  and  $k \leq d$ , let  $M(i, k)$  denote the maximum number of swing districts among all valid  $d$ -districtings of the subgraph induced by  $c_1, \dots, c_i$ . We show that

$$M(i, k) = \max_{\substack{1 \leq j < i \\ (c_{j+1}, \dots, c_i) \text{ feasible}}} \left\{ M(j, k - 1) + \mathbb{1}\{(c_{j+1}, \dots, c_i) \text{ is swing}\} \right\}, \quad (1)$$

and

$$\begin{aligned} M(i, 0) &= -\infty && \forall i \in \{1, \dots, n\} \\ M(i, 1) &= \mathbb{1}\{(c_1, \dots, c_i) \text{ is swing}\} && \text{if } (c_1, \dots, c_i) \text{ feasible, } \forall i \in \{1, \dots, n\} \\ M(i, 1) &= -\infty && \text{if } (c_1, \dots, c_i) \text{ is not feasible, } \forall i \in \{1, \dots, n\} \end{aligned}$$

Equation (1) holds since to get the maximum number of swing districts in a  $k$ -districting of  $c_1, \dots, c_i$ , we need to decide on the cell  $c_{j+1}$  that limits the last district from the left. Once the last district  $\{(c_{j+1}, \dots, c_i)\}$  is fixed, we need to pick the remaining  $k - 1$  districts from  $c_1, \dots, c_j$ . Checking if a subset of cells satisfies population constraints and induces a swing district can be done in  $O(n)$  time. Therefore, we can fill the  $n \times d$  entries of the matrix  $M(i, k)$  top down from left to right. The optimal solution is stored in  $M(n, d)$ . We have the following lemma.

► **Lemma 6.** *If  $G$  is a line on  $n$  cells, we can compute the optimal  $d$ -districting in  $O(n^2d)$  time.*

## B.2 Bounded-Degree Trees with Districts of Bounded Depth

In this subsection, we consider bounded-degree trees with the additional assumptions that districts need to have bounded depth, that is, the distance between every two cells in the same district must be less than a parameter  $d > 0$ . Consider the graph  $G$  to be a tree with  $n$  vertices and a maximum degree  $\Delta$ . Let  $\varepsilon$  be the population tolerance. We require that in a valid districting, all districts must run for a depth of at most  $d$  (i.e., the diameter of every district is less than  $d$ ).

Let  $\mathcal{D}(v, \varepsilon, d)$  be the set of  $\varepsilon$ -valid districts with depth at most  $d$  that are rooted at  $v$ .

▷ **Claim 7.**

$$|\mathcal{D}(v, \varepsilon, d)| \leq 2^{\Delta^d}$$

*Proof.* The number of vertices at a distance less or equal than  $d$  from the root  $v$  is less than  $\Delta^d$ . For every one of these vertices and every  $D \in \mathcal{D}(v, \varepsilon, d)$ ,  $D$  can either contain the vertex or not. ◁

When  $d = O(1)$  and  $\Delta = O(1)$ , we propose a polynomial time dynamic program to solve the swing district maximization problem. For a fixed  $d$  and  $k \leq d$ , let  $M(v, k)$  denote the maximum number of swing districts in a  $\varepsilon$ -valid  $k$ -districting for the subtree of  $G$  that is rooted at  $v$ . To get the optimal districting, we need to first fix the district that  $v$  will belong to in  $\mathcal{D}(v, \varepsilon, d)$ . Because of the depth constraint on districts, the number of possible choices is bounded by  $2^{\Delta^d}$ . After we fix the district of  $v$ , we need to choose the roots of the remaining  $k - 1$  districts.

Let  $G(v)$  be the subtree of  $G$  that is rooted at  $v$ . Let  $D \in \mathcal{D}(v, \varepsilon, d)$ , and let  $\mathcal{R}(D)$  be the roots of the subtrees of  $G(v) \setminus D$ . If  $|\mathcal{R}(D)| > k - 1$ , then clearly we cannot assign all the remaining cells to districts. Therefore we need  $0 < |\mathcal{R}(D)| \leq k - 1$ . Furthermore, to compute the remaining  $k - 1$  districts, we need to start from the roots in  $\mathcal{R}(D)$  such that, every vertex in  $\mathcal{R}(D)$  will give rise to at least one district. To decide how many district every tree rooted at a vertex in  $\mathcal{R}(D)$  needs to have, we assign a number  $\ell(u) \in \{1, \dots, k - 1\}$  for every  $u \in \mathcal{R}(D)$ , such that  $\sum_{u \in \mathcal{R}(D)} \ell(u) = k - 1$ , and the subtree rooted at  $u$  contains  $\ell(u)$  out of the remaining  $k - 1$  districts. This gives rise to the following dynamic program

$$M(v, k) = \max_{\substack{D \in \mathcal{D}(v, \varepsilon, d) \\ |\mathcal{R}(D)| \leq k-1 \\ \sum_{u \in \mathcal{R}(D)} \ell(u) = k-1}} \left\{ \mathbb{1}\{D \text{ is swing}\} + \sum_{u \in \mathcal{R}(D)} M(u, \ell(u)) \right\}$$

If  $r$  is the root of  $G$ , the  $M(r, d)$  will contain the optimal number of swing districts. In order to get  $M(r, d)$ , we need to top-down fill  $O(nk)$  entries of  $M$ . To fill an entry  $M(v, k)$ , we have to choose a district  $D \in \mathcal{D}(v, \varepsilon, d)$  and an assignment  $u \mapsto \ell(u)$  for  $u \in \mathcal{R}(D)$  such that  $\sum_{u \in \mathcal{R}(D)} \ell(u) = k - 1$ .

▷ **Claim 8.** Once a district  $D \in \mathcal{D}(v, \varepsilon, d)$  is fixed, the number of possible assignments  $u \mapsto \ell(u)$  for  $u \in \mathcal{R}(D)$  such that  $\sum_{u \in \mathcal{R}(D)} \ell(u) = k - 1$  is less than  $k^{2^{\Delta^d}}$ .

*Proof.* Similarly to the proof of Claim 7, we can show that  $|\mathcal{R}(D)| \leq 2^{\Delta^d}$ . The number of positive assignments  $u \mapsto \ell(u)$  such that  $1 \leq \ell(u) \leq k - 1$  is less than  $k^{2^{\Delta^d}}$ . ◁

The combination of Claim 7 and 8 show that every entry  $M(v, k)$  can be computed in  $O(2^{\Delta^d} k^{2^{\Delta^d}})$  time given that we already know the previous entries. We therefore have the following lemma.

► **Lemma 9.** *If  $G$  is a tree with a maximum degree  $\Delta$ , and districts can have at most a depth of  $d$ , we can compute the optimal  $d$ -districting in  $O(n2^{\Delta^d} k^{1+2^{\Delta^d}})$  time.*

### B.3 Convex Districtings of Grid Graphs

In this subsection, we assume that the graph  $G$  is an  $m \times n$  grid. If we require all the districts to be  $x$ -convex, that is, if two cells  $c_1$  and  $c_2$  of the same row are assigned to the same district, then all the cells between  $c_1$  and  $c_2$  of that same row are also assigned to that district. This case encompasses a compactness constraint since convexity has been used in gerrymandering studies as a measure of compactness to examine how redistricting reshapes the geography of congressional districts [11].

► **Theorem 5.** *Let  $G$  be an  $m \times n$  grid, and let  $P = \text{Pop}(G)$  be the total population on  $P$ . There exists an algorithm for computing an  $x$ -convex valid  $d$ -districting of  $G$  with maximum swing districts, with running time  $(Pm)^{O(d)}$ . In particular, the running time is polynomial when the total population is polynomial and the total number of partitions is a constant.*

**Proof.** Let  $D_1, \dots, D_k$  be an  $x$ -convex  $d$ -districting of  $G$ . For any  $i \in \{1, \dots, n\}$ , let  $C_i$  be the  $i$ -th column of  $G$ . We observe that for all  $i \in \{1, \dots, n\}$ , and for all  $j \in \{1, \dots, d\}$ , we have that  $D_j \cap C_i$  is either empty, or consists of a single rectangle of width 1. Let  $\mathcal{C}_i$  be the set of all *contiguous* partitions of  $C_i$  into exactly  $d$  (possibly empty) segments, each labeled with a unique integer in  $\{1, \dots, d\}$ . We further define

$$\begin{aligned}\alpha_{i,j} &= \text{PartyA}(D_j^* \cap (C_1 \cup \dots \cup C_i)) \\ \beta_{i,j} &= \text{PartyB}(D_j^* \cap (C_1 \cup \dots \cup C_i)) \\ \gamma_{i,j} &= \text{Swing}(D_j^* \cap (C_1 \cup \dots \cup C_i)),\end{aligned}$$

For each column  $i \in [n]$ , and each district  $j \in [d]$ ,  $\alpha_{i,j}$  (resp.  $\beta_{i,j}$ ,  $\gamma_{i,j}$ ) denote the number of Party A (resp. Party B, swing) voters in district  $D_j$  between column 1 and column  $i$ .

We can enumerate all the possible solutions starting from the first column and moving to the right as follows. For each  $i \in \{1, \dots, n\}$ , let  $I_i = \mathbb{N}^{3d} \times \mathcal{C}_i \times [m]^d$ . Let

$$X_i = (\alpha_{i,1}, \beta_{i,1}, \gamma_{i,1}, \dots, \alpha_{i,d}, \beta_{i,d}, \gamma_{i,d}, \mathcal{Z}_{i,i}) \in I_i,$$

where  $\mathcal{Z}_i$  is a  $d$ -partition of the column  $C_i$  and  $i = \{F_{i,1}, \dots, F_{i,d}\}$  is a collection of the forbidden indices for every district in the next column  $C_{i+1}$ , to ensure the connectivity of the districts as well as the  $x$ -convexity constraints.

If  $i = 1$ , we say that  $X_i$  is feasible if  $\mathcal{Z}_1 = \{\emptyset, \dots, \emptyset\}$  and, for all  $j \in \{1, \dots, d\}$ , the set  $\mathcal{Z}_1 = \{Z_1, \dots, Z_d\}$  satisfies

$$\alpha_{i,j} = \text{PartyA}(Z_{i,j}), \quad \beta_{i,j} = \text{PartyB}(Z_{i,j}), \quad \text{and} \quad \gamma_{i,j} = \text{Swing}(Z_{i,j})$$

If  $i > 1$ , we say that  $X_i$  with  $\mathcal{Z}_i = \{Z_{i,1}, \dots, Z_{i,d}\}$  is feasible if the following holds, there exists some  $X_{i-1} = (\alpha_{i-1,1}, \beta_{i-1,1}, \gamma_{i-1,1}, \dots, \alpha_{i-1,d}, \beta_{i-1,d}, \gamma_{i-1,d}, Z_{i-1,i-1}) \in I_{i-1}$  such that

## 7:22 Drawing Competitive Districts in Redistricting

$$\alpha_{i,j} = \alpha_{i-1,j} + \text{PartyA}(Z_{i,j}), \quad \beta_{i,j} = \beta_{i-1,j} + \text{PartyB}(Z_{i,j}), \quad \text{and} \quad \gamma_{i,j} = \gamma_{i-1,j} + \text{Swing}(Z_{i,j}), \quad (2)$$

$$\mathcal{Z}_{i,j} \cap_{i-1,j} = \emptyset, \quad (3)$$

$$\text{If } Z_{i,j} \not\subset Z_{i-1,j} \text{ then } F_{i,j} = (Z_{i-1,j} \setminus Z_{i,j}) \cup_{i-1,j}, \text{ else } F_{i,j} =_{i-1,j} \quad (4)$$

The constraint (2) simply states that the voter populations in columns  $1, \dots, i$  are equal to the the voter populations in columns  $1, \dots, i-1$  plus the voters from partition  $\mathcal{Z}_i$ . Constraint (3) ensure that the partition of the  $i$ -th column has to respect  $x$ -convexity and not include any of the forbidden cells from  $_{i-1}$ . The last constraint (4) shows how to update the forbidden cells for the next column  $i+1$ . If for a district  $D_j$ , the indices of the rows added from column  $i$  to  $D_j$  is included in the set of rows of  $D_j$  from column  $i-1$ , then there is no need to add any other forbidden row for the next column. If however,  $\mathcal{Z}_{i,j} \not\subset \mathcal{Z}_{i-1,j}$ , that means that the column  $i$  does not “transfer” the rows of  $D_j$  from column  $i-1$  to column  $i$ , then the “non-transferred” rows have to be forbidden in the next column.

For each  $i \in \{1, \dots, n\}$  we inductively compute the set of all feasible  $X_i \in I_i$ . This can be done in time  $(Pm)^{O(d)}$  where  $P$  is the total population of the map. ◀