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

Allan Borodin  
University of Toronto  
Toronto, Canada  
bor@cs.toronto.edu

Omer Lev  
Ben-Gurion University of the Negev  
Beersheba, Israel  
omerlev@bgu.ac.il

Nisarg Shah  
University of Toronto  
nisarg@cs.toronto.edu

Tyrone Strangway  
Ben-Gurion University of the Negev  
Beersheba, Israel  
tyrone@cs.toronto.edu

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 show there are plans which are fair from a partisan perspective, but they are far from optimally compact.

KEYWORDS
Voting, Gerrymandering, Redistricting, Compactness, Social Choice, Fairness

ACM Reference Format:

1 INTRODUCTION
In many democracies, politicians are elected to represent the people of particular geographic areas, called districts. There is country wide aggregation of votes. 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.

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, there are many competing goals when designing a districting plan [42]. One could prioritize not breaking up communities of interest, such as those with a shared culture and history.1 It may also be desirable to be as compatible as possible with established city and county boundaries, a consideration studied by Wheeler and Klein [43]. Another reasonable goal would be to obtain geographically compact regions (a goal enshrined in some US states’ laws and regulations2). 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.

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 [41]. For example, in the 2020 federal election in North Carolina, a state accused of gerrymandering (partially overturned by courts [8]), the Democratic party received 49.96% of the vote and won five districts;
we see that this advantage is robust even in a non-gerrymandered, while the surrounding rural regions lean Republican. In short, in when it comes to non-partisan redistricting (see Wasserman [42] for further discussion and a comparison of objectives).

Parallel to the political partisan redistricting process there is a more complex, ongoing population-wide process. In the US [3] and Europe [31], voters are "reorganizing" themselves for various economic and social reasons. As Figure 1 shows, in Pennsylvania voters of left leaning parties tend to cluster in dense urban centres, while voters of right leaning parties spread out into surrounding rural regions. People in the large cities of Philadelphia in the east and Pittsburgh in the west are overwhelmingly Democrat voters, while the surrounding rural regions lean Republican. In short, in many democracies around the world the left-right split can be characterized as an urban-rural divide as well.

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 5.4) 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 6, 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, optimally-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 Section 7 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.

3 MODEL

We examine gerrymandering with a graph-theoretic formulation. We shall use a 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. An edge \((u, v) \in E\) represents that precincts \( u \) and \( v \) share a physical boundary. For \( v \in V \) let \( n_v \) be the number of people who live in precinct \( v \), and \( n_{v,e} \) be the number of people who live in \( v \) and vote for party \( p \) in election \( e \). We will omit \( e \) when the context is obvious. Let \( N = \sum_{v \in V} n_v \) be the total number of people in the state. We limit our focus to two parties: the rural party (in the US, Republicans \( R \)), and the urban party (in the US, Democrats \( D \)).

Creating a districting plan requires partitioning \( G \) into \( K \) vertex-disjoint subgraphs \( G_1, \ldots, G_K \) (the districts). The number of districts \( K \) is extrinsically determined (in the US, by a census every 10 years). 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 from any point in the district to any other point in the district without crossing into another district.

Population balance-$\delta$ 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 $\delta$ required in the U.S. changes between states (and judicial decisions). Informally, districts should be as equal-sized in population as possible [25]. 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 fairer constraint than what is respected by many previously proposed automated redistricting methods.

Parties are interested in 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 $\sum_{v \in V(G_k)} n_v^{D,\text{2012}} > \sum_{v \in V(G_k)} n_v^{R,\text{2012}}$, we say the Democrats win district $k$ according to the 2012 presidential vote totals in a given state. If the inequality is reversed, we say the Republicans win the district in that election. Note that this is only one way to define “winning a district”. In Section 5.1 we calculate the probability of winning a hypothetical district, based on historic vote totals and outcomes, and find plans with several high probability wins.

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. These are states and elections for which granular, precinct-level data is available. Each of these three states has a particular election of interest. We also include the number of nodes (precincts) and edges in the graphs of each state.

Pennsylvania (PA) 2012 Sizeable Democrat advantage. The Democrat (Obama) won 51.97% of the vote, vs. 46.59% to the Republican (Romney). PA has 9,255 nodes and 25,721 edges. North Carolina (NC) 2016 Sizeable Republican advantage. The Republican (Trump) won 49.83% of the vote, vs. 46.17% to the Democrat (Clinton). NC has 2,692 nodes and 7,593 edges. Wisconsin (WI) 2016 Near tie. The Republican (Trump) won 47.22% of the vote, vs. 46.45% to the Democrat (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. 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 Voting and the Urban-Rural Divide

As noted above, a key geographic feature of US political parties is the growing divide between a more rural Republican party and a more urban Democratic party. We examined 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 (see Table 1).

<table>
<thead>
<tr>
<th>State</th>
<th>2012 correlation</th>
<th>2016 correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>PA</td>
<td>0.47</td>
<td>0.58</td>
</tr>
<tr>
<td>WI</td>
<td>0.38</td>
<td>0.57</td>
</tr>
</tbody>
</table>

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.

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 plans 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 the robust partisan bias metric introduced in Section 6), 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 (e.g., the Polsby-Popper and Convex Hull scores defined in Section 5.2). 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 now demonstrate our algorithm is capable of matching the performance of human experts when creating partisan plans (Section 5.1), and compact plans (Section 5.2). In Section 5.3 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 (for full details see Appendix A.1). Our method is based on simulated annealing, a local-search like method which can make non-improvement steps, allowing it to escape local optima. After some fixed number of iterations or elapsed time, the process ends and the best of all iterated solutions is returned. Starting from an (often random) initial plan, a step is considered by using a modification of the tree-recombination procedure proposed by the Metric Geometry and Gerrymandering Group [28]. 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. Any objective that can be expressed numerically and calculated from an arbitrary plan may be used. Additionally, any binary constraint (for which it can be checked whether a given plan satisfies) can be
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 using absentee data (which we did not have access to).

<table>
<thead>
<tr>
<th>State</th>
<th>Total seats</th>
<th>Our D</th>
<th>538 D</th>
<th>Our R</th>
<th>538 R</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>8</td>
<td>7</td>
<td>5 (8)</td>
<td>4</td>
<td>4 (4)</td>
</tr>
<tr>
<td>MA</td>
<td>9</td>
<td>9</td>
<td>9 (9)</td>
<td>0</td>
<td>0 (0)</td>
</tr>
<tr>
<td>NC</td>
<td>13</td>
<td>7</td>
<td>8 (8)</td>
<td>11</td>
<td>10 (10)</td>
</tr>
<tr>
<td>PA</td>
<td>18</td>
<td>8</td>
<td>8 (9)</td>
<td>13</td>
<td>13 (13)</td>
</tr>
<tr>
<td>WI</td>
<td>8</td>
<td>5</td>
<td>5 (5)</td>
<td>6</td>
<td>6 (6)</td>
</tr>
</tbody>
</table>

Incorporated by ensuring that the algorithm only considers steps which satisfy the given constraint.

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 create. 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 Proof of Concept: 538 Gerrymandering

Nate Silver and the election experts at 538’s gerrymandering project [37] 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 non-trivial 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 (see Appendix A.3 for a detailed description of our reconstruction). Briefly, 538 uses the Cook Partisan Voting Index (CPVI) [9], which measures a district’s $D$ party bias according to the 2012 and 2016 elections, and transforms it into the probability that the $D$ party wins that district. The R party wins it with the remaining probability. 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 (see Appendix A.5.1 for details).

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. For each state and party we ran our algorithm on 60 cores with 2.10 GHz computing power for 24 hours, though our algorithm often stopped advancing well before this deadline. Out of all solutions iterated, we took the one with the most districts above $82\%$ for the indicated party. Our results are shown in Table 2.

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 $D$s, 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.\(^5\) In NC for the $R$ party, we also outperformed 538.

5.2 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. To measure compactness in a single district, we use the following scores:

- **Polsky-Popper (PP)** A district’s PP score is its area divided by the area of the circle with the same length perimeter.
- **Convex Hull (CH)** A district’s CH score is its area divided by the area of its minimum convex hull.

For these two metrics, we use our algorithm to find a plan optimized for the mean score across all districts. From a visual standpoint (see Figure 2), our plans pass an “eye test” for looking compact, especially compared to the plans enacted in practice.

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 used it too to 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. In addition, we have the 2011 plan for all relevant states, and for NC and PA a court mandated updated plan as well.

\(^5\)In the other 4 states there are at most 0.3\% missing ballots.
Table 3: The compactness scores for various plans in Pennsylvania. The best score for each metric is bolded. For PP, CH, and DRA, higher is better, for 538 lower is better.

<table>
<thead>
<tr>
<th>Plan</th>
<th>PP</th>
<th>CH</th>
<th>538</th>
<th>DRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Polsby-Popper</td>
<td>44</td>
<td>86</td>
<td>0.31</td>
<td>93</td>
</tr>
<tr>
<td>Our Convex Hull</td>
<td>37</td>
<td>88</td>
<td>0.29</td>
<td>85</td>
</tr>
<tr>
<td>Our 2% Fair</td>
<td>26</td>
<td>76</td>
<td>0.34</td>
<td>49</td>
</tr>
<tr>
<td>538’s Compact Plan</td>
<td>34</td>
<td>87</td>
<td>0.27</td>
<td>81</td>
</tr>
<tr>
<td>DRA’s Compact Plan</td>
<td>40</td>
<td>82</td>
<td>0.29</td>
<td>70</td>
</tr>
<tr>
<td>2011 Plan</td>
<td>16</td>
<td>62</td>
<td>0.41</td>
<td>15</td>
</tr>
<tr>
<td>Updated Plan</td>
<td>32</td>
<td>78</td>
<td>0.30</td>
<td>64</td>
</tr>
</tbody>
</table>

Comparison to all of the above mentioned plans, in every state our plans had the best mean compactness scores for their respective metrics. They are compact even according to metrics they were not optimized for (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 PA, according to any metric the four compact plans have similar scores. Far more compact than the 2011 and our 2% Fair plan (the later of which will be introduced in Section 6.2). Even the Updated Plan, partially designed to address the 2011 plan’s compactness issues, is generally less compact than the compact ones. We only argue, for a variety of reasons, our algorithm is capable of creating plans just as compact as those from human experts. For the WI, NC, MA, and MD compactness tables information about the 538 and DRA metrics, see Appendix refsubsec:compactnes-tables.

5.3 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. Here, we were finalists from among almost 150 entrants. We competed in two categories. First, designing compact but gerrymandered plans (“stealth gerrymandering”), is the exact topic of Section 7. For the second, partisan fairness, we used our robust partisan bias score (explained in Section 6). Our plans, created in a day, 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 exact quantitative comparison. For more details of the contest and our submissions, see Appendix A.6. See Appendix A.5.5 and Appendix A.5.6 for optimization details.

5.4 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 behavior 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. Here, 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 it is highly modular, it can be used to quickly propose alternative plans, optimizing a diverse set of desiderata. In fact, we were invited to work with the aforementioned group from Princeton to help them and their state partners with the 2021 redistricting cycle. The plans our tool suggests may not be the final ones, but can give a sense of what is and is not achievable.

6 FAIRNESS IN DISTRICTING

![Uniform swing for R(D) in red (blue), in the 2012 presidential election in PA under our Convex Hull compact 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 (\(\frac{1}{2}, \frac{1}{2}\)).](image-url)
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 over time. See Figure 3 for the uniform swing model applied to the 2012 PA presidential election under the plan produced by our algorithm for optimizing the Convex Hull score, where the fraction of districts won by a party is plotted against the fraction of votes received by the party at different uniform swing levels.

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 [18]. This value measures the vertical displacement of the swing curve from the point \((\frac{1}{2}, \frac{1}{2})\). Intuitively, the partisan bias measure 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 \([0.4, 0.6]\) as the reasonable range (as is the case with most presidential elections, the vote shares of both parties are between 40% and 60%). The 45° line in this range is shown in green in Figure 3. 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,

\[
\int_{l}^{r} (s(x) - x) dx \quad \text{average}\]

measuring on average how much higher or lower the swing curve is over the proportional line. A positive (negative) value indicates 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,

\[
\int_{l}^{r} |s(x) - x| dx \quad \text{average}\]

which measures the average deviation from proportionally fair. That is, how much does the plan deviate from “an \(a\) fraction of the vote share should translate into an \(a\) fraction of the seats”. While these measures are similar, and often correlated, they can differ. For example, consider a plan where a vote share of 50% + \(\varepsilon\) (50% − \(\varepsilon\)) results in winning (loosing) 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. 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. 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.

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.

We note that Equation 2 is non-linear, it can not be calculated by combining scores calculated from individual districts. This is unlike the 538 partisan measure, which is just the sum of how much each district leans towards a target party. Even the compactness measures we examine are the simple means (and sums) of individual district compactness scores. Thus this measure is fairly complex to optimize for. Even the algorithm presented recently by Gurnee and Shmoys [21], which was designed with the goal of finding fair plans, would be unable to optimize our second fairness objective.

### 6.1 Compactness Can Improve Fairness

![Figure 4: Average signed distance from the R swing curve to the y = x line over the range [0.4, 0.6] (Equation 1). In each state there are four compact plans, DRA’s, 538’s, our Convex Hull, and our Polsby-Popper. And in most states two implemented ones (2011 and Updated). Each score in PA uses 2012 presidential data, while each plan in WI and NC uses 2016 presidential data. The WI 2011 plan was not struck down so there is no WI updated plan. Appendix A.4 has the scores for other elections and ranges.](image)

As mentioned, compactness is often a primary goal when redistricting. It has even been suggested it is a path to partisan neutrality [22]. A priori, it is not clear if 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. We find that optimizing for various definitions of compactness can reduce the signed partisan bias (Equation 1), relative to the plans used in real life. Despite this, we find a persistent bias to these compact plans, i.e., they favour the \(R\) party despite being optimized for a non-partisan goal.

We find that optimizing for any form of 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]). This improvement is consistent across every state and independent of the compactness measure optimized. Figure 4 shows our signed partisan bias score (closer to zero is fairer) for various plans in three states using the presidential elections of interest.
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 [8], while in PA the plan was found to disenfranchise Ds [27]. The updated plan from 2016 (with an 11% robust partisan bias towards the R party) was still almost twice as biased as any compact plan.

Interestingly, the updated plans from 2016 in NC and 2018 in PA seem dissimilar. While the updated NC plan is still significantly more R-biased than any of the compact ones, the opposite holds for the new PA plan. Its R bias is lower than 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 [7]. That said, it should be noted that neither of the compact plans are designed to optimize the Equation 1. 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 4 because the bars would be virtually invisible.

For all plans, in all states, both elections, and both ranges of comparison there is one consistent pattern: The R party always 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 compactness (supposedly the 2018 PA was also designed to consider proportional fairness). The advantage was significant. On average in PA this was a 10% (and often higher) advantage.

### 6.2 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 rural bias, a "fair" plan. We use our algorithm to optimize for the unsigned partisan bias, Equation 2 over the range [40, 60] (optimizing for Equation 1 can lead to plans with huge jumps in the swing curve). In each state the resulting plans had an unsigned partisan bias score from 3.5 to 10 times lower than any of the the existing and compact plans. Furthermore, in these plans there is near zero advantage at for either party, when measured by the signed partisan bias.

This gain in fairness comes at a cost of compactness. While we can make these fair plans more compact by optimizing for the Polsby-Popper score, with the constraint that the score from Equation 2 should remain near the best found, we are still nowhere near the best compactness scores. As Table 3 shows, in PA, if we keep the unsigned bias under 2% (the best score was 1.5%), the only plan with worse compactness scores is the 2011 plan.

Figure 5 shows this sacrifice in compactness ensures partisan fairness, over the [40%, 60%] vote range. For either party, an α% of the vote means a majority in an α% of the seats. This is unlike other plans (such as our CH plan in Figure 3) where both parties are far from proportional. An exact description of our optimization, and similar results showing a near identical tradeoff in fairness and compactness for WI and PA can be found in Appendix A.4.2.

### 7 DESIGNING PARTISAN PLANS

In this section we examine what limits compactness thresholds impose on partisan gerrymandering. For compactness, we use the average Polsby-Popper score of a plan. To measure gerrymandering ability, we use gerrymandering power (introduced in Borodin et al.
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 stretches $p$’s vote share into a disproportionately large number of districts. A low (or negative) gerrymandering power indicates $p$ is unable to stretch its vote into many extra wins (or even win a proportional number of seats).

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 (at least 50% of the vote) for party $p$. To ensure a diversity of election outcomes we use the elections specified in Section 4. As we saw, 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 6 shows, our compact gerrymander easily passes the eye test, especially when compared to the implemented plan. In NC our algorithm can stretch the number of districts $Rs$ win to all 13 using the 2016 election data, all while creating a plan more compact than the existing one (in the existing plan, $Rs$ won 10 of 13). We can also create a map, more compact than the existing, and where $Ds$ where they win 8 more seats.

We vary $k$ from the most partisan outcome (maximal number of districts won with no compactness constraint), to the most compact outcome (number of seats won by $p$ in the Polsby-Popper compact plan from Section 6).

### 7.1 Effect of Increasing Compactness

![Figure 7: Gerrymandering power in PA when faced with a minimum required Polsby-Popper score using data from the 2012 PA presidential election. $R$ in red; $D$ in blue. The vertical purple (grey) line is the Polsby-Popper score of the 2011 congressional plan (2018 court mandated plan). Average distance between the two curves is a 10.8% $R$ advantage.](image)

As Figure 7 shows in PA, increasing the required mean Polsby-Popper score lowers the gerrymandering power of both parties. But, to have an impact, a steep increase beyond what current plans use is required. WI and NC has the same results (see Appendix A.4.3).

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.

In every single state, even with the most extreme compactness requirements, $Rs$ are able to stretch their vote share beyond proportional. Conversely, $Ds$ have a negative gerrymandering power when compactness requirements are high – no legal plan reaches their proportional allocation.

## 8 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 compact districts, far more compact (according to various metrics) than the plans used in practice or the ones produced by electoral experts. While these plans, which were optimizing compactness, reduce partisan bias, we find they do not eliminate it and still always favor the rural party. Furthermore, 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. Despite this lopsidedness we show there are plans which are near totally proportionally fair, but to achieve this we must sacrifice a significant amount of compactness.

We see many applications of our algorithm for future work, including further examination of the rural party advantage. We have preliminary results on extending the gerrymandering power metric used in Section 7. We find that even if we require districts to have large margins of victory, compactness constraints still more negatively impact the urban party. We also 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 6.2, the next step would be mapping out its Pareto frontier.

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 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


A APPENDIX

A.1 Description of the GREAT Algorithm

In this section we will describe the GREAT algorithm in its general form.

Our basic approach is built around simulated annealing (SA), a general optimization technique that has found much success in various discrete optimization problems. To build our SA approach we will expand on the Markov Chain Monte Carlo (MCMC) package for redistricting known as Gerrychain provided by MGGG. Details of what they provide and our extensions will be expanded on in subsequent sections.

With hill-climbing based approaches, at every iteration we have a candidate, current, solution and we 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, but 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 with 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 and the difference in quality (also known as energy-difference) of the current and proposed solution.

Energy: The first component of a 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 node set with \( n \) nodes and they must be partitioned into \( K \) districts the energy function is:

\[
E : [n]^K \rightarrow \mathbb{R}_+ \cup \infty.
\]

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\(^7\).

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 \( S \):

\[
P : [n]^K \rightarrow [n]^K.
\]

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:

When we say a solution or district is valid we mean that it satisfies all constraints we place on districts. Note that using Kruskal’s

---

\( ^7 \) 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.

\( ^8 \) To draw a random spanning tree we find the minimum spanning tree after randomly assigning each edge a weight.

---

**Algorithm 1** recom_proposal(G, S, j):

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

We note that the recombination method proposed by MGGG only worked for recombinining 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 works well (here $(s)$ is an iteration counter):

$$T(s) = 10000 \cdot (0.99)^s$$  \hspace{1cm} (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.

**Algorithm 2**

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:

1. Let $S_0 = \text{recom\_proposal}(G, \text{None}, K)$.
2. $i = 0$
3. while $i \leq s_{\text{max}}$

4. 

5. 

6. 

7. 

8. 

9. 

10. 

11. 

12. 

13. 

14. 

15. 

16. 

end while

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 valid districting since we are recombining all of the nodes. 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 cool.

It is entirely possible that the procedure will eventually be caught in a local optimum (or even the global one) it cannot move away from with reasonable probability. This is especially true later on as the temperature cools. 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.

### A.2 Compactness Definitions

In this subsection we formally define the compactness measures used in Section 5.2.

For ease of notation we define the following functions for a district $i$ with subgraph $G_i$ and polygon $P_i$. Let the area be $A(P_i)$, let the length of the perimeter be $L(P_i)$, and let the geometric centre point be $M(P_i)$.

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 $A(P_i) / A(C_i)$. 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 $A(P_i) / A(CH_i)$. 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 was not formally explained, it is described as "the average distance between each constituent and his or her district's geographic centre". One possible interpretation of this could be $\sum_{u \in V(G_i)} d(M(P_u), M(P_i)) / n_u$, where $P_u$ refers to the polygon representing node $u$ and $n_u$ is the population in $u$. 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$^{10}$.

---

$^1$This is the point where $P_j$ would balance on a pin tip.

$^{10}$See https://medium.com/dra-2020/compactness-be4e3851126 for more information regarding their metric.
A.3 Description of 538 Reconstruction

Reconstructing the model 538 used to evaluate wins involved multiple steps and analysis.

A.3.1 The Cook PVI. 538 builds their probabilistic model on the Cook partisanship voting index (Cook PVI or PVI) published by the Cook Political Report [9] a non-partisan and independent newsletter that analyzes elections and trends in the United States. The PVI 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 PVI there needs to be a running value for how partisan the country is as a whole (call this value $\beta_D$). To calculate this we take the number of votes garnered in the two most recent presidential elections and comparing to the formula we derived.

In particular they did not make it clear if the mean of the two elections is weighted or not. We confirmed our interpretation by measuring the reported PVI in single district states and comparing to the formula we derived.

To calculate the PVI there needs to be a running value for how partisan the country is as a whole (call this value $\beta_D$). To calculate this we take the number of votes garnered in the two most recent presidential elections and comparing to the formula we derived. Note, this is an average of averages, it is not weighted by the total votes in each election. To calculate the PVI 538 used we need the 2012 election where:

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

For the 2016 election the exact results were:

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

Using the above information we get the value of $\beta_D$ would be:

$$\beta_D = \frac{65,915,795}{65,915,795+60,933,504} + \frac{65,853,514}{65,853,514+62,984,828}$$

(6)

Thus, we see $\beta_D$ is roughly 51.5385759136132%. Note because the United States uses the electoral college system, Donald Trump and Mike Pence won the 2016 election despite receiving fewer votes than Hillary Clinton and Tim Kaine. Note that while the United States is effectively a two party system there are other candidates who run for various offices including president. For example, in 2016, Gary Johnson and Joe Weldon of the Libertarian party received 4,489,341 votes (over 3% of the total vote). 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. 12

The PVI of a district is then just how partisan that district is relative to $\beta_D$. In district $i$, let the total number of Democratic votes denoted by $N_{i,1}^{D}$ and the Republican ones as $N_{i,1}^{R}$ for the last presidential election; and $N_{i,2}^{D}$ and $N_{i,2}^{R}$ the same for the presidential election before that, then the PVI is:

$$\beta_i = \frac{N_{i,1}^{D} + N_{i,2}^{D}}{2} - \beta_D$$

(7)

11The presidential election is chosen since they use the same candidate for the entire country and thus are free of any local effects.

12The Cook Political Report does not actually publish the formula or exact method for this metric. In particular they did not make it clear if the mean of the two elections was weighted or not. We confirmed our interpretation by measuring the reported PVI in single district states and comparing to the formula we derived.

Equation 7 can range from $-\beta_D \cdot 100$ for completely Republican dominated districts, to $(1 - \beta_D) \cdot 100$ 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).

A.3.2 The 538 Sigmoid. The next step in the 538 model is going from the Cook PVI for a hypothetical district to how probable it is that district elects a Democrat. We suspect 538 went with the sigmoid function (the sigmoid function is ideal since it is monotone in its inputs and the output falls in the range (0,1)). Recall, the sigmoid function takes the form:

$$\sigma(x) = \frac{1}{1 + e^{-wx}}$$

(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

Figure 8: 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.
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\textsuperscript{13}. The probability of a Democratic win, plotted against the Cook PVI (first subfigure of Figure 8), clearly shows a sigmoid shape. From here we just need to derive what weight parameter \( w \) they use is. To figure this out we first invert the sigmoid function using the log-odds (or logit) function:

\[
\text{logit}(\sigma(x)) = \log\left(\frac{\sigma(x)}{1 - \sigma(x)}\right)
\]

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 8 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. Firstly, 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 \((0, 50)\)\textsuperscript{14} 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 8. 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 0.3047121945377743 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 (the final subfigure of Figure 8).

\textbf{A.3.3 Tables for compactness.} In this section we provide the tables referenced in Section 5.2. As mentioned we created two plans using our algorithm, one for the Polsby-Popper score and one for the Convex Hull score. In addition we have a plan created by 538 to optimize for their compactness measure. Finally, in each state we have the congressional plan implemented in 2011. And for NC and PA, where the original plans were thrown out, the updated ones.

Unsurprisingly for the plans designed for compactness – our Polsby-Popper, our Convex Hull, and 538’s compactness – they are the best according to their targeted metric. Somewhat suprisingly, it is often the case one of our compact plans is better than the featured DRA plan according to the internal DRA metric. Furthermore, the measures of compactness are fairly correlated. In general a plan designed for one compact measure is very compact, near the best, for the other compactness measures. Interestingly, the implemented WI plan and update PA plan are both slightly more compact than our Polsby-Popper plans according to the 538 metric (the difference is at most marginal). For WI, we believe the odd geography, notably the large portion of lake Michigan that slices into the state, may be impacting purely geography based measures. For PA it is possible whatever compactness measure the updated plan was designed with is more correlated with the 538 metric than the Polsby-Popper metric. It is important to stress, these two Polsby-Popper plans are by no means not compact. According to the Polsby-Popper metric they are far more compact than the other plans, and they are still very compact according to the other metrics.

For getting the DRA metric we took the assignment of nodes to districts for each of our plans, and uploaded it to the DRA website. The website itself does the calculation for their internal compactness metric. Unfortunately our data does not entirely line up with the DRA data. This is not unexpected, our data was hand aggregated and provided in a downloadable format by the Metric Geometry and Gerrymandering Group (https://mggg.org), while the DRA data was put together by the people behind the DRA website. Different naming conventions for precincts, merging or breaking apart of precincts, and how to deal with water are all subjective choices that can cause conflicts. Because of this, we are only able to get the DRA score for plans in two of our states (PA and NC). Interestingly, in these two states, our compact plans beat the featured DRA plans according to the DRA metric (and our PP plan was the most compact according to the metric).

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|l|l|}
\hline
Plan & PP & CH & 538 & DRA \\
\hline
Our Polsby-Popper & 44 & 86 & 0.31 & 93 \\
Our Convex Hull & 37 & 88 & 0.29 & 85 \\
Our 2% Fair & 26 & 76 & 0.34 & 49 \\
538’s Compact Plan & 34 & 87 & 0.27 & 81 \\
DRA’s Compact Plan & 40 & 82 & 0.29 & 70 \\
2011 Plan & 16 & 62 & 0.41 & 15 \\
Updated Plan & 32 & 78 & 0.30 & 64 \\
\hline
\end{tabular}
\caption{Table showing the compactness scores for various plans in Pennsylvania. The best score for each metric is bolded.}
\end{table}

\textsuperscript{13}In total there are 2586 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.

\textsuperscript{14}There is exactly one data point with a PVI of 0 and a Democratic probability of winning of 50%.

\textbf{A.4 Additional Material for Section 6}

\textbf{A.4.1 Additional Plots for Swing Advantage.} Here we provide the plots for other data ranges and elections showing the Republican advantage in the uniform swing model. In Section 6, Figure 4 showed the advantage over the [40, 60] data range for the featured election in each state. Here we provide the complete data set for each election and each range. The 2016 election for both the [40, 60] data range (Figure 12) and [45, 55] data range (Figure 10). And the 2012 election using the [45, 55] data range (Figure 11) and the [40, 60] data range (Figure 12).
Table 5: Table showing the compactness scores for various plans in North Carolina. The best score for each metric is bolded.

<table>
<thead>
<tr>
<th>Plan</th>
<th>PP</th>
<th>CH</th>
<th>538</th>
<th>DRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Polspy-Popper</td>
<td>47</td>
<td>87</td>
<td>39924</td>
<td>97</td>
</tr>
<tr>
<td>Our Convex Hull</td>
<td>38</td>
<td>89</td>
<td>42319</td>
<td>86</td>
</tr>
<tr>
<td>Our 2.5% Fair</td>
<td>37</td>
<td>83</td>
<td>45176</td>
<td>78</td>
</tr>
<tr>
<td>538’s Compact Plan</td>
<td>36</td>
<td>86</td>
<td>37488</td>
<td>83</td>
</tr>
<tr>
<td>DRA’s Compact Plan</td>
<td>40</td>
<td>84</td>
<td>44354</td>
<td>78</td>
</tr>
<tr>
<td>2011 Plan</td>
<td>11</td>
<td>60</td>
<td>55228</td>
<td>10</td>
</tr>
<tr>
<td>Updated Plan</td>
<td>25</td>
<td>71</td>
<td>53160</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 6: Table showing the compactness scores for various plans in Wisconsin. The best score for each metric is bolded.

Because of data incompatibility between the plans we generated and the DRA website we are unable to get the DRA score.

<table>
<thead>
<tr>
<th>Plan</th>
<th>PP</th>
<th>CH</th>
<th>538</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Polspy-Popper</td>
<td>40</td>
<td>85</td>
<td>55991</td>
</tr>
<tr>
<td>Our Convex Hull</td>
<td>35</td>
<td>87</td>
<td>48729</td>
</tr>
<tr>
<td>Our 4% Fair</td>
<td>29</td>
<td>77</td>
<td>56958</td>
</tr>
<tr>
<td>538’s Compact Plan</td>
<td>25</td>
<td>83</td>
<td>47728</td>
</tr>
<tr>
<td>DRA’s Compact Plan</td>
<td>37</td>
<td>81</td>
<td>55025</td>
</tr>
<tr>
<td>2011 Plan</td>
<td>21</td>
<td>74</td>
<td>52623</td>
</tr>
</tbody>
</table>

Table 7: Table showing the compactness scores for various plans in Maryland. The best score for each metric is bolded.

Because of data incompatibility between the plans we generated and the DRA website we are unable to get the DRA score.

<table>
<thead>
<tr>
<th>Plan</th>
<th>PP</th>
<th>CH</th>
<th>538</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Polspy-Popper</td>
<td>27</td>
<td>75</td>
<td>28479</td>
</tr>
<tr>
<td>Our Convex Hull</td>
<td>18</td>
<td>76</td>
<td>29090</td>
</tr>
<tr>
<td>538’s Compact Plan</td>
<td>16</td>
<td>72</td>
<td>26165</td>
</tr>
<tr>
<td>DRA’s Compact Plan</td>
<td>14</td>
<td>72</td>
<td>27330</td>
</tr>
<tr>
<td>2011 Plan</td>
<td>4</td>
<td>4</td>
<td>37459</td>
</tr>
</tbody>
</table>

Table 8: Table showing the compactness scores for various plans in Massachusetts. The best score for each metric is bolded.

Because of data incompatibility between the plans we generated and the DRA website we are unable to get the DRA score.

<table>
<thead>
<tr>
<th>Plan</th>
<th>PP</th>
<th>CH</th>
<th>538</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Polspy-Popper</td>
<td>30</td>
<td>76</td>
<td>0.21</td>
</tr>
<tr>
<td>Our Convex Hull</td>
<td>25</td>
<td>79</td>
<td>0.21</td>
</tr>
<tr>
<td>538’s Compact Plan</td>
<td>20</td>
<td>77</td>
<td>0.19</td>
</tr>
<tr>
<td>DRA’s Compact Plan</td>
<td>19</td>
<td>71</td>
<td>0.21</td>
</tr>
<tr>
<td>2011 Plan</td>
<td>13</td>
<td>59</td>
<td>0.24</td>
</tr>
</tbody>
</table>

As mentioned earlier, in any setting the partisan bias of the compact plans is lower than that of the 2011 plans, with one exception. In Wisconsin in 2016, over the [45, 55] data range the Convex Hull plan has a marginally higher bias than the 2011 plan. Furthermore every single plan we examine shows a Republican bias, none have a democrat lean.

Figure 9: Average distance from the $R$ swing curve to the $y = x$ line over the range $[40, 60]$ for the various plans in each state using 2016 presidential election data. The WI 2011 plan was not struck down, unlike in PA and NC, thus there is no “new” plan for it.

Figure 10: Average distance from the $R$ swing curve to the $y = x$ line over the range $[45, 55]$ for the various plans in each state using 2016 presidential election data. The WI 2011 plan was not struck down, unlike in PA and NC, thus there is no “new” plan for it.

A.4.2 Additional Material for Section 6.2. In this section we discuss how we generated our fair but compact plans, and provide the swing plots for each plan. To find the fair plan we simply use Equation 2 as the energy function. Once the chain stops making improvement steps, we let the best of these iterated plans be our “fair” plan. The swing curves for each of these fair plans are given here (PA in
Figure 11: Average distance from the $R$ swing curve to the $y = x$ line over the range $[45, 55]$ for the various plans in each state using 2012 presidential election data. The WI 2011 plan was not struck down, unlike in PA and NC, thus there is no “new” plan for it.

Figure 12: Average distance from the $R$ swing curve to the $y = x$ line over the range $[40, 60]$ for the various plans in each state using 2016 presidential election data. The WI 2011 plan was not struck down, unlike in PA and NC, thus there is no “new” plan for it.

Figure 13: Uniform swing for $R(D)$ in red (blue), in the 2016 presidential election in WI under our fair 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 $(1/2, 1/2)$. The value of Equation 2 is 3.5938% and the mean Polsby-Popper score is 35.542324.

Figure 14: Uniform swing for $R(D)$ in red (blue), in the 2016 presidential election in NC under our fair 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 $(1/2, 1/2)$. The value of Equation 2 is 2.1006% and the mean Polsby-Popper score is 17.643982.

Figure 15, NC in Figure 14, and WI in Figure 13). While the swing curves here very closely follow the proportional line, the actual plans themselves are highly non-compact. In fact, these plans are less compact than the ones implemented in 2011 which had the lowest compactness score of any plan we looked at. To make these plans more compact we take the “fair” plan as the starting point for our algorithm, optimize for the Polsby-Popper score (as we did in the previous section), but with the added constraint the value for Equation 2 must stay under a fixed threshold. The swing plots for those plans are given here (PA in Figure 18, NC in Figure 17, and WI in Figure 16). We tried other thresholds for how much we allow Equation 2 to deviate, but we feel the ones we present in the main body, and here, present the best balance of compactness and fairness.

A.4.3 Missing Charts for Gerrymandering Power. In Section 7 we explored what happened to the gerrymandering power as stronger
Figure 15: Uniform swing for $R(D)$ in red (blue), in the 2012 presidential election in PA under our fair 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 $[0, 0.6]$. A green star marks the point $(\frac{1}{2}, \frac{1}{2})$. The value of Equation 2 is $1.4879\%$ and the mean Polsby-Popper score is $13.490846$.

Figure 16: Uniform swing for $R(D)$ in red (blue), in the 2016 presidential election in WI under our fair but compact 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 $[0, 0.6]$. A green star marks the point $(\frac{1}{2}, \frac{1}{2})$. The value of Equation 2 is $3.7503\%$ and the mean Polsby-Popper score is $29.420202$.

Figure 17: Uniform swing for $R(D)$ in red (blue), in the 2016 presidential election in NC under our fair but compact 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 $[0, 0.6]$. A green star marks the point $(\frac{1}{2}, \frac{1}{2})$. The value of Equation 2 is $2.3483\%$ and the mean Polsby-Popper score is $37.987309$.

Figure 18: Uniform swing for $R(D)$ in red (blue), in the 2012 presidential election in PA under our fair but compact 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 $[0, 0.6]$. A green star marks the point $(\frac{1}{2}, \frac{1}{2})$. The value of Equation 2 is $1.9865\%$ and the mean Polsby-Popper score is $26.196143$.

Figure 19 shows the effect in PA for the 2012 election, here we provide the same plots for WI and PA (which show a similar effect).
equivalent plot for WI in 2016. Note, since the 2011 plan was never struck down there is only one vertical line. Figure 20 shows the equivalent plot for NC in 2016. As was the case with PA, the 2011 plan in NC was found to be illegal. It was replaced in 2016.

Figure 19: Gerrymandering power when faced with a minimum required Polsby-Popper score using data from the 2016 WI presidential election. $R$ in red; $D$ in blue. The vertical purple line is the Polsby-Popper score of the 2011 congressional plan. Average distance between the two curves is a 4.3% advantage for the $R$s.

Figure 20: Gerrymandering power when faced with a minimum required Polsby-Popper score using data from the 2016 NC presidential election. $R$ in red; $D$ in blue. The vertical purple line is the Polsby-Popper score of the 2011 congressional plan, the vertical grey line is the Polsby-Popper score of the 2016 court mandated plan. Average distance between the two curves is a 10.4% advantage for the $R$s.

A.5 Details of optimization objectives

In this section we will provide the exact details of how we optimized for each of our redistricting goals. In general for each objective we specify, the number of cores we ran our algorithm across, and the exact setting of the energy function, and any constraints. Recall, the energy function is the objective which guides our redistricting algorithm. We specify a function,

$$E : [n]^K 	o \mathbb{R} \cup \infty,$$  \hspace{1cm} (10)

which maps an arbitrary district to a score (where lower is better). This highly modular setup allows us to efficiently optimize for almost any objective.

A.5.1 Optimizing for the 538 Goal. In this section we describe how to use Algorithm 2 to match the 538 partisan plans. For our purposes we will have our energy function be based on the expected number of districts won with one slight modification. Say we are gerrymandering for the Democratic party. If a potential solution $S$ is comprised of $K$ districts called $S_1, \ldots, S_K$ then the energy of that solution is:

$$E(S) = K - \sum_i^K v_D(S_i),$$ \hspace{1cm} (11)

where $v_D(S_i)$ is equal to:

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

Where $\sigma$ 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 $v_R(S_i)$ **15** which is defined as follows:

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

Here, $\tau$ is the threshold for what we consider a strong win. That is if a district win probability for our target party is above $\tau$, we say that is a safe win for that party. 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. For all of our simulations we copy 538 and use $\tau = 0.82$.

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 found with 60 parallel runs we were able to get the results shown in Table 2.

Intuitively this energy function prefers plans with wins just over the 82% threshold and any extra votes in the districts under this threshold, over 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.

A.5.2 Optimizing for compact districting. In this section we describe how to use Algorithm 2 to create various compact plans. These were the plans presented in 5.2. For our purposes we will have our energy function be the final objective, the mean compactness score of our plan (where the mean is the individual score of each districts). Say we are optimizing for compactness measure $M$.

**15**Recall, the probability the Republican party wins a district is just one minus the probability the Democratic party wins it.
If a potential solution $S$ is comprised of $K$ districts called $S_1, \ldots, S_K$, and $M(S_i)$ is the compactness score of district $S_i$ (where the compactness functions are defined in A.2) the energy of that solution is:

$$E(S) = \frac{\sum_i M(S_i)}{K} \quad (12)$$

The only constraints on a proposed districting were a population balance of at most half a percent from ideal, and connectedness. For each metric we optimize for we ran our algorithm for 48 hours across 288 cores.

### A.5.3 Optimizing for compact gerrymandering

In 7 we used our algorithm for stealth gerrymandering. We use our method to generate compact, but partisan, districts. First, for each party, we generate highly partisan outcomes. That is, given a partitioning of the nodes of $G$ into $S = (S_1, \ldots, S_K)$ set Equation 3 as follows (assuming we are gerrymandering for the Democrats):

$$E(S) = K - \sum_i v_D(S_i), \quad (13)$$

where $v_D(S_i)$ is equal to:

$$v_D(S_i) = \begin{cases} \frac{N^D_{S_i}}{N^D_{S_i} + N^R_{S_i}} & N^D_{S_i} \leq \tau \\ \frac{N^R_{S_i}}{N^D_{S_i} + N^R_{S_i}} & \text{otherwise} \end{cases}$$

Here $N^D_{S_i}$ is the total Democrat vote in district $i$ ($N^R_{S_i}$ 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:

$$v_R(S_i) = \begin{cases} \frac{N^R_{S_i}}{N^D_{S_i} + N^R_{S_i}} & N^R_{S_i} \leq \tau \\ \frac{N^D_{S_i}}{N^D_{S_i} + N^R_{S_i}} & \text{otherwise} \end{cases}$$

We set $\tau = 0.5$, that is we require a simple majority of the vote for a win. 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, 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.

For the rest of this section, 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 $W_{max}$. In addition, for each state we ran 288 executions of our code to optimize for the Polsby-Popper score (just like we did in Section 6). Then for $P$ in each state, we have a range of potential win values $\{w_{max}, \ldots, w_{min}\}$ (where $w_{max}$ is the total Democrat votes previous attorney general and governor election. We set the victory threshold $\tau$ at $0.55$, that is we were optimizing for a 10% margin of victory. As we usually do, we started each run of our algorithm from a randomly generated 17 district initial plan. We ran this process on 96 cores for 24 hours. Afterwards we took the lowest energy plan, which secured 15 “wins” for the Democrats, and used it as the initial plan for three separate setups. Then keeping $k$ wins at $\tau = 0.55$ we maximizing the Polsby-Popper score. We tried $k = 11, 12, 13$. In the end we submitted the lowest energy plan
for \( k = 12 \) since we felt it got the best blend of compactness and gerrymandering.

In addition both steps had the usual constraints of connectedness and deviation from the ideal population of at most half a percent.

**A.5.6 Optimizing for Princeton’s Great American Mapoff (Wisconsin).** In this section we describe how to use Algorithm 2 to create plans for Princeton’s gerrymandering competition, specifically the Wisconsin competition. The objective was to create plans which are “fair” from a partisan perspective. Here we used the exact same process as A.5.4. The only difference was we replaced presidential vote totals with the Dave’s Redistricting App composite (as we did in A.5.5). The initial fair outcomes were found by running our algorithm for 24 hours on 96 cores. The lowest energy of these found solutions was then used as the starting point and optimized for the Polsby-Popper score for another 24 hours across 96 cores.

In addition both steps had the usual constraints of connectedness and deviation from the ideal population of at most half a percent.

**A.6 Details Princeton Redistricting**

Here we provide more details of The Great American Mapoff, hosted by Princeton University’s Gerrymandering Project (discussed in Section 5.3). This competition was meant to raise awareness around the upcoming redistricting cycle in the United States and to help recruit members to their Mapping Corps. The Mapping Corps is a group of volunteers who will work with Princeton University and their state partners 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\(^{17}\). 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.

We briefly note that the data used for the rest of this paper was 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 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 was the mean of, the mean of the previous two presidential elections, the mean of the previous two senate elections, and the mean of the previous governor and attorney general elections. As far as we can tell there is no presented evidence that this composite score is a better predictor of future elections than something more simple. We decided to not work with the most recent release (which is still somewhat incomplete) for this paper.

\(^{17}\)See https://gerrymander.princeton.edu/map-contest for details.

**A.6.1 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 achieves this by avoiding several blow-out wins for the Democrats, using these votes to convert tossups and Republican wins to Democrat wins.

Our gerrymander (Figure 21) is far harder to detect than the existing one (Figure 22). It is far more subtle than the existing one and easily passes the “eye test”. Our least compact district (Polsby-Popper 21%) is more compact than all but 4 of the existing districts. Other compactness measures lead to similar results.

**A.6.2 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). Our districts have average population deviation is 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 24), our plan’s swing curves tightly follows the proportional line (Figure 23). In our plan, if either party gets \( x \% \) of the vote, then they should hold a majority in \( x \% \) of the districts (for \( x \in [40, 60] \)).
Congressional maps. Blue districts have at least 55% Democrat vote, red no more than 45% Democrat vote, grey otherwise. All values calculated using the default composite vote total specified by The Princeton Gerrymandering Project.

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.