Little House (Seat) on the Prairie: Compactness, Gerrymandering, and Population Distribution

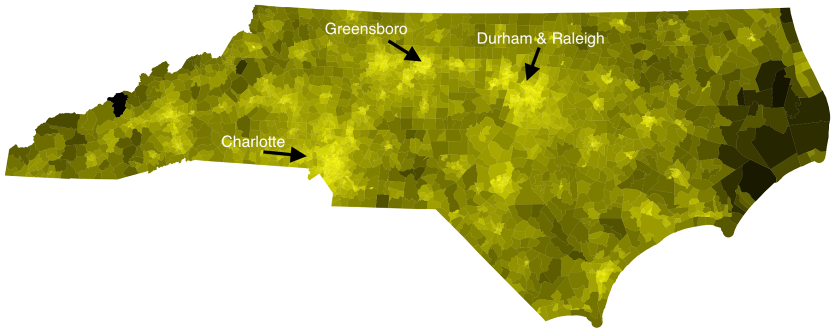
Allan Borodin[a], Omer Lev[b], Nisarg Shah[a], Tyrone Strangway[b]

[a] *University of Toronto, Ontario, Canada*
[b] *Ben-Gurion University of the Negev, Beersheba, Israel*

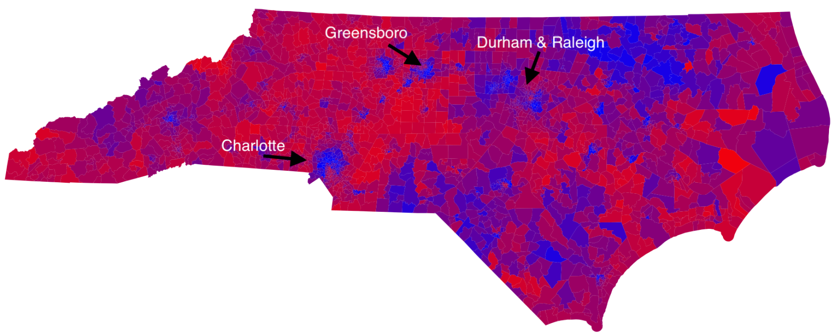
Abstract

Gerrymandering is the process of creating electoral districts for partisan advantage, allowing a party to win more seats than what is reasonable for their vote. While research on gerrymandering has recently grown, many issues are still not fully understood such as what influences the degree to which a party can gerrymander and what techniques can be used to counter it. One commonly suggested (and, in some US states, mandated) requirement is that districts be "geographically compact". However, there are many competing compactness definitions and the impact of compactness on the gerrymandering abilities of the parties is not well understood. Also not well understood is how the growing urban-rural divide between supporters of different parties impacts redistricting.

We develop a modular, scalable, and efficient algorithm that can design districts for various criteria. We confirm its effectiveness on several US states by pitting it against maps "hand-drawn" by political experts. Using real data from US political elections we use our algorithm to study the interaction between population distribution, partisanship, and geographic compactness. We find that compactness can lead to more fair plans (compared to implemented plans) and limit gerrymandering potential, but there is a consistent asymmetry where the party with rural supporters has an advantage. We also propose and explore a definition of compactness, one based around creating political districts where the residents all are similar with respect to their housing density. A priori, it is not clear if such a plan should help combat gerrymandering, or actually inflame the issue. Finally, we show there are plans which are fair from a partisan perspective, but they are far from optimally geographically compact.



(a) NC population density



(b) NC voting intention

Figure 1: Top, the natural log of population density in NC (population divided by area) according to the 2010 census. More yellow hue indicates higher density, while black indicates lower density. Bottom, the partisan vote share of each precinct in the 2012 presidential election; more red favours the Republican party, more blue favours the Democratic party. The three largest metropolitan areas, by population, are indicated on both figures.

1. Introduction

In many democracies, politicians are elected to represent the people of particular geographic areas, called districts.[1] There is no global aggregation of votes over the whole country, and instead voters within a district pick a winner from the alternatives vying to represent their district. Political power is based on the number of districts won by each party. This method of decision-making, using bottom-up structures, is not unique to countries, and can be seen in organizations (e.g., universities reaching decisions by approving them at the

Email addresses: `bor@cs.toronto.edu` (Allan Borodin), `omerlev@bgu.ac.il` (Omer Lev), `nisarg@cs.toronto.edu` (Nisarg Shah), `tyrone@cs.toronto.edu` (Tyrone Strangway)

[1]Many countries use different names for these, such as constituencies or ridings. In this paper, we will use the term "district".

departmental level, and if enough departments support, at the Faculty level, etc.), and other structures where sub-unit divisions make sense.

How voters are partitioned into these districts directly affects the makeup of the legislative body. The partitioning is often governed by hard constraints. For example, most jurisdictions require that the districts be geographically connected (with certain exceptions) and have roughly equal populations. In addition to connectivity and population balance constraints, there are many competing goals when designing a districting plan [1]. One could prioritize not breaking up communities of interest, such as those with a shared culture and history.[2] It may also be desirable to be as compatible as possible with established city and county boundaries, a consideration studied by Wheeler and Klein [2]. Another reasonable goal would be to obtain geographically compact regions (a goal enshrined in some US states' laws and regulations[3]). A less defensible goal is *gerrymandering*: designing districts for partisan gain, i.e., creating districts which help a particular party gain a number of seats beyond its popular support. Such a goal is termed gerrymandering after Massachusetts governor Elbridge Gerry, who approved a supposedly salamander-shaped district for such motives in the early 19th century.

While many countries have tried to take district drawing out of political hands, partisan districting is still a common, legal practice in some countries, most evidently in the US congressional system. In the US, following every 10-year census, state legislatures decide their new federal congressional and state legislative districts, and partisan concerns are often part of the consideration [3]. For example, in the 2020 federal election in North Carolina, a state accused of gerrymandering (partially overturned by courts [4]), the Democratic party received 49.96% of the vote and won five districts; the Republican party received 49.41% of the vote and won eight districts. Of course, gerrymandering is just one reason for problematically formed districts – prior to the Great Reform Bill of 1832, British district boundaries had remained static for hundreds of years. Due to changing population patterns, this resulted in "rotten boroughs" – districts containing very few (e.g., single digit) voters. That being said, in the United States, strict requirements on population balance between districts would prevent such situations. As noted above, due to the many competing goals, even with clearly stated goals it is not clear what is fair or optimal when it comes to non-partisan redistricting (see Wasserman [1] for further discussion and a comparison of objectives). Indeed, it may be impossible to satisfy all these goals simultaneously.

Parallel to the political partisan redistricting process there is a more complex, ongoing population-wide process, where a person's neighbourhood is correlated with their political leanings. As Figure 1 shows, in North Carolina voters of the Democratic party tend to cluster in dense urban centres, while

[2]Regarding ethnic minorities, this is required by the US' Voting Rights Act of 1965.

[3]For example, California's constitution states, in article XXI, "districts shall be drawn to encourage geographical compactness".

voters of the Republican party are spread out in the surrounding rural regions. We will explore and quantify this density and partisan relationship later on in North Carolina and several other states. This urban-rural divide in voting intentions is the norm in the United States [5], Britain [6], and in several European countries [7, 8]. In democracies around the world voters are "reorganizing" themselves for various economic and social reasons, creating this striking divide. Along with anger over lagging rural economies, the divide has been blamed for the resurgence of populist politicians and the success of movements like the United Kingdom leaving the European Union [9].

Our contribution. Our work explores aspects of both of these processes – the immediate partisan one and the process of population dynamics. In the first part of our work (Section 5), we introduce our automated redistricting procedure, which is flexible and can be used to design plans for various objectives, both partisan and nonpartisan. To prove the utility of our algorithm, we compare its performance against hand-drawn plans from election experts. We also explore (Section 6) the social contribution of our algorithm.

Once we show the power of our algorithm, we begin using it to understand the interplay between population distribution, geographic compactness constraints, and political power. Our algorithm allows us to examine possible requirements that have been suggested as a means to mitigate or eliminate gerrymandering. In particular, we study the impact of a compactness requirement. In Section 7, a few compactness measures are considered and we see that in the US, the more rural party (Republicans) still consistently outperforms the more urban party (Democrats). Moreover, we see that this advantage is robust *even in a non-gerrymandered, geographically-compact plan.* This advantage is not due to political gaming of the division process, but rather due to the geographic spread of each party's supporters. That being said, we show the existence of "fair" but not ideally compact plans. In addition, we introduce a novel metric for compactness, not one based on geography, but instead on population density. Taking inspiration from the idea of keeping like minded communities together, we create plans which place people of similar housing density into districts with each other. How the urban-rural divide and districts designed in such a way would impact outcomes is not obvious.

In Section 8 we examine how compactness constraints affect gerrymandering possibilities. We show that demanding stringent compactness constraints reduces the ability of parties to reach extreme gerrymanders. However, in most cases, the compactness requirement allows for relatively greater rural-party gerrymandering. Indeed, under the most stringent compactness constraints, the urban party sometimes cannot even achieve its vote proportion.[4]

[4]This is a significantly expanded version of a paper published in AAMAS 2022 [10]. Many details were added, including detailed descriptions of the algorithm, as well as new metrics, such as the density metric, and new simulations run, allowing for more wide-ranging analysis.

2. Related Work

Political districting, and in particular gerrymandering, has long been the examined in the humanities and social sciences. There have been legal discussions [11, 12, 13], and recently the US Supreme Court ruled [14] partisan gerrymandering cannot be addressed by federal courts. Gerrymandering has long been studied by historians [15, 16] and political scientists [17, 18, 19, 20].

From a technical perspective there is a long history of trying to detect gerrymandering by proposing and comparing various scores in potential plans [3, 21, 22, 23]. Recently, the AI community has become interested in redistricting. Pegden et al. [24] proposed treating the redistricting process as a repeated "cut-and-choose" game, while Bachrach et al. [25] measure how much districts can distort voter representation. Slightly more further afield, there has been research on allowing agents (or voters) to move between districts [26, 27, 28], and research on geographical manipulation of locating polling stations, influencing voters' willingness to travel to their location [29].

Most related to our work is automated redistricting, which originally was proposed as a solution to gerrymandering by Vickrey [30]. A popular automated redistricting technique uses Markov Chains to generate a sequence of related districts (see Fifield et al. [23], DeFord et al. [31] for two types of chains). These chains allow us to see where values from a plan, such as various measures of partisanship or compactness, fall within the distribution of a random ensemble of them. Outlier values may be a sign of poor district design or gerrymandering. One such analysis was used to help with the 2018 Pennsylvania redistricting [32]. Another line of automated redistricting work focuses not on generating ensembles of plans, but instead designing one optimized plan. Within the AI community many published automated methods violate basic legal requirements such as population balances [33, 34]; or are unable to scale to real-life sized problems [35]. Some other methods that produce legal plans for real data are only able to design plans for objectives based on linear combinations of individual district scores [36]. These shortcomings are not surprising since finding an optimal gerrymandering, or even certain graph partitions (a prerequisite for any division process), is known to be NP-hard in various settings [34, 37, 26]. For a recent survey on various automated redistricting methodologies we refer the reader to Becker and Solomon [38].

As discussed, the growing urban-rural divide and its impact on redistricting is an inspiration for our work. Commentators [39, 40] have argued this divide is amplifying the effects of partisan gerrymandering. Using simulated data, Borodin et al. [35] argue the rural party can stretch its vote share more effectively when gerrymandering than the urban party – although at extreme levels of the divide the opportunity to gerrymander is limited. Bishop [5] blames US cultural cleavages and population moves for this divide, while Rodden [41] goes further, arguing that the urban-rural divide is a fundamental disadvantage to the urban party in almost every scenario, though his argument boils down to the advantage of rural areas having more urban party supporters than vice versa.

3. Model and Background

We examine gerrymandering with a graph-theoretic formulation. We shall use US-oriented terminology (states, precincts, etc.), but the formulation represents most geographic districting settings. A state is an undirected graph $G(V, E)$, and each node $v \in V$ represents a precinct, a small geographic region where votes are tallied.[5] An edge $(u, v) \in E$ represents that precincts u and v share a physical boundary. Creating a districting plan requires partitioning G into K vertex-disjoint subgraphs $G_1, \cdots, G_K$ (the districts). We will denote the precincts of district i (i.e., the vertices of G_i) as $V(G_k)$. The number of districts K is extrinsically determined (in the US, by a census every 10 years). In all of our analysis we limit our focus to two parties: the rural party (in the US, Republicans (R)), and the urban party (in the US, Democrats (D)). While there are other minor parties, and even some independent politicians who have held office, the United States is effectively a two party system, and the influence of other parties is minimal.

3.1. Legal Requirements for Redistricting

For $v \in V$ let n_v be the number of people who live in precinct v. Let $N = \sum_{v \in V} n_v$ be the total number of people in the state. There are two widely accepted requirements for legal districts in the US and elsewhere:

Contiguity For each $k \in [K]$, G_k must form a connected subgraph of G. In the real world, this translates to being able to walk (or swim) from any point in the district to any other point in the district without crossing into another district.

Population balance-δ Given $\delta > 0$, for each $k \in [K]$,

$$1 - \delta \leq \frac{\sum_{v \in V(G_k)} n_v}{N/K} \leq 1 + \delta.$$

The exact value of δ required in the U.S. changes between states (and judicial decisions). Informally, the criteria is that districts should be as near equal-sized in population as possible [42]. We take $\delta = 0.005$, so that the maximum population deviation between any two districts is at most 1% of the state's population. This is the legal requirements in some states, and a far tighter constraint than implemented by many previously proposed automated redistricting methods.

These two constraints, even strictly enforced, still allow for an exponential number of plans in the number of nodes and districts. This is far too many to enumerate, and as we will see it allows for a diverse set of plans all expressing different combinations of goals such as partisanship, fairness, and compactness.

[5] In the US, census block data is more fine grained, but there is no voting information at this level, so it is not useful for our needs.

3.2. Measuring Partisanship

There are many ways of measuring how partisan a district is. Often, the first choice is what historical data we use. In the United States there are elections for various levels of government and offices. Unless otherwise stated we will use presidential election data when measuring how partisan a hypothetical district is[6]. For a particular election e and party p we let $n_v^{p,e}$ be the number of people who live in precinct v and vote for party p in election e. We will omit e when the context is obvious.

While not their only goal, a primary one of political parties is winning as many districts as they can. The party with the most voters in a district is typically said to win that district. For example, if $\sum_{v \in V(G_k)} n_v^{D,President-2012} > \sum_{v \in V(G_k)} n_v^{R,President-2012}$, we say the Democrats win district k according to the 2012 presidential vote. If the inequality is reversed, we say the Republicans win the district in that election.

When actually designing a district for partisan reasons a party would most likely not be satisfied that they would have hypothetically won it by 1 vote based on previous election data. Instead they would almost certainly design districts with a "safe" margin of victory. What is considered safe is debatable, and later on we will explore the effects of different margins of victory. For now we say that for a particular election e with two parties p_1 and p_2 and margin of victory $\tau \in \mathbb{R}_{\geq 0}$, p_1 wins district k if $\sum_{v \in V(G_k)} n_v^{p_1,e} > \sum_{v \in V(G_k)} n_v^{p_2,e} + \tau$. If $\sum_{v \in V(G_k)} n_v^{p_2,e} > \sum_{v \in V(G_k)} n_v^{p_1,e} + \tau$, we say p_2 won district k. If neither party has a margin of victory of at least τ we call district k a tossup.

We finally note that measuring partisanship using only one previous election is not a requirement. Sometimes we will use a composite, or even a function of the composite, of previous elections. In Section 5.2 we calculate the probability of winning a hypothetical district, based on historic vote totals and outcomes using the 2012 and 2016 presidential elections. In Section 5.4 we take a composite of various elections for various levels of government. There can be an argument that averaging previous elections is more robust. That is, they make a better predictor of future elections than any single election. However, there is a certain ad-hoc nature to these composites, and we are unable to find any results which formally, or informally, argue their advantage. Thus we only use these composites when comparing our work against previous works on redistricting which used said composites. Outside of these comparisons, we stick to using single elections, making as few assumptions as possible.

3.3. Uniform Swing and Proportionality

At several points in this work we will want to measure how much a particular plan deviates from the "fair outcome". Formally, in the ideal fair plan, the fraction of districts won by each party should match – as closely as possible –

[6] Because of non party influences, such as a congressional candidate's popularity and history with their district, presidential data is often considered less locally-biased as a measure of partisanship.

its fraction of overall votes. That is, want to develop metrics which measure how much a plan deviates from what is proportionally fair.

The most obvious way of measuring the deviation from what is proportionally fair would be taking the difference between the fraction of votes won and the fraction of seats won for a fixed election, or even for some composite of elections. However, measuring how proportional a plan is based on a fixed data point is not very robust to changes in public opinion. To address this, the *uniform swing* model [17] is widely used. In this model, hypothetical elections are generated starting from a baseline election, or composite of elections, by shifting the vote shares of the parties. Specifically, the vote share of a given party is increased or decreased by an equal amount in every district. The fraction of districts won by each party is then measured in these hypothetical elections in order to measure the amount by which proportionality would likely be violated if vote shares change in the future. See Figure 2 for an example of uniform swing in WI using 2016 presidential election data with the 2011 implemented plan.

Formally, the swing model is seeded with an election e (or some composite of elections), these are the dots in Figure 2. To measure a uniform swing of $t \in [-1, +1]$ for a particular party p, in every district k we take $p's$ vote fraction (in $[0, 1]$) for that district and add t to it. We then measure what fraction of districts p has a majority in after a swing of t. Note a uniform swing of $+t$ ($-t$) for party p is the equivalent of a uniform swing of $-t$ ($+t$) for the other party. Figure 2 shows what happens after a +5% uniform swing for the Ds in WI using the 2016 presidential election data with the 2011 implemented plan. In this particular example the Ds would have won two districts under the 2016 presidential election. In their closest two losses the Ds had a 44% and 47% share of the vote, so a +5% uniform swing would turn one more district into a win for them.

There are several metrics that use the uniform swing model to measure the partisan bias in a given plan. We are interested in the *partisan bias* score [43]. This value measures the vertical displacement of the swing curve from the point $(1/2, 1/2)$.[7] Intuitively, the partisan bias measures the divergence from the idea that "half the votes should translate to half the seats". More generally, we can measure the vertical displacement from any point (a, a) for $a \in [0, 1]$. We introduce a robust version of this metric. Fixing a line segment $[l, r]$ ($l, r \in [0, 1], l < r$), we measure the average vertical distance from the swing curve to the line $y = x$ over this line segment. We use $[0.45, 0.55]$ or $[0.4, 0.6]$ as the reasonable ranges (i.e., the vote shares of both parties are between 45% and 55% or 40% and 60%).[8] The 45° line in the $[0.4, 0.6]$ range is shown in green in Figure 2. There are two ways to measure the distance between a party's swing curve ($s(x), x \in [0, 1]$) and a line segment $[l, r]$. The first is a signed version,

[7] We can take the distance of the swing curve of either party as both distances are guaranteed to be identical, see below.

[8] Most presidential elections fall within the smaller range, and almost every presidential election falls within the larger range.

$$\frac{\int_l^r (s(x) - x)dx}{r - l} \tag{1}$$

measuring on average how much higher or lower the swing curve is over the proportional line. A positive (negative) value for the *signed partisan bias* indicated this party, over the range of reasonable vote shares, can expect more (fewer) seats than what is proportionally fair. Alternatively we could take the unsigned version,

$$\frac{\int_l^r |s(x) - x|dx}{r - l} \tag{2}$$

which measures the average deviation from proportionally fair. The *unsigned partisan bias*, tells us how much does the plan deviate from "an α fraction of the vote share should translate into an α fraction of the seats".

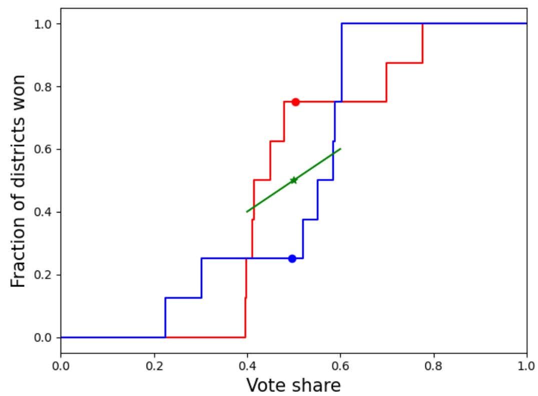
Both of these measures provide important information. For a fixed party p and its swing curve we quantify its partisan advantage (disadvantage), over $[l, r]$, by how positive (negative) Equation 1 is. That is, there may be portions $[l, r]$ where p's swing curve is above the green line (p is getting more than what is fair) and portions where it is under the green line (p is getting less than what is fair). Equation 1 tells us if the advantage or disadvantage is more prominent, and by how much. On the other hand, Equation 2 tells us, over $[l, r]$, on average how disproportionate p's outcomes are, its total advantage plus total disadvantage. That is, Equation 2 tells us how much, in both directions, p deviates from what is proportionally fair.

The reader will notice that for any fixed plan Equation 2 is always at least as large as the absolute value of Equation 1. While these measures are similar, and often correlated, they can differ. For example, consider a plan where a vote share of $50\% + \epsilon$ ($50\% - \epsilon$) results in winning (losing) each district. For any symmetric range about 50% vote share, this plan has the best possible score for Equation 1 and the worst possible score for Equation 2. Thus, it need not be that for a fixed election the optimal plan for Equation 1 bears any resemblance to the optimal plan for Equation 2.

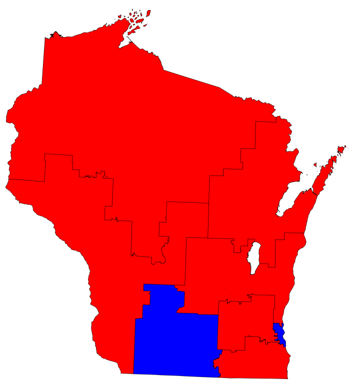
For any $t < 0.5$, the fraction of districts won by one party with $0.5 - t$ vote share is exactly one minus the fraction of districts won by the other party with $0.5+t$ vote share. Hence, for a symmetric range around the 0.5 vote share point, their swing lines are mirrors of each other about the point $(0.5, 0.5)$. Thus for both parties, the value of Equation 1 is identical in magnitude (but opposite in sign), and the value of Equation 2 will be identical. As mentioned, we only consider our two measures for symmetric ranges about the point $(0.5, 0.5)$.

3.4. Compactness Definitions

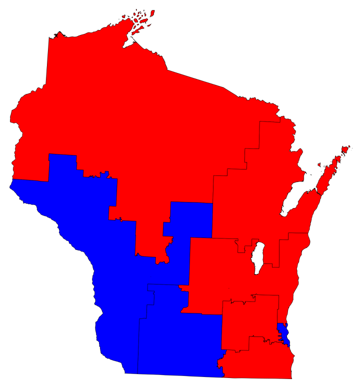
There are several accepted ways of measuring how compact a particular plan is. Some are geography based, some population based, and some are a mixture of the two. In this work we use two common geographic measures, the Polsby-Popper and Convex Hull scores. We also use a custom compactness



(a) Uniform swing in WI



(b) WI 2011 election outcome



(c) WI 2011 election outcome with D swing

Figure 2: Top row, uniform swing for $R(D)$ in red (blue), in the 2016 presidential election in WI using the implemented plan. Vertical axis shows the fraction of districts won; horizontal axis the vote fraction. The dots on the party curves indicate the actual election outcome (0 swing). The green line is the range of proportional outcomes on the range $[0.4, 0.6]$. A green star marks the point $(1/2, 1/2)$. Bottom row, WI districts implemented in 2011 (R wins in red; D wins in blue). Left, the data using 2016 presidential election data; right, the same figure, but after a $+5\%$ uniform swing for the Ds.

score which blends geography and population, designed by 538, a large and well-regarding website analyzing elections (with a US focus). We also use a custom geographic score from Dave's Redistricting App (DRA), a popular open-source redistricting tool. This score blends existing geographic scores, such as Polsby-Popper. Finally we use a population density measure, which is agnostic to the shape of districts, which we designed.

Let us formally define each of the compactness metrics we will use. For ease of notation, we define the following functions for a district i with subgraph G_i and polygon P_i: Let the district's area be $A(P_i)$, let the length of the perimeter be $L(P_i)$, and let the geometric centre point be $M(P_i)$[9]. These measures can be defined for any arbitrary $2D$ polygon (not necessarily just those that belong to a district). Let the straight line distance between two points $a, b \in \mathbb{R}^2$ be $d(a, b)$.

Polsby-Popper (PP) Let C_i be the circle where $L(C_i) = L(P_i)$. The Polsby-Popper score is equal to $\frac{A(P_i)}{A(C_i)}$. This value ranges from the least compact 0 (a district with no area), to the most compact 1 (a circle-shaped district). For reporting we scale the value to lie on the range $[0, 100]$. A plan's PP score is simply the mean of each district PP score.

Convex Hull (CH) Let CH_i be the convex shape which bounds P_i and has the minimal value for $A(CH_i)$. The Convex Hull score is equal to $\frac{A(P_i)}{A(CH_i)}$. This value ranges from the least compact 0 (a district with no area), to the most compact 1 (a convex district). For reporting we scale the value to lie on the range $[0, 100]$. A plan's CH score is simply the mean of each district CH score.

538 metric The 538 metric is not formally explained, but it is described as "the average distance between each constituent and his or her district's geographic centre" [44]. One possible interpretation of this could be $\sum_{v \in V(G_i)} n_v d(v, M(P_i))$. But it is also possible there are other interpretations of what centre means.

DRA metric The DRA metric is described as a blend of compactness scores normalized by historical data, and optimal values. Because of the ambiguity of its definition, we don't actually calculate the score ourselves. Instead, we upload our plans (when possible) to the DRA website and have them calculate it for us[10].

Density Deviation Finally, we introduce a novel metric based on the variance of the population density of the precincts within a district. For a district with subgraph G_i, for a node $v \in V(G_i)$ with population n_v, let its polygon

[9] This is the point where P_i would balance on a pin tip if it was a two dimensional object.

[10] See `https://medium.com/dra-2020/compactness-8e0ee3851126` for more information regarding their metric.

be P^v and the area of the polygon be given by $A(P^v)$. We define the density of v as $\gamma_v = \frac{n_v}{A(P^v)}$. We define the Density Deviation of a district i as $\frac{StandardDeviation_{v \in V(G_i)}(\gamma_v)}{mean_{v \in V(G_i)}(\gamma_v)}$. This value, standard deviation divided by mean, is known as the coefficient of variation. A plan's Density Deviation is the mean district Density Deviation.

While the first four measures may not be perfectly correlated with each other, what is considered compact by one is often considered compact by another, though there can be exceptions to this. Some states have a geography which may be incompatible with "nicely" shaped districts according to certain definitions. For example, Figure 2 shows how WI has a large jagged bay and lakes slicing and poking holes in the state geography. Because a large population centre, Green Bay, is located at the base of this bay, drawing population balanced districts which are close in shape to a circle (optimized for the Polsby-Popper score) may be difficult.

The reader may ask why we define a district's Density Deviation as the coefficient of variation instead of the standard deviation. The reason is because when finding a plan which optimizes for this value we aim to minimize the mean Density Deviation across all districts in the plan. That is, we want to find plans where each node in a district is of a similar density. By dividing the standard deviation of the densities by the mean density we normalize the scores, and make the various districts comparable on the same scale. If we just compared standard deviations of densities then the scores of districts with a large mean node density would dominate the metric.

Unlike Polsby-Popper, Convex Hull, and DRA's measure, which are pure district shape based metric, and 538's metric which is a a measure of population travel distance (hence also district shapes), our novel deviation metric does not factor in the shape of a district. We find this interesting for a few reasons. First, as mentioned in WI the geography may make designing geographically compact districts difficult, our measure presents a novel alternative for defining compactness. Secondly, a common goal in redistricting is keeping communities of interest together. How one defines a community of interest is up for debate, but it certainly is not a stretch to say similar densities of neighbourhoods could be a factor. It also allows us to examine if keeping urban areas separate from rural ones (as optimizing for this metric will tend to do) helps in ameliorating rural gerrymandering advantage.

4. Election Settings

In this paper, we use election data from three US states — Pennsylvania, North Carolina, and Wisconsin — from the 2012 and 2016 presidential elections.[11] These are states and elections for which granular, precinct-level, data

[11] In Sections 5.2 and 5.3 we also use data from Maryland and Massachusetts for a proof of concept. But, because of missing votes and an overwhelming Democratic lean respectively,

was available. The actual aggregation of data was done by the Metric Geometry and Gerrymandering Group (MGGG), a multidisciplinary research group that focuses on redistricting at Tufts university[12]. Each of these three states has a particular election of interest. For reference of scale, we also include the number of nodes (precincts) and edges in the graphs of each state.

Pennsylvania (PA) 2012 *Sizeable Democratic advantage.* The Democratic candidate (Obama) won 51.97% of the vote, vs. 46.59% to the Republican candidate (Romney). PA has 9,255 nodes and 25,721 edges.

North Carolina (NC) 2016 *Sizeable Republican advantage.* The Republican candidate (Trump) won 49.83% of the vote, vs. 46.17% to the Democratic candidate (Clinton). NC has 2,692 nodes and 7,593 edges.

Wisconsin (WI) 2016 *Near tie.* The Republican candidate (Trump) won 47.22% of the vote, vs. 46.45% to the Democratic candidate (Clinton). WI has 6,634 nodes and 18,126 edges.

In addition to these elections, they provide a good mix of geographic features. WI, for example, has its north-east corner carved up by lake Michigan, forming a jagged bay. PA and NC, on the other hand, have a much more convex shape[13]. Furthermore, the population distribution is varied: PA's large urban centres are in its east and west edges, whereas in NC, the urban centres are concentrated in the middle of the state.

4.1. Quality of Voting Data

Because each state has different standards for reporting voting data, and the data was aggregated from various sources by MGGG, there is a small amount of missing vote in each election. In each state and each election we look at, for each party there is over 99.4% of the total vote accounted for. Generally this vote aggregation is without issue, but there is one case we should discuss further. In PA, because the voting data and census data do not exactly overlap in their geographic divisions, there are some cases of nodes with an adult population but no vote. These vote-free nodes are because a few nodes in the voter data may overlap with several nodes in the census data. In these ambiguous cases, MGGG opted to assign voters to the census node which shared the most geographic overlap with the voter node. In total, just under 2.5% of the populated nodes in our data have no votes in at least one of the two presidential elections we look at. While there are other reasonable methods for dealing with this ambiguity we opt to use the MGGG data as is. We felt making further assumptions without more information was not justified. Furthermore, the issue only impacts a very small fraction of the population and nodes.

---

we omit them from subsequent experiments.

[12] Data from MGGG (`https://github.com/mggg-states`).

[13] The state government shape files for the precincts in North Carolina consider the water features as contributing to that node's geography. Thus the overall shape of the state is fairly convex.

4.2. Voting and the Urban-Rural Divide

State	2012 correlation	2016 correlation
NC	0.79	0.80
PA	0.47	0.58
WI	0.38	0.57

Table 1: Spearman correlation between a precinct's fraction of D party votes and its density (total population divided by area) in three states and two elections.

As noted above (and explored visually in Figure 1), a key geographic feature of US political parties is the growing divide between a more rural Republican party and a more urban Democratic party. As we have cited above, there are numerous works that examine the geographic history and begin to explore the consequences of this divide.

Unsurprisingly, the urban-rural divide is present in the data we work with. We focus on three different states, with differing ethnic makeup, education patterns, and history, but this feature was common across all our data: densely populated urban centres favour the Democrats, while sparsely populated rural regions favour the Republicans. This correlation is consistent and undeniable, as Table 1 shows. In every state this correlation increased from 2012 to 2016, showing the effect of the urban-rural divide has only strengthened over the last decade.

5. The GREAT Algorithm

To study the role of compactness and population distribution in gerrymandering we need an algorithm that can optimize for various compactness and partisan fairness metrics (or handle them as constraints) on real-world data. To that end, we introduce our Goal-based Redistricting for Elections Automatically using Technology (GREAT) algorithm, that can create districts from graph representations. As we will demonstrate, our algorithm, with minimal engineering effort, can be used to optimize various measures of partisan fairness (e.g., to minimize either of our robust partisan bias metrics), partisan gain (e.g., the number of districts won by a given party either by achieving a plurality of votes, or at least a threshold fraction of votes, or with at least a certain probability), and compactness (both geographic and density based). Furthermore, the algorithm can optimize towards one of these goals while satisfying strict constraints on other metrics (e.g., optimize compactness while ensuring that a given party wins at least a fixed number of districts).

To show its capabilities, we will demonstrate our algorithm is capable of matching the performance of human experts when creating partisan plans (Section 5.2), and compact plans (Section 5.3). In Section 5.4 we show our algorithm is able to compete with human experts in a prestigious redistricting competition.

First, we give a brief overview of our algorithm. Our method is based on simulated annealing, a common local-search-like method which can make non-improvement steps, allowing it to escape local optima.

At the core of any SA algorithm, or any local search algorithm, is the energy function which scores each potential plan the algorithm proposes. By associating every potential plan with a numeric score, where lower is better, a local search algorithm can make small modifications to the current solution and propose a neighbouring solution. If this neighbour has a lower energy, it is better for our objective, so we make it our current solution. If the neighbour has higher energy, we may make it the current solution (with probability that decreases as the process continues and decreases if the difference in quality is large). Any objective that can be expressed numerically and calculated from an arbitrary plan may be used; that is, we only require that the energy function which maps plans to numbers is "efficiently" computable. This gives us considerable flexibility in optimizing, allowing for complex and highly non-linear objectives. After some fixed number of iterations or elapsed time, the process ends and the best of all iterated solutions is returned.

Additionally, any binary constraint (for which it can be checked whether a given plan satisfies) can be incorporated by ensuring that the algorithm only considers steps which satisfy the given constraint. This ensures our iterated, and ultimate, solutions can guarantee things like a certain number of wins for a target party, or that the solution satisfies compactness requirements.

For any given step, to generate a neighbouring plan we use a modification of the tree-recombination procedure proposed by the Metric Geometry and Gerrymandering Group [45]. Briefly, the method takes a set of m adjacent districts from the current solution, and recombines and redivides the nodes in them to form m new districts. This is done by drawing random spanning trees over the precincts of the m districts and cutting random edges in the trees to separate the nodes into the desired number of districts. For efficiency reasons, we generally use $m = 2$. Using larger m values did not noticeably improve the results.

The final piece of our method is how do we provide the initial plan to start the process. Almost always, we can use existing plans for our starting point. But for a diversity of starting points, in general, we generate a random plan by using the tree-recombination algorithm on the entire state. That is we partition every single node of the graph by iteratively drawing $k - 1$ spanning trees, each of which carves off a new district. We find that if our only constraints are population and connectedness, this method is more than able to partition the state into dozens of legal districts. If we have more stringent constraints, such as target party wins, or fairness requirements, we first use our algorithm to optimize for the constraints, then use this solution as the starting point for another run of our algorithm.

Section 5.1 below contains further details regarding the algorithm, including pseudo code and descriptions of each step of the process. For readers familiar with simulated annealing and the MGGG recombination algorithm, this section may be skipped.

Like the work before us, we are unable to provide guarantees (with respect to optimal solutions) on our method's performance. Instead, we compare against the best plans human experts have created. As far as we are aware, we are the first to publish work that compares against, let alone matches, state of the art hand-drawn plans.

5.1. *A Detailed Description of the* GREAT *Algorithm*

The GREAT algorithm is built around simulated annealing (SA), a general optimization technique that has found much success in various discrete optimization problems. For those familiar with simulated annealing, one can ignore the general overview of the algorithm below, and examine only the details on the specifics of energy and temperature functions. In a sense, SA extends hill-climbing, which is the discrete version of gradient descent. In hill-climbing we have a current state and examine some neighbouring solution. We need to decide if we should move to this neighbouring solution or not. In standard hill-climbing methods, like (greedy) local search, only moves which improve the solution are accepted. At a high level, simulated annealing-based optimization is essentially a hill-climb in which sometimes moves are allowed towards inferior solutions. The ability to accept non-improvement steps, i.e., objectively worse solutions, becomes increasingly less permissible as the optimization proceeds. The logic of allowing such non-improvement steps is that they allow the procedure to escape local optima earlier in the process. If the space of solutions is non-convex (with respect to solution quality) these local optima can act as sinks for procedures which only allow improvement steps. The ability to accept a non-improvement neighbour is controlled by two parameters, the "temperature" of the system (a parameter set by us, allowing us to determine a baseline willingness to accept such moves) and the difference in quality (also known as energy-difference) of the current and proposed solution.

Energy:. The first component of an SA based approach is the *energy* of a solution. The energy of a solution is a function which maps a potential solution to a numeric measure of quality, for our work we consider the set of solutions to be all legal districting plans. That is, if a graph of a state has a node set with n nodes and they must be partitioned into K districts. An energy function is a map:

$$E : [n]^K \to \mathbb{R}_+ \cup \infty. \tag{3}$$

It is standard for lower energy values to correspond to superior solutions and for zero energy to be the best any solution can take on[14].

Proposal:. The second component of the SA based approach is the *proposal* function. A proposal function P takes in a potential solution S and picks a *neighbour* S' of S:

$$P : [n]^K \to [n]^K. \tag{4}$$

[14]The optimal solution for a particular instance could have non-zero energy, zero just serves as a lower bound. Generally invalid solutions have infinite energy.

There is no fixed definition of what a neighbour is and this can vary from domain to domain, or even within a problem itself. For our work we will use the following recom-proposal function, which was first suggested by MGGG. The recom-proposal is presented in the following algorithm 1:

Algorithm 1 recom-proposal(G, S, j):

Let S be the current solution (districting plan which partitions the vertices into k components).

1: Pick at random $i \in \{2, \cdots, j\}$ connected districts from S (where $j \leq k$). ▷ Let R denote the precinct nodes in these i districts.
2: **for** $t \in \{1, \cdots, i-1\}$ **do**
3:     Draw a random spanning tree using only the nodes of R. Call this spanning tree T_t.
4:     Sample a random edge e that has yet to be picked (see details in text) in T_t. This divides T_t into 2 connected components.
5:     **if** T_t beneath e forms a valid district **then** Make it one of our new districts, remove these nodes from R.
6:     **else if** there are edges yet to be sampled and T_t beneath e is not a valid district **then** Repeat step 4.
7:     **else if** all of the edges of T_t have been sampled and no valid district was ever found in T_t **then** repeat step 3.
8:     **end if**
9: **end for**
10: Let the remaining nodes of R be the final new district.
11: Let S' be the solution identical to S but where R has been redistricted according to steps 2-10.
12: **if** S' is a valid solution **then**
13:     Return S'
14: **else**
15:     Retry the algorithm from step 1.
16: **end if**

When we say a solution or district is valid we mean that it satisfies all constraints we place on districts. We use Kruskal's algorithm for drawing spanning trees (the drawing of T_t) by randomly assigning each edge a weight and finding a minimal spanning tree. Thus, it is done in time linear to the number of nodes left in R.

Each time through the for loop at step 2 the algorithm may pick several edges in the spanning tree T_t (step 4). For each iteration of the loop, we shall require the first such e to be connected to a leaf of T_t (the first time step 4 is executed for each loop iteration). If the node under e is not a valid district then another random edge is chosen (step 6). We shall require this next edge to be either another edge connected to a leaf or the edge directly above e in T_t. In subsequent steps (if they are required) the algorithm shall pick an edge e that has not been previously selected in T_t with two conditions on this edge

e: First, this edge is either connected to a leaf node or is the direct ancestor of a previously chosen edge. Second, the algorithm requires that all of the edges under e in T_t have previously been selected. Intuitively, this process works by bubbling up through the various branches for T_t, trying edges until a sufficient one is found.

Thus, each time through the for loop the algorithm should find one of the districts we need. It is possible some iteration of the for loop will fail to find a valid division, sampling every edge in T_t (step 7). In this case this iteration of the for loop restarts, finding a new spanning tree, but all districts which were found up to this point are still kept.

It is also possible that for some iteration of the for loop no spanning tree can lead to a valid districting. That is, it will just draw new spanning trees forever. We are unaware of any method that can detect this scenario, short of sampling every spanning tree. As a heuristic solution, we put a time limit on the algorithm. We found the algorithm tends to find solutions within 20 seconds for the most complex instances we work with. If after 1000 seconds we do not have a solution we restart the entire algorithm.

We note that the recombination method proposed by MGGG only worked for recombining two districts at a time, whereas we extended it to work for any number. In the step where we pick i random districts for recombination we do so by sampling uniformly at random from the set of all sets of connected districts up to size j. The intention of the MGGG method seems to be the same (for $j = 2$), but their code shows that they pick districts by uniformly sampling from all edges which cross district boundaries. This will favour picking pairs of districts which share large boundaries (in terms of nodes). In general we found that increasing the number of merged districts beyond two did not improve our solution quality (but it did slow the procedure down).

Temperature:. The third part of the SA approach is the *temperature*, which acts as a control for how likely negative moves are at a given state of time. Generally the temperature is a decreasing function of the number of iterations so far in the optimization. While there are many temperature functions and choosing the ideal one is somewhat of a black-box in optimization, we've found the following temperature function (where s denotes the iteration count) works well:

$$T(s) = 10000 \cdot (0.99)^s \tag{5}$$

This is known as the exponential cooling schedule. From the initial temperature of $10,000$ at every step we retain 99-percent of the remaining heat until we eventually cool to a temperature of 0.

5.1.1. The simulated annealing method

The simulated annealing method is as follows for a graph $G = (V, E)$ which is to be partitioned into K districts:

In the first step $None$ refers to the districting which makes no assignments. To find the initial partition we do not need to provide the sub-routine with a

Algorithm 2 simulated annealing for gerrymandering(G):

Let $S_0 = recom\_proposal(G, None, K)$.
$i = 0$
while $i \leq s_{max}$ **do**
  $S' = recom\_proposal(G, S_i, j)$
  if $E(S_i) \geq E(S')$ **then**
    $S_{i+1} = S'$
    $i = i + 1$
  else
    Let $\Delta E = E(S_i) - E(S')$
    Let r be drawn uniformly at random from $[0, 1]$.
    if $\exp \frac{\Delta E}{T(i)} \geq r$ **then**
      $S_{i+1} = S'$
      $i = i + 1$
    end if
  end if
end while

valid districting since we are recombining all of the nodes. In the later iterations we can set j (the number of created districts) to whatever value, but as we noted above, $j = 2$ works well. Intuitively, the algorithm will always move to a lower energy solution and will move to a higher energy solution with high probability if the increase in energy is not too high and the temperature is not too low.

It is possible that the procedure will eventually end up in a local optimum (or even a global optimum) it cannot move away from with reasonable probability. This is especially true later on as the temperature drops. If this is the case the main loop will, with very high probability, make no progress to completion. Because of this we often set a hard time limit and cut off the procedure after this point. In general with SA, or any random algorithm, one needs to run many parallel executions of the procedure, and each of these will iterate over many potential solutions. The best of all iterated solutions will be chosen as the returned solution.

For the remainder of this paper, when we introduce a task for our algorithm we will describe how we optimize for it by specifying, the number of cores we ran our algorithm across, how long we ran the algorithm for, the exact setting of the energy function, and any constraints. This highly modular setup allows us to efficiently optimize for almost any objective. Furthermore, unless otherwise stated, we always start our algorithm off by building a random initial solution. The only exceptions to this will be when we use our algorithm to optimize for multiple objectives. For example, to gerrymander for partisan gain while maintaining compact districts, we use a two step approach: First, we use our algorithm to create a partisan plan, winning as many as many districts as we can for our target party. Then, using our partisan plan as the starting point for another run of our algorithm, we optimize for some compactness scores, but with the constraint that we do not consider plans with too few wins for our target party.

5.2. Proof of Concept: 538 Gerrymandering

Nate Silver and the election experts at 538's gerrymandering project [46] drew thousands of hand-crafted districts for various objectives. While there is no guarantee their plans are optimal, they do serve as an excellent, and publicly available, benchmark.

As noted above, winning a plurality of votes is just one of the measures of what it means to win a district. At 538, they took a probabilistic view, designing partisan plans that maximized the number of districts that were won with a sufficient probability. This expanded measure of victory also serves as an ideal goal to show the modularity of our algorithm. Unfortunately, they released few details regarding their method. However, we believe we were able to reconstruct it using released results.

Briefly, 538 uses the Cook Partisan Voting Index (CPVI) [47], which measures a district's D party bias according to the 2012 and 2016 elections and then transforms this CPVI into a probability that the Democrats win the district. First, we show how to determine the general formula for the CPVI of a district. After, we will show how 538 transforms the CPVI into the probability of a Democratic win.

The CPVI:. The CPVI is a metric which measures how partisan a group of voters, in particular those who form a congressional district, are relative to the average voter in the United States. To calculate the CPVI there needs to be a running value for how partisan the country is as a whole (call this value β_D). To calculate this we take the votes in the two most recent presidential elections[15] and see what fraction of these votes belong to the Democratic party. The partisan skew expresses the average of the vote fractions for the Democrats in the last two presidential elections as an average of averages (not weighted by the total votes in each election). To calculate the PVI 538 used we need the 2012 election, in which:

- Barack Obama and Joe Biden of the Democratic party: $65,915,795$ votes.
- Mitt Romney and Paul Ryan of the Republican party: $60,933,504$ votes.

For the 2016 election the exact results were:

- Hillary Clinton and Tim Kaine of the Democratic party: $65,853,514$ votes.
- Donald Trump and Mike Pence of the Republican party: $62,984,828$ votes.

[15] The presidential election is chosen since they use the same candidate for the entire country and thus are free of any local effects.

Using the above information we get the value of β_D would be:

$$\frac{\frac{65{,}915{,}795}{65{,}915{,}795+60{,}933{,}504} + \frac{65{,}853{,}514}{65{,}853{,}514+62{,}984{,}828}}{2} \tag{6}$$

Thus, we see β_D is roughly 51.53%. While the US is effectively a two party system, there are other candidates – Gary Johnson and Joe Weld of the Libertarian party received $4,489,341$ votes (over 3% of the total vote) in the 2016 elections. Since the Cook PVI is meant to be a direct comparison between the Democratic and Republican party it does not factor in third-party votes. The PVI of a district is how partisan that district is relative to β_D. In district i, let the total number of Democratic votes denoted by $N_i^{D,1}$ and the Republican ones as $N_i^{R,1}$ for the last presidential election; and $N_i^{D,2}$ and $N_i^{R,2}$ for the presidential election before that, then the PVI is:

$$100 \cdot \left(\frac{\frac{N_i^{D,1}}{N_i^{D,1}+N_i^{R,1}} + \frac{N_i^{D,2}}{N_i^{D,2}+N_i^{R,2}}}{2} - \beta_D \right) \tag{7}$$

Equation 7 can range from $-100\beta_D$ for completely Republican dominated districts, to $100(1-\beta_D)$ for districts with only Democratic voters, or 0 for districts which match the national average in the last two presidential elections. Intuitively, a district with a very positive PVI should be safely Democratic. Even if there is a uniform swing towards Republican sentiments this particular district should lean Democratic (the same is true for Republicans and districts with a very negative PVI).[16]

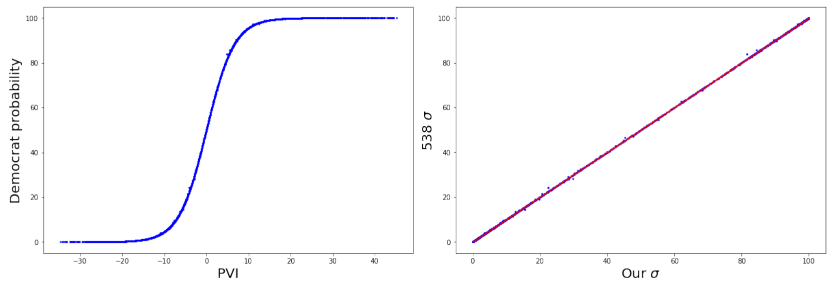
The 538 probability:. Next, 538 transforms the CPVI into the probability that the D party wins that district. The R party wins it with the remaining probability. To find this probability we believe 538 used a sigmoid function (the inverse of the log-odds function). We now describe how we reconstructed the sigmoid function. Recall, the sigmoid function takes the form:

$$\sigma(x) = \frac{1}{1+e^{-w\cdot x}} \tag{8}$$

This function is fitted to data (x) by adjusting the weight parameter w. Unfortunately 538 was not specific on what exact data was used to fit the sigmoid, or if regularization terms were included in the fitting. Luckily, 538 did publicly report the Cook PVI and their derived probability of a Democratic win for all the districts in their catalogue for each state.[17] The probability of

[16] While The Cook Political Report does not actually publish the formula or exact method for this metric, we confirmed our interpretation by measuring the reported PVI in single district states and comparing to the formula we derived.

[17] In total there are 2568 districts. These districts are the entirety of all of their created plans. This includes plans such as the partisan plans, competitive plans and plans that emphasize compactness. They also include the current congressional plans.



(a) 538 model

(b) Our model vs 538's

Figure 3: First figure shows the reported Cook PVI for each district created by 538 vs their estimation of the probability that the Democrats will win that district. Second figure shows the output of our reconstruction of the 538 model vs the 538 model itself, the inputs to these two models were each of the districts created by 538.

a Democratic win, plotted against the Cook PVI (Figure 3a), clearly shows a sigmoid shape. From here we just need to derive what the weight parameter w they use is. To figure this out we first invert the sigmoid function using the log-odds (or logit) function:

$$logit(\sigma(x)) = \log_e \frac{\sigma(x)}{1 - \sigma(x)} \tag{9}$$

Inverting the sigmoid function with the logit function would produce a line given by $y = w \cdot x$, thus we simply need to invert any two data point in the second subfigure of Figure 3 and take the slope of the resulting line as w (since this is a linear function of one variable any two distinct points are sufficient to determine it). Briefly, we mention two important points. First, the sigmoid, and hence the line from the logit, may have a bias term associated with them. We found 538 did not include one since their sigmoid passes through $(0, 50)$[18] and the resulting logit line passes through $(0, 0)$. Secondly, the points 538 published do not perfectly follow a sigmoid, instead there is a small amount of "jitter" on some of the points in the first subfigure of Figure 3. This deviation could simply be a rounding issue or minor transcription errors, in either case the points still very closely follow the sigmoid pattern. Because of the small amount of noise the resulting inverted plot found with the logit function will not be perfectly linear. Thus our choice of the two points for the inference of w would (very slightly) influence the outcome. To mitigate this issue we take the ordinary least squares (OLS) regression line, also known as the line of best fit, for all of the points (after inverting them with the logit). We found the slope of the OLS line was 0.304 which we ended up using for the w parameter in our sigmoid model. Our resulting model is a near perfect fit for the 538 model since they form the line $y = x$ when plotted against each other (Figure 3b).

[18] There is exactly one data point with a PVI of 0 and a Democratic probability of winning of 50%.

Optimizing to match 538. When gerrymandering for party P, 538's objective was to maximize the number of districts for which P's probability of winning was at least 82%. To guide our method, we used a combination of the expected number of districts won by P and the total number of districts won with at least 82% probability. Thus, our energy function is based on the number of districts won at 82% with one slight modification. Say we are gerrymandering for the Democratic party. If a potential solution S is comprised of K districts called $S_1, \cdots, S_K$, and we want to have many districts which we win with probability τ or more (for all of our simulations we follow 538 and use $\tau = 0.82$), then the energy of that solution is:

$$E(S) = K - \sum_i v_D(S_i), \tag{10}$$

where $v_D(S_i)$ is equal to :

$$v_D(S_i) = \begin{cases} \sigma(S_i) & \sigma(S_i) \leq \tau \\ 1 & otherwise \end{cases}$$

Where σ is the sigmoid function we derived from the 538 data. If the target party is Republican party we can replace $v_D(S_i)$ with $v_R(S_i)$ [19] which is defined as follows:

$$v_R(S_i) = \begin{cases} 1 - \sigma(S_i) & 1 - \sigma(S_i) \leq \tau \\ 1 & otherwise \end{cases}$$

Intuitively, our function is aiming to maximize the number of safe wins for the target party. Our method would prefer a solution with several borderline safe wins over a solution with fewer very safe wins, i.e., winning with just over the 82% threshold and the extra votes in the districts under this threshold are preferable to plans which use the extra votes to push the same number of wins well over the 82% threshold. The idea here is that these loosing districts under the threshold can become more competitive and eventually wins if we move the extra votes into them.

Our only constraints were that districts must be connected and within half a percent of the ideal population. For each state and party we found well before our cutoff of 24 hours the algorithm had stopped making steps. We also limited ourselves to 60 parallel runs for each state and party combination.

Almost as good as 538. The availability of presidential election data at the precinct level is inconsistent, so we are unable to compare against 538 in all states. There are five states for which we have publicly available data, and for each of them we optimized for the 538 objective for each party. Our results are shown in Table 2.

[19]Recall, the probability the Republican party wins a district is just one minus the probability the Democratic party wins it.

State	Total seats	Our D	538 D	Our R	538 R
MD	8	7	5 (8)	4	4 (4)
MA	9	9	9 (9)	0	0 (0)
NC	13	7	8 (8)	11	10 (10)
PA	18	8	8 (9)	13	13 (13)
WI	8	5	5 (5)	6	6 (6)

Table 2: First column is the number of seats in the state. Second and third columns are the number of districts D take with over 82% probability with our algorithm and the 538 optimally-gerrymandered plans, respectively. Fourth and fifth columns are the same for the R party. The 538 numbers show the number of districts won according to their districting based on our election data. In parentheses are 538's results from their website using absentee data (which we did not have access to).

To actually make comparisons we needed to transfer the 538 plans into our data format. Briefly, we needed to map the polygons describing 538's districts onto the precinct level data we had. To do so, we used a tool from MGGG called MAUP. This tool assigns each precinct level polygon to exactly one of the district level polygons. In most situations, the assignment is unambiguous, a precinct polygon is entirely contained in a district polygon. In a few cases the data did not line up, and a precinct could belong to several districts. In this situation we use the default MAUP behaviour, assign the precinct to the district with which it has the most geographic overlap. Because of these ambiguous situations, there were some non-contiguous assignments created. For us, this was not an issue. We only needed to count the number of districts the 538's plans won using our data.

Overall, there was only one case, NC for D, where we did not match 538. Even here, we only missed by one district out of the 13. We did outperform 538 in Maryland for the Ds, but we caution we were missing 25% of their vote for each party – the absentee data (mail-in ballots), for which we have no precinct level data.[20] In NC for the R party, we also outperformed 538, although we caution, small differences in the voting data may account for this.

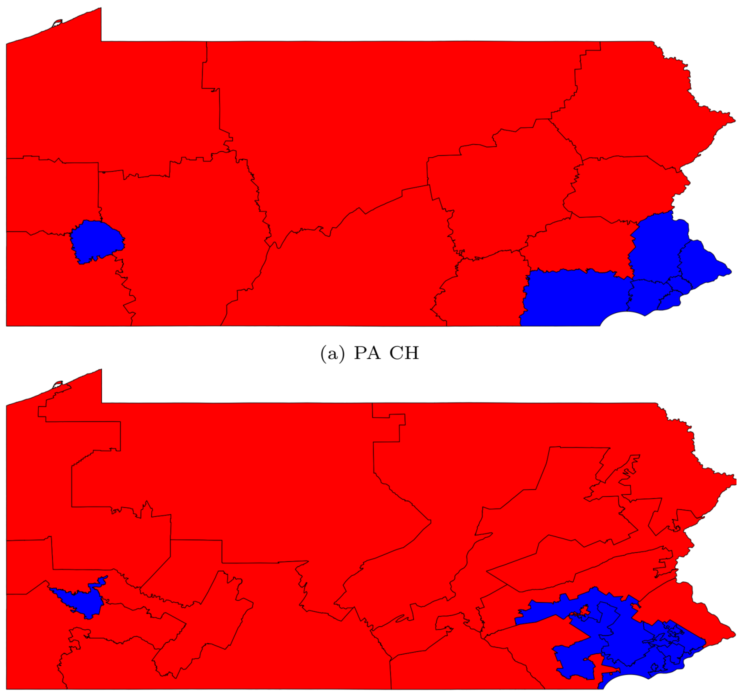
5.3. Proof of Concept: Compact Redistricting

As mentioned, compactness is often a legislated requirement, even if the mathematical definition and the required levels are not specified. Despite this ambiguity our algorithm is able to easily optimize for a variety of compactness scores.

We use our algorithm to find two plans, one optimized for the mean score across all districts of the Polsby-Popper score, and one for the mean score across all districts for the of the Convex Hull score.

Optimizing for Compact Districts. To use Algorithm 2 to create the Polsby-Popper and Convex Hull compact plans we had our energy function be the

[20] In the other 4 states there are at most 0.3% missing ballots.



(a) PA CH

(b) PA 2011

Figure 4: PA districts (R wins in red); (D wins in blue) based on the 2016 PA election data. Top, our plan, optimizing the convex hull score; bottom, PA's actual 2011 districts.

	MD			MA			NC				PA				WI		
	PP	CH	538	PP	CH	538	PP	CH	538	DRA	PP	CH	538	DRA	PP	CH	538
Our PP	**27**	75	28479	**30**	76	0.21	**47**	87	39924	**97**	**44**	86	0.31	**93**	**40**	85	55991
Our CH	18	**76**	29090	25	**79**	0.21	38	**89**	42319	86	37	**88**	0.29	85	35	**87**	48729
Our Fair		N/A			N/A		37	83	45176	78	26	76	0.34	49	29	77	56958
538	16	72	**26165**	20	77	**0.19**	36	86	**37488**	83	34	87	**0.27**	81	25	83	**47728**
DRA	14	72	27330	19	71	0.21	40	84	44354	78	40	82	0.29	70	37	81	5025
2011 Plan	4	47	37459	13	59	0.24	11	60	55228	10	16	62	0.41	15	21	74	52623
Updated		N/A			N/A		25	71	53160	36	32	78	0.30	64		N/A	

Table 3: Table containing the compactness scores for various plans in each state according to different metrics. Each row is a plan, each group of columns is a state, each column is a compactness metrics. Within each state, for each compactness metric, the plan with the best score is bolded. The PP, CH, and DRA, scores are on a scale of 0 to 100, where 100 is the most compact. The 538 metric is on a scale of 0 to ∞, where 0 is the most compact; due to different state data formats, each state is scaled differently. For Maryland and Massachusetts we were missing some voter data, so we did not create a Compact and Fair plan. Because of differences in our data and the data on the DRA website we are only able to calculate the DRA score for Pennsylvania and North Carolina. Only the Pennsylvania and North Carolina 2011 plans were overturned, so there is no Updated Plan in the other states.

final objective: the mean compactness score of our plan (where the mean is taken over the individual score of each districts). Say we are optimizing for compactness measure M. If a potential solution S is comprised of K districts called $S_1, \cdots, S_K$, and $C(S_i)$ is the compactness score of district S_i the energy of that solution is:

$$E(S) = \frac{\sum_i C(S_i)}{K} \tag{11}$$

The only constraints on a proposed districting were a population balance of at most half a percent from ideal, and connectedness. For each metric for which we optimized, we ran our algorithm for 24 hours across 288 cores.

Compact Districts. From a visual standpoint, our plans (see Figure 4a) pass an "eye test" for looking compact, especially compared to the plans enacted in real life (see Figure 4b).

As was the case for gerrymandering, 538 implemented a compact plan for each state. These plans were designed to minimize "the average distance between each constituent and his or her district's geographic centre". In addition, we have plans created by the public using Dave's Redistricting App (DRA). DRA is the most popular, open source tool for redistricting, and 538 also used the DRA to help create all their plans. Amongst all plans ever published on DRA, the website features the most compact (according to their internal metric) for each state[21]. In addition, we have the 2011 plan for all relevant states, and for NC and PA a court mandated updated plan as well.

Compared to *all* of the above mentioned plans, in *every* state our plans had the best mean compactness scores for their respective metrics. They are sometimes compact even according to metrics for which they were not optimized;

[21] For all plans see `https://davesredistricting.org/`. See `https://medium.com/dra-2020/compactness-8e0ee3851126` for a high level description of DRA's compactness metric.

in PA and NC, our PP plans have the best DRA score). We are not claiming the compactness measures we chose are superior to others. As Table 3 shows, in each of the five states we examined, the four compact plans (PP, CH, 538, DRA) have similar scores in each metric (PP, CH, 538). We only argue that for a variety of measures our algorithm is capable of creating plans just as – or more – compact as those from human experts.

5.4. Proof of Concept: Princeton Redistricting

Finally, we used our algorithm for a redistricting competition, *The Great American Mapoff*, hosted by Princeton University's Gerrymandering Project. This competition was meant to raise awareness around the 2021 redistricting cycle in the United States and to help recruit members to their Mapping Corps (a volunteer group to help study and design potential redistricting plans). The competition involved designing plans for several states and goals. We saw this as an excellent proof of concept for our algorithm, and ultimately a chance for it to improve social good.

We used our algorithm to enter the *Stealth Gerrymandering* for Illinois and *Partisan Fairness* for Wisconsin categories. Here, we were finalists from among almost 150 entrants[22]. Our plans, created in days, were judged by human experts to be among the best, as good as the handcrafted plans submitted by other participants. Because the contest goals were open ended, we are unable to make a quantitative comparison, instead we can qualitatively describe our plans and what we did. We were also invited to join the Mapping Corps.

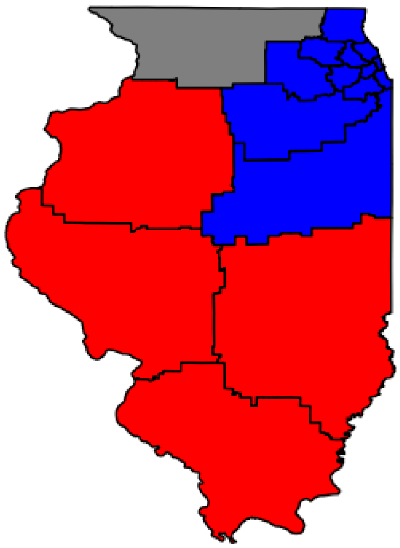
The data used elsewhere in this paper is not the same as the data the competition used. The competition used the updated 2020 precinct shape files and interim 2019 census data. This data is still being updated to the 2020 census data, thus it is currently in flux. Furthermore we ultimately wish to make comparisons against the large amount of plans, including the actual implemented ones, published using the 2010 census and shape files.

It is also worth noting DRA, and thus the competition, didn't use raw vote totals in their evaluation. They used a composite score which we will describe in detail briefly. As far as we can tell there is no presented evidence that this composite score is a better predictor of future elections than simpler metrics.

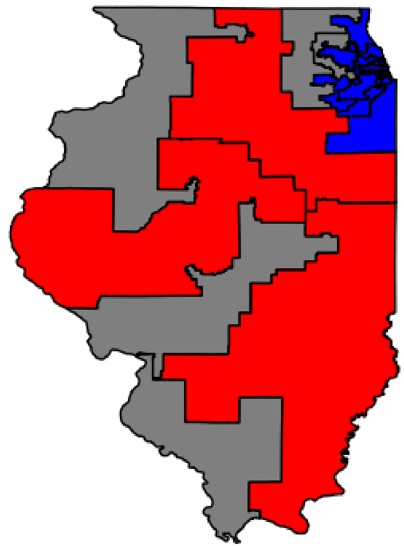
5.4.1. Illinois Stealth Gerrymander

This plan is designed to gerrymander for the Democrats while maximizing the Polsby-Popper compactness score. It does so with a population deviation of only 0.75%. We used the DRA's definition of winning a district, which is that at least 55% of the composite vote is needed to win a district. Our map secured 12 districts for the Democrats, with one additional district being a tossup. Our gerrymandered map compares favourably to the existing Illinois map that has one more district (18), only 10 of which are Democratic at the 55% threshold, and is considered to be highly gerrymandered for Democrats. Our algorithm

[22]See `https://gerrymander.princeton.edu/map-contest` for details.



(a) IL our stealth gerrymander



(b) IL existing plan

Figure 5: Congressional maps. Blue districts have at least 55% Democratic vote, red no more than 45% Democratic vote, grey otherwise. All values calculated using the default composite vote total specified by the Princeton Gerrymandering Project.

achieves this by avoiding several blow-out wins for the Democrats, using these votes to convert tossups and Republican wins to Democratic wins.

Our gerrymander (Figure 5a) is far harder to detect than the existing one (Figure 5b). It is far more subtle than the existing one and easily passes the "eye test". Our least compact district (Poslby-Popper 21%) is more compact than all but 4 of the existing districts. Comparing according to other compactness measures lead to similar results.

How we did stealth gerrymandering in Illinois. First, for our target party, Democrats, we generated highly partisan districts. That is, given a partitioning of the nodes of G into $S = (S_1, \cdots, S_K)$ set Equation 3 (the energy function) as follows:

$$E(S) = K - \sum_i v_D(S_i), \tag{12}$$

where $v_D(S_i)$ is equal to :

$$v_D(S_i) = \begin{cases} \frac{N_i^D}{N_i^D + N_i^R} & \frac{N_i^D}{N_i^D + N_i^R} \leq \tau \\ 1 & otherwise \end{cases}$$

Here N_i^D is the total Democratic vote in district i (N_i^R is the total Republican vote in district i). We set $\tau = 0.55$, as required. This is similar to our method for emulating 538, but now the sigmoid function's contribution to the energy has been replaced by a linear distance to winning the district. We tried other definitions of $v_D(S_i)$ and $v_R(S_i)$, such as exponentially decreasing energy as one gets closer to winning the district, a small decreasing contribution to energy even if the target party is winning a district, and modifications of the sigmoid. In the end we found the presented definitions worked best.

To generate the gerrymandered, but not compact, plans we used our algorithm for 24 hours. This time limit was more than sufficient for the convergence of the various processes. In the end, the plan with the most Democratic wins had 13 districts at 55% or better for them. We took the plan with the most wins as the starting point for new runs of our algorithm and optimized for the Polsby-Popper score. This required us to set the energy function to be the mean Polsby-Popper score across all districts. We added the additional constraint that for a proposal to be considered in algorithm 1 we required that the number of districts won by D at 55% is at least 13. We repeated this exact process twice, but with the win constraint lowered to 12 and 11 Democratic wins at 55%. In the end we decided the best solution was that found by the 12 win setting, as it offered, in our view, the best blend of compactness and partisanship.

As mentioned, DRA uses a vote composite, and not raw presidential votes (as we do elsewhere in the paper). Thus to calculate the Democratic "vote" total N_i^D for a district i we defined it, following DRA's instructions, as the mean of,

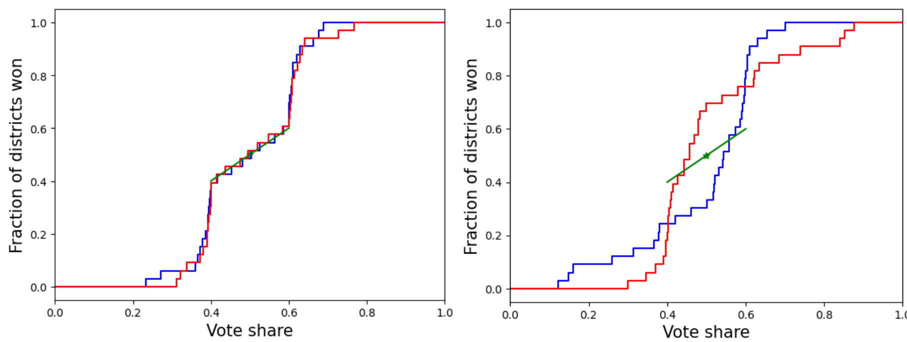
1. The mean of the Democratic votes in the previous two presidential elections.
2. The mean of the Democratic votes previous two senate elections.
3. The mean of the Democratic votes previous attorney general and governor election.

in district i. To calculate the Republican vote total for district i (N_i^R) we use the same formula, but instead count Republican votes.

5.4.2. Wisconsin Partisan Fairness

Unlike all other plans in this work, which are congressional plans, this is a state senate plan. This plan is designed with fairness in mind, but in addition, we managed to create *compact* districts (according to all plans published on DRA, this is the most compact WI state senate map). In our districts, we maintain an average *population deviation* of only 1% (so at most 0.5% from ideal – far better than the 5% required).

Our fairness metric was to aim for proportionality in the range where most vote splits happen – where each party has 40-60% of the vote. Therefore, our districts are robust to uniform swings in the electorate, and maintain their relative proportionality. Unlike the existing plan, where the swing curves show a huge Republican advantage over the range we optimize for (Figure 6b), our



(a) Our fair plan's swing

(b) Existing plan's swing

Figure 6: Win loss Seat-Vote Curves. Red for Republicans, blue for democrats, and green is proportional over the range $40\% - 60\%$. All figures and statistics calculated using published statistics and default composite vote totals specified by The Princeton Gerrymandering Project.

plan's swing curves tightly follows the proportional line (Figure 6a). In our plan, if either party gets $x\%$ of the vote, then they should hold a majority in $x\%$ of the districts (for $x \in [40, 60]$).

How we optimized for fairness in Wisconsin. First, as we did with Illinois, we used the composite vote totals that DRA recommends. We first generate "fair" plans for Wisconsin. Our energy function is simply Equation 2 (our unsigned partisan bias score) over the range $[40\%, 60\%]$. The constraints are the normal ones of a maximum half percent population deviation from ideal and connectedness..

Afterwards we take the lowest energy solution from the previous step as the starting point for another run of our algorithm. Now, with an additional constraint of keeping a fixed maximum value for Equation 2, we optimize for the Polsby-Popper score. Again, this was simply done by equating the energy function with the mean Polsby-Popper score of the districts.

6. The Ethics of Automated Redistricting

Before discussing our main results, we wish to touch upon the ethics of automated redistricting and its implications. There is an understandable concern our tool could be used to advance partisan interests. This point is especially salient for our tool, which, in hours, can match what human experts take much longer to produce.

However, the actual redistricting process takes years, and is only done once every ten years in the United States (and in many other democracies). In these situations, partisan groups would have years – and near unlimited resources – to have experts craft plans by hand, limiting the utility of an automated tool for gerrymanderers. Furthermore, the actual redistricting process involves a certain human element. When crafting a plan there is bargaining and dealing between the various interested actors. To protect their position within their district, a representative of one party may wish to keep communities of similar ethnicity, income, or shared history together. Thus they may bargain with and make concessions to members of the other party. While this behaviour would be interesting to model, it is not something a one shot algorithm is capable of.

We see our tool as something researchers can use to study redistricting. In this work, we use it to explore the impact and limitations of compactness requirements. Furthermore, it can be used to help combat gerrymandering: if a plan is as biased as the highly partisan plans produced by our tool, then there is strong evidence of gerrymandering. Because our tool is highly modular, it can be used to quickly propose alternative plans, optimizing a diverse set of desiderata.

Finally, we wish to address the recent criticisms of the Princeton Gerrymandering Project, challenging the Project's impartiality with regard to the New Jersey redistricting cycle, as they were brought in as independent advisors [48]. As noted above, we were invited to work with the Princeton group to help them and their state partners with the 2021 redistricting cycle. Because of other commitments our involvement with the project was limited, we only submitted one plan for Illinois. This plan was designed to maximize compactness, respect county boundaries, and ensure minority representation. We have never been involved in any activities regarding the New Jersey map, and at no time during our involvement were we aware of any partisan activities. As stated, our goal is to use this tool to help advance the knowledge of redistricting. We have no desire to see our tool used to advance any partisan interests.

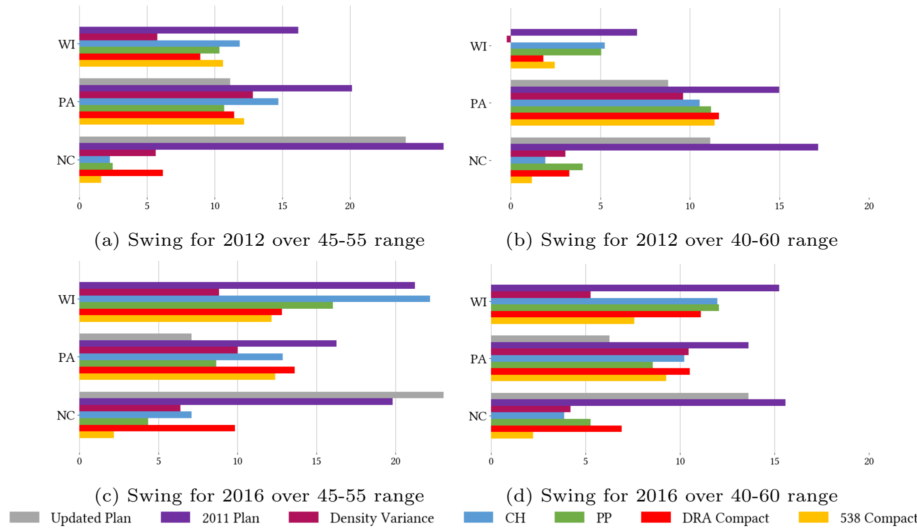
7. Fairness in Districting

We now examine the interactions of fairness and compactness using the uniform swing model, and our two robust partisan bias measures (Equations 1 and 2). Recall that the uniform swing model involves modeling hypothetical elections by adjusting actual election results by increasing (or decreasing) the vote share a target party gets in each district by the same amount.

7.1. Geographic Compactness Can Improve Fairness

As previously discussed, geographic compactness is often a primary goal in redistricting processes, and has even been suggested it is a path to partisan neutrality [49, 50]. A priori, it is not clear if geographically compact plans are more free of partisan bias than less compact ones. Thus, in this section we study plans designed to optimize various notions of compactness, contrasting them with the currently used plans. To that end we use the two compact plans from 538 and DRA. We also use the Polsby-Popper and Convex Hull compact plans from our algorithm.

We find that optimizing for any form of geographic compactness yields plans that have improved partisan fairness relative to the plans enacted in 2011, according to our signed partisan bias score (over the range $[0.4, 0.6]$ or over the



(a) Swing for 2012 over 45-55 range

(b) Swing for 2012 over 40-60 range

(c) Swing for 2016 over 45-55 range

(d) Swing for 2016 over 40-60 range

Figure 7: Average signed distance from the R swing curve to the $y = x$ line over the indicated range in the specified election (Equation 1). In each state there are four compact plans, DRA's, 538's, our Convex Hull, and our Polsby-Popper. In PA and NC two implemented ones (2011 and Updated), the WI 2011 plan was not struck down so there is no WI updated plan. Finally, there is our plan which ensures node density within a district is homogenous

range $[0.45, 0.55]$). This improvement is consistent across all states and independent of the measure optimized. Figure 7 shows our signed partisan bias score (closer to zero is fairer) for various plans in three states using either presidential election or either comparison range.

This improvement is sometimes extreme: in NC, the 2011 districting (with a 17% robust partisan bias towards the R party) is more than two times as biased as any of the compact plans. It is worth noting that both NC and PA 2011 plans were struck down by the courts for being overly biased. The NC 2011 plan was found to disenfranchise minority voters [4], while in PA the plan was found to disenfranchise Democrats [51]. The Republican advantage, and improved fairness with compact plans, continue to hold for each state when using using different election data and swing ranges. When we use $[45, 55]$ instead of $[40, 60]$, the only time the 2011 plan is less R-biased is the 2016 election in WI when compared to the CH optimal plan (and this is only a 1% difference).

Interestingly, the updated plans from 2016 in NC and 2018 in PA seem dissimilar. The updated NC plan is still significantly more R-biased than any of the compact ones, the opposite holds for the new PA plan. The R-bias of the PA plan is lower than in the compact plans, although it is, of course, less compact according to almost any metric (Table 3). It has been suggested the new PA plan was designed with partisan proportionality in mind [52], though the plan designers have not commented on their process. In any case, of course none of the compact plans are designed to optimize for Equation 1 (signed partisan bias). In each state, when we use our algorithm to optimize for this metric specifically we find plans that have near-zero bias according to Equation 1. These are not shown in Figure 7 because the bars would be virtually invisible.

In general, it was not the case that one definition of compactness was always superior to others in terms of partisan fairness. For example, in NC the 538 compact plan was the closest to zero for the signed partisan bias score (Equation 1), but the opposite happened in PA. We speculate that the difference political geography of NC and PA causes these differences. Recall, PA's large urban centres are in the east and west of the state. On the other hand, NC has its urban centres in the middle of the state. This, amongst other factors, may impact which compact plan is the fairest according to our definition. That being said, in all states using any comparison range and either presidential election, all of the compact plans are similarly fair. And we reiterate, they are all very fair compared to the plans implemented in 2011.

For all plans, in all states, both elections, and both ranges of comparison there is one near consistent pattern: The R party has a positive score in our metric, and from symmetric considerations noted above, this means a negative D score. That is, the more rural party can expect to gain more seats than its proportional voter share. This includes every single plan designed to optimize some notion of geographic compactness. And the 2018 PA plan which was supposedly also designed to consider proportional fairness. The only exception to this was Wisconsin, in 2012, over the 40-60 comparison range, with our plan that optimizes for districts having nodes of similar population density. Here, the Democrats had a slight partisan advantage beyond what is fair. We will explore these density plans in more detail in Section 7.2.

7.2. Voting With Similar People Can Improve Fairness

	NC	PA	WI
Updated Plan	1.21	1.28	N/A
2011 Plan	1.22	1.25	1.60
Density Variance	**1.02**	**1.07**	**1.39**
CH	1.28	1.27	1.61
PP	1.23	1.30	1.66
DRA Compact	1.23	1.29	1.62
538 Compact	1.27	1.29	1.68

Table 4: Measure of our density variance metric for each plan (rows) in each state (columns). Bolded row is the best (lowest) score in each state.

We now turn to explore plans optimized for our density deviation metric. These plans aim to place precincts of similar density together in the same district. In other words, dense urban precincts should be in districts together, and sparse rural precincts should be in districts together. Due to the rural/urban divide, one could argue that such a metric would effectively be gerrymandering, but beyond the qualitative argument above (such districts would maintain communities of a a similar type together), one can argue this is not a real gerrymandering strategy: by "packing" rural voters together and urban voters together, there is no clear advantage to anyone of those, and voters in mid-density areas (which are also packed together) are not closely affiliated with a partisan hue.

To do the actual optimization we used an identical technique as we used when optimizing for our Convex Hull and Polsby-Popper plans (see Section 5.3).

Let us first examine the density deviation value for each plan of interest. Unsurprisingly, as Table 4 shows, our plan optimized for the density variance is the best for that metric. Also interesting, existing plans are the second for our metric in each state (though there was not a large variance between the non-density-optimized plans). While some of this may be the byproduct of packing for gerrymandering (i.e., putting all of a party's opponents in districts where they achieve high – but useless – voting margins), we can not assign all of the "blame" on it. A very reasonable goal in redistricting is keeping communities of interest together. Although what constitutes a community of interest is often vaguely defined, keeping similar neighbourhoods, where people live in similar situations, and thus in similar densities, is a reasonable interpretation of communities of interest. Keeping smaller towns, which may be of similar density, with historic ties together is also reasonable interpretation of communities of interest. Both of these goals, and many variants of these goals, would likely contribute to lowering the score for our density deviation metric.

Regardless of the potential gerrymandering concerns, by optimizing for low density variance we do end up with more "fair" plans. As Figure 7 shows, in every situation our density deviation plans are far closer to fair according to Equation 1 than the implemented 2011 plans in every state, and the updated plan in North Carolina. In fact, the values for Equation 1 are almost always in line with those of the compact plans. Interestingly, in Wisconsin, in every situation, our density deviation plan gave the lowest Republican advantage. In one situation, the $[40, 60]$ range in 2012 the Democrats actually had a slight advantage. Why our Density Deviation plans in Wisconsin behaved differently is hard to explain. One obvious explanation could be the very non-convex shape of the state: the state's north-east is carved up by Lake Michigan, limiting the impact of geographic compactness measures. Other issues may be at play, including how rapidly voter preferences change as one moves from urban to rural are different in each state. Finally, we note that while our density deviation plan showed a slight Democratic advantage over the $[40, 60]$ range in 2012 it was fairly small. Furthermore, if we limit the range to $[45, 55]$ the the Republicans again have the advantage. That is, only when we consider the most extreme of the reasonable elections, do we see a Democratic advantage.

7.3. An α% of the Vote can be an α% of the Seats

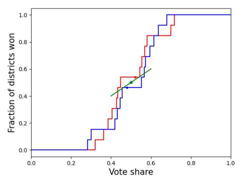
As we saw with the updated PA plan and the compact plans, optimizing purely for compactness may not be the be the most effective way to eliminate partisan bias. For each state we use our algorithm to show there is a plan that effectively has no bias, a "fair" plan. We use use our algorithm to optimize for the unsigned partisan bias, Equation 2, over the range $[40, 60]$ (optimizing for the signed version, Equation 1, can lead to plans with huge jumps in the swing curve).

To actually create the fair and compact plans seen in Figure 10 we use the following procedure: For each state and its associated election of interest, we use our algorithm on 96 cores for 24 hours. Here we set our energy function to be Equation 2 (the unsigned partisan bias score). The lowest energy plan for each state is our fair plan. In each state we then take this fair plan, and use it as the starting point for another run of our algorithm. Now, our goal is to make these fair plans more compact. We do this by setting the energy function to be the mean Polsby-Popper score across each district. In addition we add a constraint that the value found for Equation 2 can not become much higher than the value in the initial plan. For Wisconsin the initial plan had a value of 3.5% and our limit was 4%, North Carolina was initially 2.1% and limited to 2.5%, and Pennsylvania was initially 1.4% and limited to 2%. For each state we run our algorithm across 48 cores for another 24 hours.

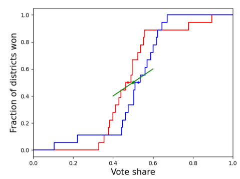
In each state the resulting plans (Figure 10) had an unsigned partisan bias score from 3.5 to 10 times lower than any of the the existing plans (Figure 9) and compact plans (Figure 8). Furthermore, in these plans there is near zero advantage at for either party, when measured by the signed partisan bias. Beyond the value of signed/unsigned bias (Equations 2 and 1) the results are striking. In our fair and compact and fair plans, either party having an α fraction of the vote (for $\alpha \in [0.4, 0.6]$) means they would have had a majority in an α fraction of the districts. This is unlike the existing and compact plans where for almost every value around the target range, either one or both parties would get representation far beyond their vote share.

The gain in fairness over the implemented and compact plans, comes at a certain cost of compactness. While we did make these fair plans more compact, they are, naturally, not as compact as our most compact plans. As Table 3 shows, in each state our fair but compact plans were more compact than the existing plans, though not as compact as the plans designed to optimize for a particular compactness metric.

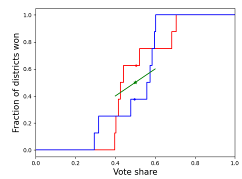
Looking at the non-optimized plans, unlike when measuring partisanship with signed bias metric (Equation 1), the fairness advantage of the compact plans with the unsigned metric (Equation 2) is reduced. In North Carolina the compact plans generally score lower than the two implemented plans. But in Pennsylvania and Wisconsin the gap between compact and implemented is far lower, and sometimes is even reversed. As was the case with signed bais there wasn't any one compact plan that was universally optimal for unsigned bias. In North Carolina our Polsby-Popper plan was by far the best out of the implemented and existing plans. In Pennsylvania all the compact and implemented plans had similar scores (the DRA plan had the lowest value). Finally, in Wisconsin the 538 plan had the lowest score. But, all in all, both compact and implemented plans had similar scores (with the single exception of Wisconsin's convex hull plan, which had a very large value compared to the rest).



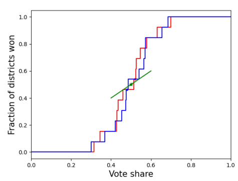
(a) NC PP 2016 (8.9%)



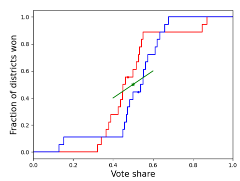
(b) PA PP 2012 (14.8%)



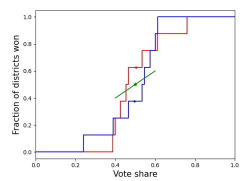
(c) WI PP 2016 (15.4%)



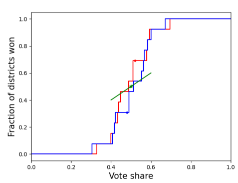
(d) NC 538 2016 (13.3%)



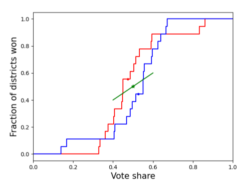
(e) PA 538 2012 (16.5%)



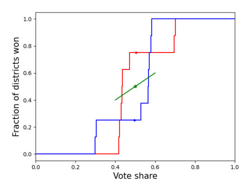
(f) WI 538 2016 (13.6%)



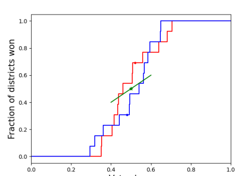
(g) NC CH 2016 (12.8%)



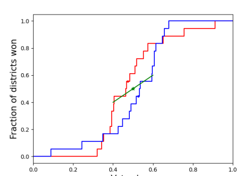
(h) PA CH 2012 (15.2%)



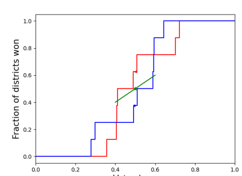
(i) WI CH 2016 (21.7%)



(j) NC DRA 2016 (12.2%)

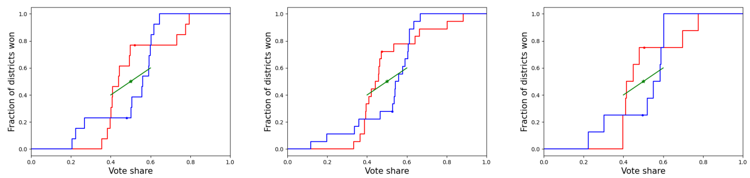


(k) PA DRA 2012 (12%)



(l) WI DRA 2016 (13.1%)

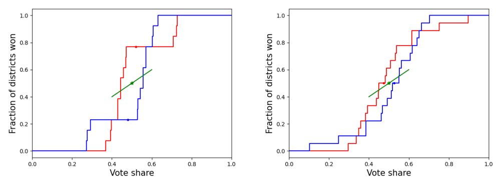
Figure 8: Uniform swings for $R(D)$ in red (blue), in indicated state, compact plan, and presidential election year. The value for Equation 2 (unsigned bias metric) is given in the parentheses. Vertical axis shows the fraction of districts won; horizontal axis the vote fraction. The dots on the party curves indicate the actual election outcome (0 swing). The green line is the range of proportional outcomes on the range $[0.4, 0.6]$. A green star marks the point $(1/2, 1/2)$.



(a) NC original 2016 (16.4%)

(b) PA original 2012 (15.9%)

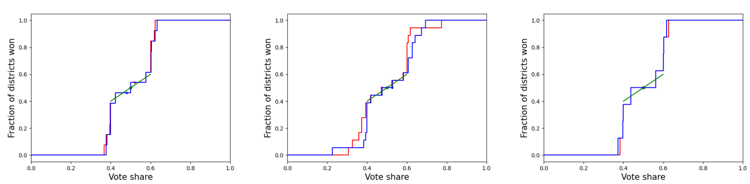
(c) WI original 2016 (17.1%)



(d) NC updated 2016 (19.6%)

(e) PA updated 2012 (12.5%)

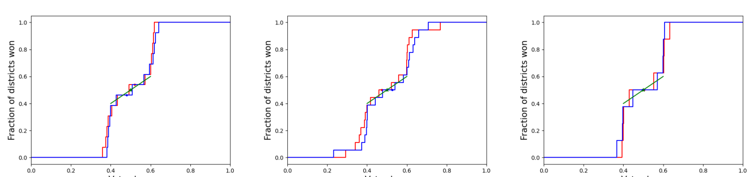
Figure 9: Uniform swings for $R(D)$ in red (blue), in indicated state, plan (either existing or updated where applicable), and presidential election year. The value for Equation 2 (unsigned bias metric) is given in the parentheses. Vertical axis shows the fraction of districts won; horizontal axis the vote fraction. The dots on the party curves indicate the actual election outcome (0 swing). The green line is the range of proportional outcomes on the range $[0.4, 0.6]$. A green star marks the point $(1/2, 1/2)$.



(a) NC fair 2016 (2.1%)

(b) PA fair 2012 (1.4%)

(c) WI fair 2016 (3.5%)

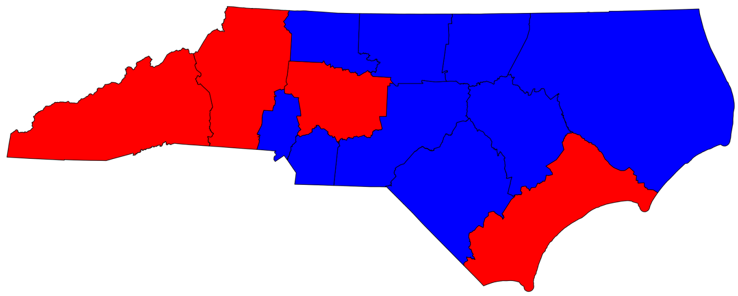


(d) NC fair & compact 2016 (2.3%)

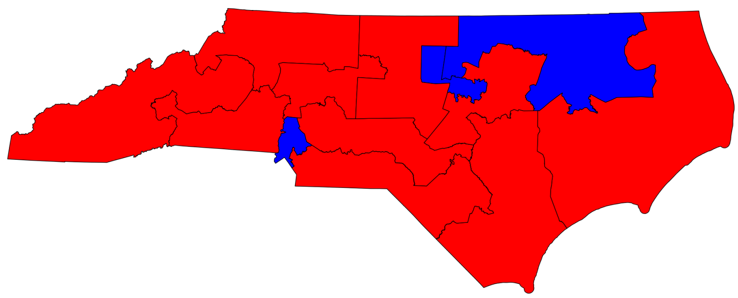
(e) PA fair & compact 2012 (1.9%)

(f) WI fair & compact 2016 (3.7%)

Figure 10: Uniform swings for $R(D)$ in red (blue), in indicated state, plan (either our fair or our fair and compact), and presidential election year. The value for Equation 2 (unsigned bias metric) is given in the parentheses. Vertical axis shows the fraction of districts won; horizontal axis the vote fraction. The dots on the party curves indicate the actual election outcome (0 swing). The green line is the range of proportional outcomes on the range $[0.4, 0.6]$. A green star marks the point $(1/2, 1/2)$.



(a) Our NC 9 D wins



(b) 2016 NC plan

Figure 11: NC districts (R wins in red); (D wins in blue) based on the 2016 NC election data. Top, our plan, optimizing the Polsby-Popper score while maintaining 9 D wins (where a win is simple majority). Bottom, NC's 2016 court mandated plan.

8. Designing Partisan Plans

We now focus on the limits compactness thresholds impose on partisan gerrymandering. As we saw, compactness as a goal can lead to more balanced outcomes. However, if our goal is to use compactness as a constraint to improve representability, rather than as an objective, it is not clear which restrictions are necessary. For compactness, we shall use the average Polsby-Popper score of a plan. Initially, to measure gerrymandering ability, we use *gerrymandering power* (introduced in Borodin et al. [35]). For a particular election, the gerrymandering power of party p is defined as the difference between the share of seats it can optimally gerrymander to win and the seat share it would have received in a purely proportional election. A high gerrymandering power indicates there is a plan that uses p's vote geographic layout to produce a disproportionately large number of districts. A low (or negative) gerrymandering power indicates p's geographic spread is such that it is unable to stretch its vote into many extra wins (or even win a proportional number of seats). An extreme example is when the population is completely homogenous: if each house contains 2 R voters and 3 D voters, the R party can never win any district, and has a negative gerrymandering power, while the D party will always win 100% of districts. We will also explore if gerrymandering while maintaining a large margin in the districts a party has won substantially impacts gerrymandering ability, and if compactness restrictions are any more effective in this situation.

To gerrymander for party p while staying compact, we run our algorithm with the objective of generating plans which are as compact as possible while maintaining k wins[23] for party p. To ensure a diversity of election outcomes we use the elections specified in Section 4. As we saw in previous sections, our algorithm is capable of generating highly compact districts and highly partisan districts. Unsurprisingly, we find it performs quite well when combining these goals. As Figure 11 shows, our compact gerrymander for the Democrats (Figure 11a) easily passes the eye test, especially when compared to the implemented plan (Figure 11b). In NC our algorithm can stretch the number of districts the Rs win to all 13 using the 2016 election data, all while creating a plan more compact than the existing one (in the existing plan, the Rs won 10 of 13). We can also create a map for the Ds where they win 8 more seats than in the existing plan, while being more compact than it.[24].

To produce these maps we first generate highly partisan outcomes: assuming we are gerrymandering for the Democrats, given a partitioning of the nodes of G into $S = (S_1, \cdots, S_K)$ we set Equation 3 to:

$$E(S) = K - \sum_i v_D(S_i), \tag{13}$$

where $v_D(S_i)$ is equal to :

$$v_D(S_i) = \begin{cases} \frac{N_i^D}{N_i^D + N_i^R} & \frac{N_i^D}{N_i^D + N_i^R} \leq \tau \\ 1 & otherwise \end{cases}$$

Here N_i^D is the total Democratic vote in district i (N_i^R is the total Republican vote in district i). If we want to gerrymander for the Republicans, replace $v_D(S_i)$ with $v_R(S_i)$ which is defined as follows:

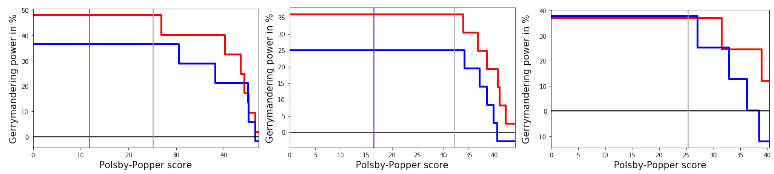
$$v_R(S_i) = \begin{cases} \frac{N_i^R}{N_i^D + N_i^R} & \frac{N_i^R}{N_i^D + N_i^R} \leq \tau \\ 1 & otherwise \end{cases}$$

We do three runs, $\tau \in \{0.5, 0.53, 0.56\}$, for defining victory as simple majority and strong wins. This is similar to our method for emulating 538, but now the sigmoid function's contribution to the energy has been replaced by a linear distance to winning the district.

Without loss of generality, assume we are gerrymandering for party P. For our first phase, in each state for P we run our method 288 times for 48 hours. This time limit was more than sufficient for the convergence of the various processes. This first phase gives us several runs that have the most possible wins for P in each state, call this set of solutions W_{max}. Then for P in each

[23]By win we mean party p has at least a $(50 + \tau)\%$ of the vote, for various settings of a safety parameter τ.

[24]For these values, we are referring to a win as a simple majority of the vote.



(a) NC 2016 (b) PA 2012 (c) WI 2016

Figure 12: Gerrymandering power when faced with a minimum required Polsby-Popper score using data from the 2012 PA presidential election and 2016 NC and WI presidential elections. R in red; D in blue. The vertical purple (grey) line is the Polsby-Popper score of the 2011 congressional plan (court mandated plan). Average distance between the two curves is a 10.8% R advantage in PA, 10.4% R advantage in NC, and 4.3% R advantage in WI.

state, we have a range of potential win values $\{w_{max}, \cdots, w_{min}\}$ (where w_{max} is the most number of wins found for P in the first phase and w_{min} is the number of wins for P in the optimally compact solution from the Polsby-Popper plan in Section 5.3).

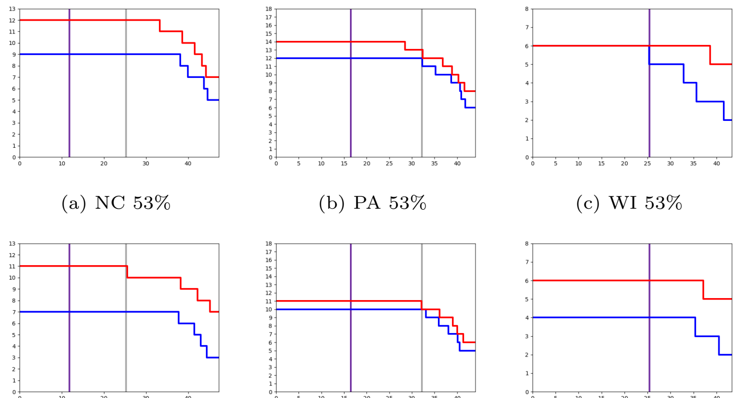
For each value $w \in \{w_{max}, \cdots, w_{min}+1\}$ we execute the following procedure: dividing the solutions of W_{max} among 96 cores as evenly as possible, we run our algorithm using these solutions as the initial plans and optimize for the Polsby-Popper score (as usual, by setting the energy function to be the mean Polsby-Popper score across all districts). We add the additional constraint that for a proposal to be considered in algorithm 1 we require that the total number of wins for P is at least w (keeping the same τ used to calculate w_{max}). We found the algorithm converged to a solution well before the 24 hour cutoff for each of these simulations.

8.1. Effect of Increasing Compactness on Gerrymandering Power

While increasing the required mean Polsby-Popper score lowers the gerrymandering power of both parties, to have an impact, a steep increase beyond what current plans use (and often near the most compact) is required. In each state Figure 12 shows existing plans do not have a compactness score that constrains any party's gerrymandering power.

Additionally, compactness requirements are unable to entirely remove the urban disadvantage. For almost any Polsby-Popper score, the R gerrymandering power is well above the D one. In PA there is no requirement level where the Ds have an advantage. In NC and WI there is a brief period of near maximum compactness requirements where the Ds have a small, temporary, advantage. In WI, when the compactness requirement is lower than that of the current plan the democrats can have a minuscule advantage, but more stringent requirements give a large R advantage. The average distance between the two curves shows a 10% R advantage in gerrymandering power in PA and NC, and 4% in WI.

Moreover, in every single state, even with the most extreme compactness requirements, Rs are able to stretch their vote share beyond proportional. On the other hand, Ds, even if they can have any legal plan they desire, have



(a) NC 53%　(b) PA 53%　(c) WI 53%

(d) NC 56%　(e) PA 56%　(f) WI 56%

Figure 13: Ability to gerrymander at various levels of win robustness (indicated in the parentheses) when faced with a minimum required Polsby-Popper score. Vertical axis is the number of winnable districts. Using data from the 2012 PA presidential election and 2016 NC and WI presidential elections. R in red; D in blue. The vertical purple (grey) line is the Polsby-Popper score of the 2011 congressional plan (court mandated plan). First row, was with a 53% win threshold, second row was with a 56% win threshold.

a negative gerrymandering power when compactness requirements are high – they cannot even reach their proportional allocation.

8.2. Robust Wins While Staying Compact

In Section 8.1, we assumed a margin of victory of a single vote was sufficient. In reality, when assessing the future performance of a plan, historical vote data is only an estimate. Thus, to protect against small (or even large-ish), swings against their party, parties design plans with a robust margin of victory. What is considered robust enough is a matter for debate – 538's probabilistic model considered an Republican advantage of about 5% from the 2012 and 2016 presidential elections a safe margin for them, while for the Democrats a 12% advantage was considered safe. On the other hand, Dave's Redistricting App considers a 10% advantage in their vote composite, which is built from presidential, senate, gubernatorial, and attorney general elections, as safe for either party.

We opt for a simple interpretation of safe. We say a 6% advantage in a presidential election of interest is safe, and a 12% advantage (in the same presidential election) is very safe. We now repeat the exact same set up from Section 8.1, but set the victory thresholds that parties are aiming for at 53% and 56%.

The advantage the urban-rural divide provides to the more rural Republican party is even more stark here. As Figure 13 shows, in every state for either win threshold level, there is no point where the Ds are able to gerrymander to win more seats at than the Rs can gerrymander. That is, no matter how we define wins, no matter how much compactness we require, the Democrats can never outperform the Republicans, and often trail them significantly. There is the occasional point in WI and PA where – when faced with the same Polsby-Popper requirement – both parties can win the same number of seats (at the same victory threshold). But this is always temporary, The Rs always recover their advantage. This is especially shocking since in PA in 2012 the Ds have a significant vote advantage.

These results (and those concerning the gerrymandering power in the previous section) do not mean the Democrats are incapable of gerrymandering. Indeed, it is quite the opposite. In every state, even when requiring a 56% win threshold the Ds (and hence the Rs) are able to win at least half the seats. At a 52% threshold the Ds can always take at least 65% of the seats. These facts are even true in NC where the Ds have a severe vote disadvantage. Even if we require the plan be at least as compact as any of the implemented ones the Ds are able to gain at least half the seats at a 56% threshold and a strict majority at 53% a threshold. These results only show that the Rs always have a higher ceiling for gerrymandering potential than the Ds. For example, as Figure 13d shows in NC the most compact outcome gives the Rs as many wins at 56% as the Ds get when the Ds engage in unrestricted gerrymandering.

9. Discussion

In this work we introduced a modular and powerful automated redistricting technique. Our technique can generate plans comparable to ones from human experts for both partisan and non-partisan goals. Our method is able to generate geographically compact districts, far more compact (according to various metrics) than the plans used in practice or the ones produced by electoral experts. While the plans, which were optimizied for geographic compactness, reduce partisan bias, we find they do not eliminate it and still always favor the rural party. We also used our algorithm to explore density deviation, a novel definition of compactness based around how homogenous a district is relative to the density of the precincts that make it up. We also saw optimizing for our novel metric can lead to more fair plans, and in one rare case, a plan that favours the more urban Democrats. Despite this rural-favouring voter geography, we show there are plans which are near totally proportionally fair, but to achieve this we must sacrifice geographic compactness to some extent. Finally, we use our algorithm to explore the effects of an often proposed solution to gerrymandering: compactness restrictions. We find that while this can reduce the ability of either party to gerrymander, the potential for some degree of gerrymandering remains, and the rural party can still gerrymander more than its urban counterpart. These results contribute to growing evidence that the urban-rural divide leads to imbalanced outcomes that disadvantage the urban party.

We intend to use our algorithm to explore the tradeoff between partisan fairness and non-partisan goals, such as not splitting counties. We began exploring the tradeoff between fairness and compactness in Section 7.3, the next step would be mapping out its Pareto frontier.As we saw, our maps which aim to ensure districts are homogenous in terms of population density often had low partisan measures, and in one situation even favoured the more urban Democratic party. We believe this effect is worth exploring further. It could be other deviation metrics would see similar results; for example, we could have instead used the Gini coefficient for measuring the variance in a district's density. More generally, perhaps other ways of grouping similar voters together would lead to similar results. If we design plans where each district is homogenous according to income, or minority population, or even vote intention, would we see similar results?

We are also exploring other, novel, definitions of compactness. Many such measures take advantage of the underlying precinct graph structure. Perhaps the compactness of a district could be defined as the number of edges the average person needs to traverse to reach each other person in their district. Leaving compactness aside, the possibility of gerrymandering under multi-party systems is not well understood, and even the effects of minor parties (as exist even today in the US) on gerrymandering should be examined.

Possibly more important than partisan considerations is ensuring that minority voices are heard in the political process. We are investigating criteria for ensuring that minority voters receive their deserved representation in redistricting. Beyond the basic requirements of majority-minority districts that satisfy the US 1965 Voting Rights Act, the function for measuring minority representation could be quite intricate and difficult for human experts to analyze and optimize for. We believe that such non-trivial objective functions, along with restrictions such as compactness, make this problem an ideal application of our algorithm.

References

[1] D. Wasserman, Hating gerrymandering is easy. fixing it is harder, FiveThirtyEight, 2018. URL: `https://fivethirtyeight.com/features/hating-gerrymandering-is-easy-fixing-it-is-harder/`.

[2] A. Wheeler, P. N. Klein, The impact of highly compact algorithmic redistricting on the rural-versus-urban balance, in: SIGSPATIAL '20: 28th International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, November 3-6, 2020, 2020, pp. 397–400.

[3] S. S.-H. Wang, Three tests for practical evaluation of partisan gerrymandering, Stanford Law Review 68 (2016) 1263–1321.

[4] Common Cause v. Lewis, Common Cause v. Lewis, 834 S.E.2d 425, Wake County Superior Court, 2019.

[5] B. Bishop, The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart, Mariner Books, 2009.

[6] N. Ward, Representing rurality? new labour and the electoral geography of rural britain, Area 34 (2002) 171–181.

[7] M. Kenny, D. Luca, The urban-rural polarisation of political disenchantment: an investigation of social and political attitudes in 30 European countries, Cambridge Journal of Regions, Economy and Society 14 (2021) 565–582.

[8] R. Noack, The urban-rural divide that bolstered Trump isn't just an American thing; it's prevalent in Europe, too, Washington Post (2016).

[9] D. E. Valero, From brexit to vox: Populist policy narratives about rurality in Europe and the populist challenges for the rural-urban divide, Rural Sociology 87 (2022).

[10] A. Borodin, O. Lev, N. Shah, T. Strangway, Little house (seat) on the prairie: Compactness, gerrymandering, and population distribution, in: Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Auckland, New Zealand, 2022, pp. 154–162.

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

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

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

[14] Rucho v. Common Cause, Rucho v. Common Cause, number 18-422 in 588 U.S., Supreme Court of the United States, 2019.

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

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

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

[18] B. Grofman, W. Koetzle, T. Brunell, An integrated perspective on the three potential sources of partisan bias: Malapportionment, turnout differences, & the geographic distribution of party vote shares, Electoral Studies 16 (1997) 457–470.

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

[20] D. S. Felsenthal, N. R. Miller, What to do about election inversions under proportional representation?, Representation 51 (2015) 173–186.

[21] B. Grofman, G. King, The future of partisan symmetry as a judicial test for partisan gerrymandering after LULAC v. Perry, Election Law Journal 6 (2007) 2–35.

[22] C. Puppe, A. Tasnádi, A computational approach to unbiased districting, Mathematical and Computer Modelling 48 (2008) 1455–1460.

[23] B. Fifield, M. Higgins, K. Imai, A. Tarr, A new automated redistricting simulator using markov chain monte carlo, 2018.

[24] W. Pegden, A. D. Procaccia, D. Yu, A partisan districting protocol with provably nonpartisan outcomes, ArXiv:1710.08781, 2017.

[25] Y. Bachrach, O. Lev, Y. Lewenberg, Y. Zick, Misrepresentation in district voting, in: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI), New York, 2016, pp. 81–87.

[26] R. van Bevern, R. Bredereck, J. Chen, V. Froese, R. Niedermeier, G. J. Woeginger, Network-based vertex dissolution, SIAM Journal on Discrete Mathematics 29 (2015) 888–914.

[27] S. Bervoets, V. Merlin, Gerrymander-proof representative democracies, International Journal of Game Theory 41 (2012) 473–488.

[28] O. Lev, Y. Lewenberg, "Reverse gerrymandering": a decentralized model for multi-group decision making, in: Proceedings of the 33rd Conference on Artificial Intelligence (AAAI), Honolulu, Hawaii, 2019, pp. 2069–2076. doi:https://doi.org/10.1609/aaai.v33i01.33012069.

[29] Z. Fitzsimmons, O. Lev, Selecting voting locations for fun and profit, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), Kyoto, Japan, 2020, pp. 224–230. doi:https://doi.org/10.24963/ijcai.2020/32.

[30] W. Vickrey, On the prevention of gerrymandering, Political Science Quarterly 76 (1961) 105–110.

[31] D. DeFord, M. Duchin, J. Solomon, Recombination: A family of markov chains for redistricting, Harvard Data Science Review (2021). Https://hdsr.mitpress.mit.edu/pub/1ds8ptxu.

[32] M. Duchin, Outlier analysis for Pennsylvania congressional redistricting, Technical Report, Metric Geometry and Gerrymandering Group, Tisch College, Tufts University, 2018. URL: https://mggg.org/publications/md-report.pdf.

[33] A. Cohen-Zemach, Y. Lewenberg, J. S. Rosenschein:, Gerrymandering over graphs, in: Proceedings of AAMAS, Stockholm, Sweden, 2018, pp. 274–282.

[34] Y. Lewenberg, O. Lev, J. S. Rosenschein, Divide and conquer: Using geographic manipulation to win district-based elections, in: Proceedings of the 16th International Coference on Autonomous Agents and Multiagent Systems (AAMAS), São-Paulo, Brazil, 2017, pp. 624–632.

[35] A. Borodin, O. Lev, N. Shah, T. Strangway, Big city vs. the great outdoors: Voter distribution and how it affects gerrymandering, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI), Stockholm, Sweden, 2018, pp. 98–104.

[36] W. Gurnee, D. B. Shmoys, Fairmandering: A column generation heuristic for fairness-optimized political districting, 2021. arXiv:2103.11469.

[37] M. Dyer, A. Frieze, On the complexity of partitioning graphs into connected subgraphs, Discrete Applied Mathematics 10 (1985) 139–153.

[38] A. Becker, J. Solomon, Redistricting algorithms, Political Geometry (2021).

[39] B. Klaas, Gerrymandering is the biggest obstacle to genuine democracy in the united states. so why is no one protesting?, Washington Post (2017).

[40] New York Times Editorial Board, A shot at fixing american politics, New York Times (2017) A22.

[41] J. Rodden, Why Cities Lose: The Deep Roots of the Urban-Rural Political Divide, Basic Books, 2019.

[42] J. Levitt, 2020, Where are the lines drawn?, URL: https://redistricting.lls.edu/redistricting-101/where-are-the-lines-drawn/.

[43] B. Grofman, Measures of bias and proportionality in seats-votes relationships, Political Methodology (1983) 295–327.

[44] A. Bycoffe, E. Koeze, D. Wasserman, N. Silver, 2018, We drew 2,568 congressional districts by hand. here's how, URL: https://fivethirtyeight.com/features/we-drew-2568-congressional-districts-by-hand-heres-how/.

[45] MGGG, Metric Geometry and Gerrymandering Group; https://mggg.org, 2020. Accessed: 2021-01-01.

[46] N. Silver, 538: The gerrymandering project, https://fivethirtyeight.com/tag/the-gerrymandering-project/, 2020.

[47] C. Cook, The Cook political report, https://cookpolitical.com/, 2018. Accessed: 2021-01-01.

[48] D. Wildstein, Claiming partisan fairness models are proprietary, princeton gerrymandering project won't show their work, New Jersey Globe (2022).

[49] C. Ingraham, This computer programmer solved gerrymandering in his spare time, Washington Post, 2014. URL: https://www.washingtonpost.com/news/wonk/wp/2014/06/03/this-computer-programmer-solved-gerrymandering-in-his-spare-time/.

[50] C. Ingraham, This is actually what america would look like without gerrymandering, Washington Post, 2016. URL: https://www.washingtonpost.com/news/wonk/wp/2016/01/13/this-is-actually-what-america-would-look-like-without-gerrymandering/.

[51] LWV v. Pennsylvania, League of Women Voters of Pennsylvania v. Commonwealth of Pennsylvania, number 159 in MM 2017, Supreme Court of Pennsylvania, 2018.

[52] N. Cohn, M. Bloch, K. Quealy, The new Pennsylvania congressional map, district by district, New York Times, 2018.