Mathematically Quantifying Non-responsiveness of the 2021 Georgia Congressional Districting Plan

Zhanzhan Zhao[†]
zhanzhanzhao@gatech.edu
Georgia Institute of Technology
School of Computer Science
USA

Cyrus Hettle[*]
chettle@gatech.edu
Georgia Institute of Technology
School of Mathematics
Atlanta, Georgia, USA

Swati Gupta
swatig@gatech.edu
Georgia Institute of Technology
School of Industrial and Systems Engineering
USA

Jonathan C. Mattingly
jonm@math.duke.edu
Duke University
Department of Mathematics and Department of Statistical Science
USA

Dana Randall
randall@cc.gatech.edu
Georgia Institute of Technology
School of Computer Science
USA

Gregory J. Herschlag[†]
gjh@math.duke.edu
Duke University
Department of Mathematics
USA

ABSTRACT

To audit political district maps for partisan gerrymandering, one may determine a baseline for the expected distribution of partisan outcomes by sampling an ensemble of maps. One approach to sampling is to use redistricting policy as a guide to precisely codify preferences between maps. Such preferences give rise to a probability distribution on the space of redistricting plans, and Metropolis-Hastings methods allow one to sample ensembles of maps from the specified distribution. Although these approaches have nice theoretical properties and have successfully detected gerrymandering in legal settings, sampling from commonly-used policy-driven distributions is often computationally difficult. As of yet, there is no algorithm that can be used off-the-shelf for checking maps under generic redistricting criteria. In this work, we mitigate the computational challenges in a Metropolized-sampling technique through a parallel tempering method combined with ReCom[11] and, for the first time, validate that such techniques are effective on these problems at the scale of statewide precinct graphs for more policy informed measures. We develop these improvements through the first case study of district plans in Georgia. Our analysis projects that any election in Georgia will reliably elect 9 Republicans and 5 Democrats under the enacted plan. This result is largely fixed even as public opinion shifts toward either party and the partisan outcome of the enacted plan does not respond to the will of the people. Only 0.12% of the ∼160K plans in our ensemble were similarly non-responsive.

[*]Second first author.
[†]Corresponding authors.
This work was supported by NSF Award CCF-2106687, CCF-1733812, CMMI-2112533 and DMS-1613337.

This work is licensed under a Creative Commons Attribution International 4.0 License.

EAAMO '22, October 6–9, 2022, Arlington, VA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9477-2/22/10.
https://doi.org/10.1145/3551624.3555300

CCS CONCEPTS

• **Theory of computation** → **Theory of randomized search heuristics**; • **Applied computing** → **Voting / election technologies**.

KEYWORDS

redistricting, polarization, gerrymandering, parallel tempering

ACM Reference Format:
Zhanzhan Zhao, Cyrus Hettle, Swati Gupta, Jonathan C. Mattingly, Dana Randall, and Gregory J. Herschlag. 2022. Mathematically Quantifying Non-responsiveness of the 2021 Georgia Congressional Districting Plan. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), October 6–9, 2022, Arlington, VA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3551624.3555300

1 INTRODUCTION

Gerrymandering is the process of manipulating political districts either to amplify the influence of a political group or suppress the representation of various demographic groups. Over recent redistricting cycles, mathematicians, political scientists, and lawyers have begun to develop effective methodologies to understand and uncover the intent and effects of gerrymandered districts [4, 8, 10, 12, 17, 30]. The basic idea behind these methods is to compare a redistricting plan to a large collection of neutrally drawn alternative plans that comply with preferences motivated by legal and policy considerations. In this work, our main goal is to analyze gerrymandering in the 2021 Georgia congressional districts. We also endeavour to refine and codify the intellectual framework one ideally would use in such analyses and analyze the plans with the Georgia congressional redistricting criteria.

A key approach in detecting gerrymandering that has been upheld in various state courts [15, 16, 25, 34] is as follows: one uses non-partisan criteria reflected by a plan of interest, M, to draw a collection of plans, referred to as an *ensemble*, from a specified distribution, and then checks whether M is an outlier with respect to partisan properties, such as election outcomes. Creating a collection of plans reflecting the non-partisan properties of M is necessary

to be able to directly compare the ensemble to the enacted plan; for example, if the ensemble has districts that are significantly less compact than the enacted plan, one could not assess if differences in partisan behavior came from explicit manipulation or were necessitated by the increased compactness. Therefore, one may wish to "tune" (i.e., control) the average compactness[1] of the districts in the ensemble of plans to align with the compactness of M.

In generating ensembles, Metropolized-sampling approaches transparently allow one to codify preferences between plans in a policy-driven framework, providing a flexible and explicit distribution on the space of plans. However, sampling from many commonly-used policy-driven distributions is computationally difficult and, as of yet, there is no algorithm that can be used off-the-shelf for plans in generic political environments. This is due to barriers in some sampling procedures that can cause extremely slow convergence rates, as well as using proposal chains or generative methods that are close to singular with respect to the desired measure.

Although currently there is no federal prohibition on partisan gerrymandering, the state court cases mentioned above have overturned enacted plans based on these methods. Furthermore, these methods provide a robust and solid footing in how to think about and understand extreme partisanship in redistricting which can serve to inform public debate and policy surrounding redistricting practices.

The key contribution of our work is to mitigate the challenges with sampling from policy-driven distributions via a parallel tempering scheme; parallel tempering is a standard tool to bridge target measures with measures that are easier to sample from, however it does not always work in practice. It has been used previously in redistricting on diffusive boundary methods [12, 13], however these methods have not been demonstrated to scale to the size of statewide graphs. Here, we employ parallel tempering with recombination; to our knowledge, this is the first time these two methods have been combined and we demonstrate their efficacy on statewide graphs for particular measures. We implement our algorithm via a case study of detection of gerrymandering in the 2021 enacted congressional redistricting plan in Georgia. We summarize our main findings below:

(1) **Accelerated Sampling:** In sampling redistricting plans on Georgia, we tune a distribution to sample from an ensemble of plans that have a comparable level of compactness as the 2021 enacted plan. We bridge algorithmically accessible measures to policy-informed measures via parallel tempering using Metropolized ReCom chains. Although theoretically sampling from this distribution is feasible, in practice, using single chains we find that these show no signs of mixing after days of run time. Using the resulting ensemble, we compare the partisan properties of the ensemble with those of the 2021 enacted plan as explained next.

(2) **Non-responsiveness:** We find that the 2021 plan will likely be highly *non-responsive* to changing opinions of the electorate. The enacted plan is structured so that it will reliably elect nine Republicans and five Democrats over a wide range of studied voting patterns, with statewide Democratic vote fraction percentages ranging from the mid 40s to the low 50s, as has been typical in recent Georgia elections (see Figure 2). In contrast, over this same range, the non-partisan plans in our ensemble do react to the changing voter preferences by shifting the partisan make-up of those elected. Only 0.12% (186 out of 159997) of the plans in the ensemble exhibit the same extreme non-responsiveness as the enacted plan by producing a single election outcome over the seventeen elections considered (between 2016 and 2020). Moreover, when considering the effects of modifying statewide vote fractions to start strongly favoring either party, the number of officials that would be elected for that party under the enacted plan tend to be systematically smaller than what is projected under plans from our ensemble (See Figure 3 and Figures 8 and 9 in Supplementary Information (SI)).

(3) **Polarization in Competitive Districts:** We find the major cause of non-responsiveness in the enacted 2021 plan is *polarization of* voters across the more competitive districts. Specifically, there are five districts that would be significantly more competitive under plans with only our non-partisan considerations, but have been shifted to become more Republican. Similarly, there are three districts that have more Democratic voters than is typical. The effect is that districts that could be more responsive to the changing opinion of the voters have become more solidified in their partisan lean and consequently the enacted plan is non-responsive to shifts in voting patterns.

(4) **Packing in Democratic Strongholds and Cracking in a Republican Stronghold:** Utilizing a combination of our ensemble statistics and spatial analysis, we find evidence of where the most impactful changes are. We show that the heavily Democratic 4th, 5th and 13th Congressional Districts and the heavily Republican 9th District contain a significantly larger number of Democrats than typical plans in the ensemble. The consolidation or *packing* of Democratic voters in the 4th, 5th, and 13th Districts creates significant numbers of wasted votes and dilutes their voting power. Furthermore, based on comparisons with both the previously enacted 2011 plan and the ensemble of plans, we find that a significant number of Democratic voters have been safely added to the solidly Republican 9th District, replacing Republican voters who have been moved to the 6th and 10th Districts. This redrawing substantially weakens, or *cracks*, the potential influence of Democratic votes in these districts.

We summarize how the above contributions fit into a more general framework of detecting gerrymandering in a given plan, M, in Figure 1. This process entails three major steps: (i) designing the distribution, or family of distributions, that is used to quantify the non-partisan redistricting criteria (Section 2); (ii) randomly sampling from the space of compliant redistricting plans according to

[1]In the current literature, there is an ongoing debate over which measure of compactness is appropriate in redistricting; the debate includes using spanning trees, "cut" edges [11], ML/human learning approaches [23], and more traditional measures such as Reock and Polsby-Popper. Despite known issues with the traditional measures [3], they are still most often used in the practice of redistricting. This is partially because nearly all measures have known issues. In our case study, we will use the Polsby-Popper measure of compactness.

our preferences/distribution to generate a large, non-partisan collection (or ensemble; Section 2.1); and (iii) comparing the collection of plans to a particular plan of interest, M (Sections 3 and 4).

The first two steps are performed iteratively, as distribution parameters are tuned so that the non-partisan criteria in the ensemble are comparable to those of a plan of interest. When sampling from a measure, we draw samples until convergence can be established. Once a reasonable ensemble is sampled (and its convergence verified), we move to the analysis; when investigating partisan behavior of plans, we analyze both the ensemble plans and the enacted plan with past voting data. For example, we examine the distribution of elected Democrats under the ensemble coupled with a particular set of voting data; we compare this distribution with the result for the enacted plan. Of course, in this entire pipeline, decisions must be made such as defining compactness as well as the adjacency structure and regions of the plan that may be assigned to different districts. However, the process described above is largely what has been employed in states including North Carolina [18, 30], Ohio [33], Pennsylvania [31, 34], Virginia [10], Maryland [24], and Wisconsin [20]. Importantly, we note that this framework does not make reference to any concept of proportionality in redistricting and elucidates how the spatial distribution of the state's electorate interacts with the redistricting process.

2 GENERATING THE ENSEMBLE

In this work, we focus on the Metropolized sampling methods that are capable, in theory, of sampling from any specified distribution on the space of redistricting plans. Specifically, these include tree based approaches [1, 2]. These methods also include diffusive boundary approaches, however they often run into mixing problems due to energetic barriers [12, 19, 30]. Yet other approaches rely on Metropolized methods that do not require mixing but rather compare local chain properties with a plan of interest [7, 8]. Metropolization is not the only way to sample from a known distribution; recently a sequential Monte Carlo method was introduced that is also capable of sampling from generic measures, although it faces the same practical mixing challenges as the tree based methods [31].

Algorithm-dependent methods are another popular choice for generating redistricting ensembles. These approaches tend to be more computationally efficient. The choice of the algorithms used to sample the plans dictates the properties of generated plans (e.g., [5, 6, 9, 10, 26]). Despite being highly computationally efficient, these methods may introduce unintended and hard-to-identify biases into the collection of plans. Nonetheless, these approaches are becoming popular in practice due to the ease and efficiency in implementation.

In each state, the methods are adapted to account for the state's specific redistricting requirements. In each of the above listed works, the measure being sampled from was either ignored or in some way altered for the sake of computational efficiency. To our knowledge, there have been only two state-wide investigations that have sampled from a measure constructed solely around existing policy [27–29]. These works sample the state legislative districts in North Carolina by breaking the state into small sub-regions; each of the sub-regions is more manageable to sample from using standard Metropolis-Hasting based sampling methods.

One promising avenue to sample from a desired meassure is to employ parallel tempering, which samples from a range of distributions that interpolate between a measure that is efficiently sampled and a target measure. Parallel tempering has been used previously in redistricting [12, 13]. In these works, parallel tempering was coupled with diffusive boundary Markov chains which worked well for smaller graphs, but was not shown to effectively scale to larger graphs such as statewide precinct graphs. In this work, in contrast, we use recombination techniques; these techniques can make significant changes to a redistricting plan in a single step of the Markov chain. The challenge in using them is that it is natural to use them to focus on spanning forest measures rather than partition measures (see below and SM section B). Our major contribution is to show how parallel tempering coupled with recombination methods can effectively refocus a measure around relevant redistricting criteria in an enacted map.

We next discuss the steps that we take to generate our ensemble of plans[2], starting with defining a graph, a family of probability distributions, and our accelerated sampling method.

Defining the graph. We define a congressional redistricting plan in Georgia as a balanced graph partition of 14 elements on a graph in which nodes (roughly) represent Georgia's precincts and edges represent precincts with shared geographic boundaries. In general, redistricting processes will preserve voting precincts. However, precincts may be comprised of discontiguous regions and ensuring that these regions all belong to the same district may require a large number of neighboring precincts (and a large number of voters) to all be confined to the same districts. In these cases, practical considerations may lead mapmakers to assign the distinct regions making up a precinct to different districts. We discuss the exceptions and modifications to this rule in Section 1 of SI. We work with a planar graph, which ensures there are no discontiguous districts, and resolve some of the contiguity or density issues on a case-by-case basis. Such issues are quite infrequent, and therefore, these choices have negligible impact (if any) on the analysis.

Defining a family of probability distributions. It is important to note that we are not just generating a collection of "random redistricting plans," but rather sampling a *distribution* on redistricting plans which encapsulates the laws and preferences for a redistricting. This makes the preferences between plans explicit, so that they can be discussed and critiqued. Additionally, one is free to use different algorithms to sample a fixed distribution. If we only describe the distribution implicitly through an algorithm, we risk introducing unforeseen biases.

The redistricting plans in our ensemble satisfy the following:

- *Contiguity:* All districts consist of one contiguous region.
- *Population balance:* The total population in each district is within 1% of the ideal district population. In [20], the authors verify that the small changes needed to make districting plans have perfectly balanced populations do not have significant impact on the partisan results of our ensembles.

[2]Our code and data can be accessed at https://git.math.duke.edu/gitlab/quantifyinggerrymandering/2020-analysis/ga-ensemble-analysis

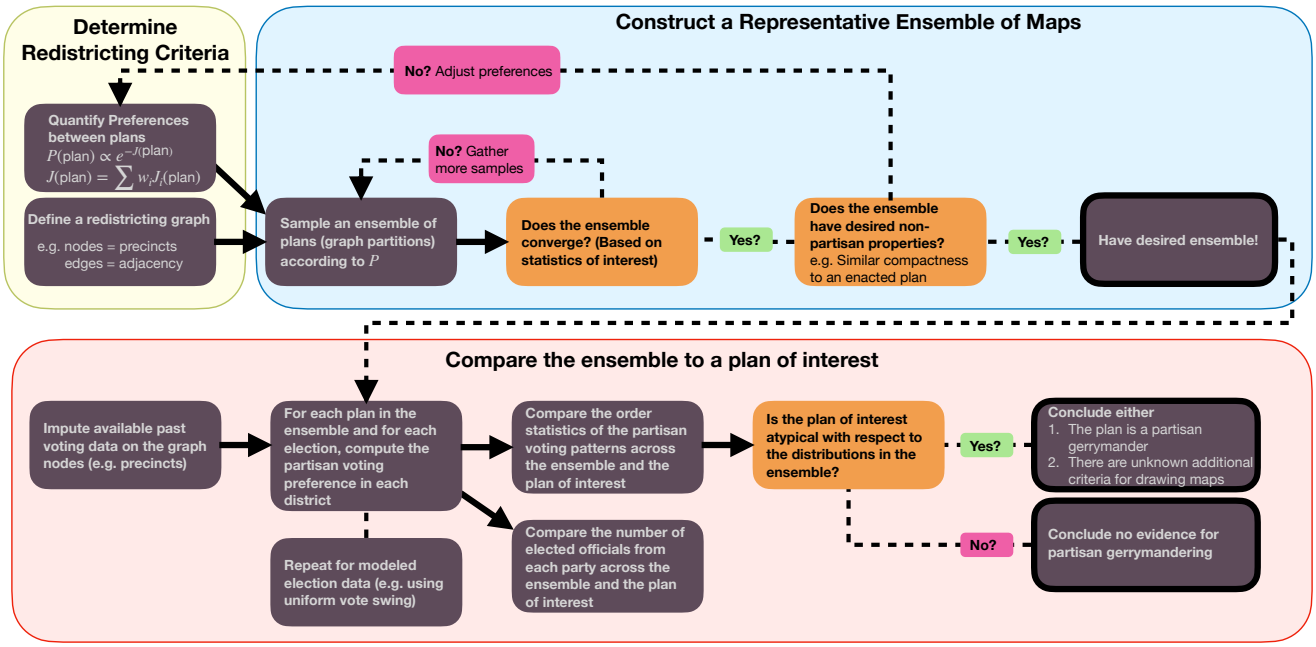

Figure 1: Outline of our procedure for detecting partisan gerrymandering.

- *Maximum splits:* The plan splits at most 21 counties. The number of 21 is chosen in accordance with the number of county splits in the 2021 plan.[3]
- *Traversing boundaries:* Districts traverse each county boundary at most once; when a district splits a county, it may not form two discontinuous regions when restricted to that county.
- *Compactness:* The distribution is concentrated on more compact plans; this reflects the General Assembly's guidelines that the plans should be compact. We have tuned the distribution so that it yields plans of a similar compactness to those of the 2021 plan [14]. We measure compactness with the Polsby-Popper score, which is a commonly used measure of assessing compactness. See Section 5 of SI for formal definitions of the compactness and a comparison with the enacted plan (Figure 7 in SI).

Our chosen probability distribution over redistricting plans, from which we draw plans in our ensemble, prioritizes the desired policies and complies with legal considerations. To mathematically account for compactness of plans, we use a target measure that includes the compactness score. We target these scores so that the distribution of compactness scores in our ensemble is close to that of the enacted plan from 2021. See Section 3 of SI for more details.

2.1 Accelerated sampling from the distribution

In order to obtain random plans from our chosen probability distribution, we run a Markov chain with Metropolized transition probabilities with parallel tempering. In this work, we use a Metropolized tree-based sampling method. Tree-based methods have shown promise to mix[4] when used to define a Markov chain that merges adjacent districts, draws a spanning tree on the merged space, and cuts the tree into two subtrees that each represent a district [11].

The work of DeFord et. al [11] has been employed as a proposal kernel and Metropolized by Autry et. al [1] and has also inspired a sequential Monte Carlo method [31]. Modified Metropolized methods have been shown to efficiently sample on the uniform measure of balanced spanning forests (i.e., each tree in the forest represents a district or partition with roughly equal population; for more details see [2] and Section 2 of SI). Theoretically, these methods can sample from any measure, but in practice, the chains may not mix in reasonable time. For example, in our case study for Georgia, the measure of uniform spanning forests lead to plans that were significantly less compact than the enacted plan; however, sampling from a measure with greater affinity for compactness was unable to mix even after several days using the same method on a single chain.

Parallel tempering is a class of Markov chain Monte Carlo algorithms that constructs a path of distributions interpolating between a tractable reference distribution which mixes quickly and the target distribution which we want to sample from but unable to mix. By swapping states along the path with Metropolis probability, the mixing of the target distribution may be improved [37]. Autry et. al [1] proposed parallel tempering as a potential mechanism to successfully access new measures, however it was never directly implemented or tested in this context. In this work, we implement such a method, adopting the multi-scale approach presented in [2] in order to efficiently preserve counties. We begin by gathering plans sampled from the uniform distribution of spanning forests

[3] One can expect the properties of the ensemble to change if the number of allowed county splits are increased. Setting the number of splits equal to he proposed plan allows us to mimic the non-partisan properties of the plan in question.

[4] The time needed for the observables of interest to be sufficiently close to the goal distribution [22].

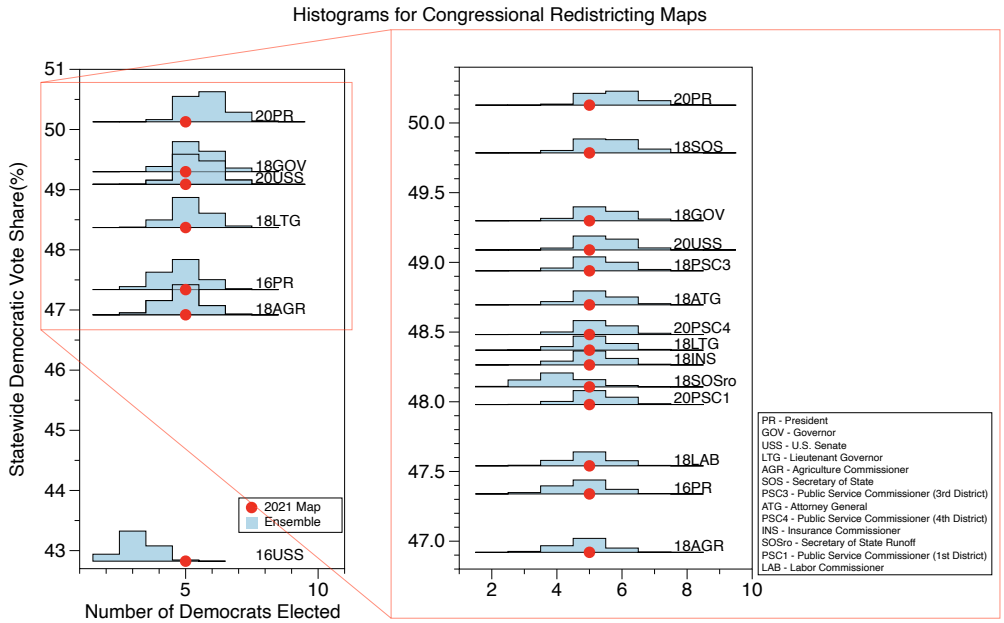

Figure 2: Each histogram represents the range and distribution of possible Democratic seats won in the ensemble of plans; the height is the relative probability of observing the result. We include only a selection of the historic vote counts for clarity so that we have a representative election for vote fractions that are spaced out. On the left axis, we provide selected Democratic statewide vote percentages. The red dots represent the reslts from the enacted 2021 plan under the various historic votes.

on the hierarchical structure; this measure has been shown to be efficiently sampled by these methods and we find the same in the case of Georgia. We launch four such independent chains with random initial conditions and run each chain for 10 million proposals, saving the state every 25 proposals. We find strong agreement between the violin plots (see Figure 3 in Section 4 of SI) which means that the observable agrees in distribution and become independent of the chain's starting point. We later use the samples we get at this measure as independent and identically distributed (i.i.d.) samples for tempering runs at base measure of $\gamma = 0$, so that instantaneous mixing can be obtained by swapping with the base measure.

Formally, the reference distribution that is uniform on hierarchical forests, which is easy to sample from is given as

$$P_F(T) \propto \mathbb{1}_C(T),$$

$$P_F(M) \propto \mathbb{1}_C(M) \sum_{T_1 \in T(M_1)} \cdots \sum_{T_{14} \in T(M_{14})} P_F(T) = \mathbb{1}_C(M) \prod_{i=1}^{14} \tau(M_i), \tag{1}$$

where $T(M_i)$ is the collection of spanning trees associated with a district M_i, T_i is a particular spanning tree on district M_i, $T = (T_1, \ldots, T_{14})$ is a spanning forest with trees on each of the districts in M, $\tau(M_i)$ is the number of hierarchical spanning trees associated with district M_i, and M is a plan consisting of 14 districts. Finally, $\mathbb{1}_C$ is the indicator function that is 1 when M is in the set of constraints listed in Section 2. Ideally, the target distribution for sampling would yield

$$P(M) \propto \mathbb{1}_C(M) \exp(-w \cdot J(M)), \tag{2}$$

where J is a score function measuring the level of compactness in M, and w is the weight we can tune to match the compactness of our ensemble with the enacted plan. However, this target distribution is currently infeasible. As shown in [1], one can interpolate between the tractable reference distribution and the intractable target distribution via a sequence of distributions parameterized by γ, given by

$$P_\gamma(M) \propto \mathbb{1}_C(M) \prod_{i=1}^{14} \tau(M_i)^{1-\gamma} \exp(-\gamma \cdot w \cdot J(M)), \tag{3}$$

where $\gamma \in [0, 1]$. When $\gamma = 0$, we sample from the uniform measure on hierarchical forests P_F in (1), and when $\gamma = 1$ we sample from the target measure P in (2).

We begin by naively implementing a parallel tempering scheme [36] on 32 cores, setting the sequence of $\gamma_i = (i-1)/31$, where $i = 1, 2, \ldots, 32$. Furthermore, we use the previously sampled and converged ensemble at $\gamma = 0$ as an i.i.d. sampler. This confers the advantage that our chains instantaneously mix when they exchange with the $\gamma = 0$ measure, and we later referred to it as the heat bath method. Unfortunately, we do not observe mixing in this case. To estimate the number of needed cores, we can draw random pairs from our samples at each level of γ and determine the spacing of the next γ that would be needed to ensure at least a certain percent swap probability. In a similar example, we have found that the required spacing is nearly the same for each level of γ, and that we would need somewhere between 1000 and 10000 cores to effectively implement a parallel tempering scheme, which is infeasible given our resource limits. Hence we instead make a concession to draw

distributions from

$$P'_{\gamma}(M) \propto \mathbb{1}_C(M) \prod_{i=1}^{14} \tau(M_i) \exp(-\gamma \cdot w \cdot J(M)), \tag{4}$$

which favors plans with higher spanning tree counts but also has bigger probability weights for plans with our desired compactness weight w. In this way, our target distribution to sample from now yields

$$P'(M) \propto \mathbb{1}_C(M) \prod_{i=1}^{14} \tau(M_i) \exp(-w \cdot J(M)). \tag{5}$$

Sampling from this new measure is computationally feasible given our available resources, and relaxes the tension between sampling from a desired distribution and the computational feasibility of sampling. With this setup, we find strong evidence of mixing using only 10 levels of γ (i.e., 11 cores with $\gamma_i = (i-1)/10$, where $i = 1, \ldots, 10, 11$ and the i.i.d. samples drawn at heat bath of $\gamma = 0$). See Section 4 of SI for a convergence study of this method.

Although our ensemble still favors plans with higher spanning tree counts, the above scheme provides a significant advance to previous implementations in the literature. First, it focuses on a tunable measure that cannot be readily sampled with published methods. The tempering scheme, coupled to the heat bath method at the lowest γ, provides an efficient algorithm that can converge in an achievable amount of wall clock time with feasible computational resources.

3 NON-RESPONSIVENESS OF THE 2021 PLAN

We see that the enacted Georgia plan does not produce different partisan results, in terms of the representative elected, over the range of recent voting patterns. In contrast, plans in the ensemble typically respond to the changing popular vote over these same voting patterns. Furthermore, when the Democratic vote share grows, the enacted plan systematically under-elects Democrats. There is a smaller range of election environments where the plan underelects Republicans.

Note that a plan that reacts by changing representation when the number of votes for a particular party changes sufficiently is a minimal requirement of a democratic process responsive to the changing will of the people. To investigate this, we analyze the enacted Georgia plan by evaluating how many Democrats it would elect under a number of statewide voting patterns and compare this with the ensemble of plans. We demonstrate this through a type of plot we call *collected seat histograms*. The election data we use is either a set of historical elections (Section 3.1) or data generated by applying a uniform swing to a particular historical election[5] (Section 3.2). Both kinds of collected seat histograms are effective at identifying plans that are non-responsive or under-respond to changing voter opinions.

[5]In accordance with the uniform swing hypothesis, we take a single election and then uniformly increase or decrease the vote percentage for a given party across all the districts. This creates a new set of voting data with the same spatial structure but a different statewide partisan percentage for each party.

3.1 Collected seat histograms under historic elections

In this section, we plot a *collected seat histogram (CSH)* for historic elections data , i.e., the number of Democrats elected by the 2021 plan, and compare this against the distribution of Democrats elected across all the plans in our ensemble for historic elections, which provide different voting patterns (see Figure 2). These CSH plots illustrate the level of responsiveness to changes in the votes one should expect of plans drawn without a partisan bias[6].

We see that the enacted plan elects 5 Democrats and 9 Republicans under all 17 of the historic elections we have examined. We check whether this lack of response to voting patterns is commonly observed in the ensemble of plans. When looking at the statewide Democratic vote share, 16 of the 17 elections are clustered around a statewide Democratic vote share of 46.5%-50.5%. The one exception to this is the 2016 US Senate race which has an atypically high Republican vote share in which less than 43% of the vote went to the Democratic candidate. When looking at all but the the 2016 US Senate votes, only 2.1% (3428 of the 159997) of the plans in the ensemble elect the same number of officials from each party under all 16 historic voting patterns. If we include the US Senate 2016 votes, only 0.12% (186 out of 159997) of the plans in the ensemble elect a fixed number of Democrats under all 17 historic voting patterns. In short, the plans in the ensemble are nearly always more responsive than the enacted plan.

It may be temping to look at Figure 2 and conclude that the enacted plan is fairly typical of the ensemble, as it is in the center of the histograms over the majority of the elections. However, the plans in the ensemble elect 4-6 Democrats under the 2018 Commissioner of Agriculture election (at a statewide Democratic vote fraction of just under 47%) and 5-7 Democrats under 2020 Presidential election (at a statewide Democratic vote fraction of just over 50.2%). This shift reflects that under a typical plan it is normal for the composition of the delegation to change as the vote profile does. It is highly unusual for the composition to not change over the range of elections, as seen in the enacted plan.

In particular, the enacted congressional plan is stuck electing five Democrats in the fourteen districts, despite shifts in the statewide vote fraction and the distribution of votes across the state. Over these elections, with a statewide vote Democratic vote ranging from 42.8% to 50.1%, the number of Republicans and the number of Democrats elected does not change at all. This shows the enacted plan to be highly non-responsive to the changing opinion of the electorate, and without holding the election, one largely knows that 9 Republicans and 5 Democrats will be elected.

3.2 Uniform swing analysis

In addition to using historical statewide votes to produce our collected seat histograms, we create a set of collected seat histograms built from a single historical vote which is shifted in accordance with the "uniform swing hypothesis" to produce a new collection

[6]Throughout, we normalize election results by using the fraction

$$\text{Democratic vote share} = \frac{\text{number of Democratic votes}}{\text{number of Democratic votes} + \text{number of Republican votes}},$$

since the fraction of third party votes varies for each election. To determine the number of seats won by each party, we compare this fraction to 0.5.

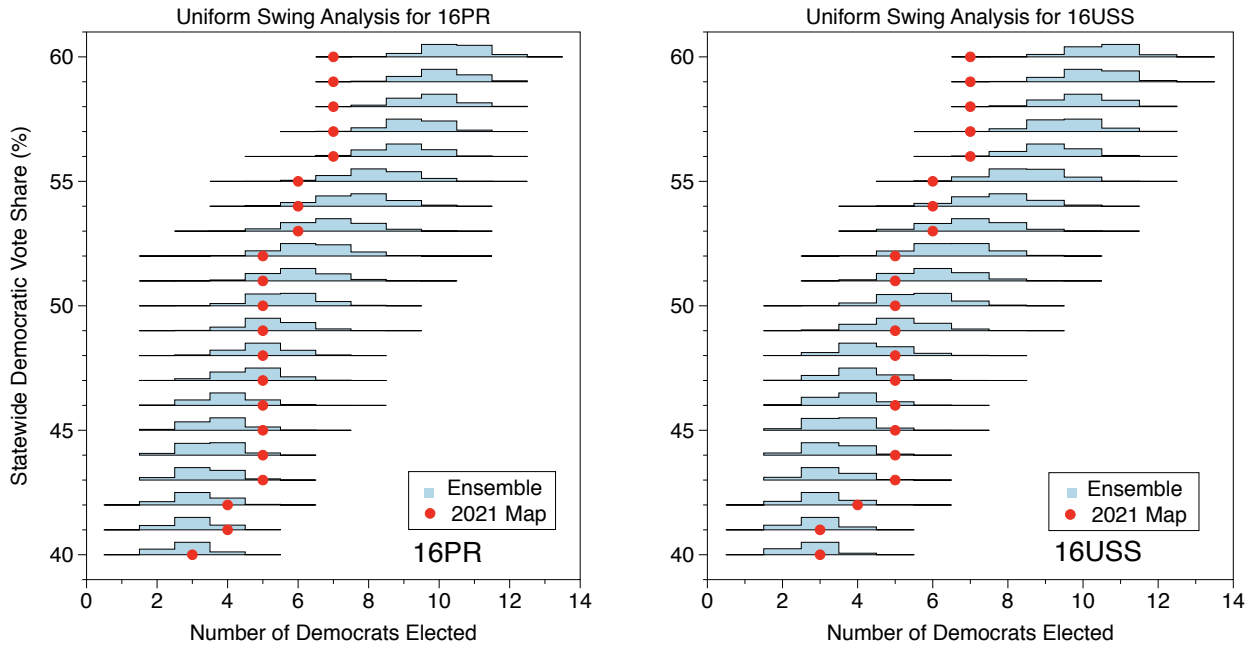

Figure 3: The individual histograms give the frequency of the Democratic seat count in the ensemble for each of the shown statewide elections, with a uniform swing. The histograms are organized vertically based on the swung statewide partisan vote fraction. The red dots denote the Democratic seat count for the enacted plan for each of the swung vote profiles.

of votes [21]. This preserves the relative voting pattern across the state while allowing us to study the effect of shifting the partisan tilt of the election.

In Figure 3 for 2016 elections (and Figures 8 and 9 of SI, for 2018 and 2020 elections respectively), we see that the non-responsiveness phenomenon from Figure 2 is repeated much more severely. In many cases the enacted plan fails to respond to the shifting will of the electorate, or is significantly less responsive than the plans produced by the ensemble. Our analysis shows that even if the Democratic vote share were to increase to greater than 54%, the enacted plan under-elects Democrats. For example, when the statewide Democratic vote fraction for the 2016 presidential election (16PR) and 2016 U.S. senate election (16USS) elections is swung to between 54% and 60%, the ensemble elects the same or fewer Democrats in 0.28% and 0.15% of the plans respectively. Though Georgia has not historically seen such large Democratic swings, this plan may serve to solidify these districts against future demographic changes. Furthermore, there are regions in which the enacted plan elects more Democrats then expected; for example, when the 2016 Presidential votes are swung to a Democratic statewide vote share between 43%-45% (see left Figure 3). This non-responsiveness is due to polarized districts that abnormally separated Democratic and Republican voters; we demonstrate this in the next section. See Section 6 of SI for additional plots for 2018 and 2020 statewide elections. Moreover, it underlines the increased polarization we see in many districts.

4 POLARIZATION IN COMPETITIVE DISTRICTS

In addition to looking at the number of elected representatives from each party, we examine the margins of victory within races between 2016-2020. To this end, we examine box plots that show the rank-ordered marginal distributions of the partisan vote fraction across the plans. These plots help identify when the plan contains districts with abnormally many Democrats or Republicans. This is done by considering the partisan vote fraction for one of the political parties (Democrats) in each of the districts for a given redistricting plan. These marginal vote fractions are then ordered from smallest to largest, i.e., from the most Republican district to the most Democratic district. These ordered fractions are then tabulated over all of the plans in the ensemble and used to form order statistics over the ensemble (see Figure 4 and 5 for 2020 and 2016 elections). Qualitatively similar results are seen for 2018 elections (Section 7 of SI).

The rank-ordered marginal box plots show the typical range of the most Republican district to the most Democratic district. Ranges are represented by box plots. In these box plots, 50% of all plans have corresponding ranked districts that lie within the box; the median is given by the line within the box; the ticks mark the 2.5%, 10%, 90% and 97.5% quartiles; and the extent of the lines outside of the boxes represent the range of results observed in the ensemble. Any box that lies above the 50% line on the vertical axis corresponds to a (ranked) district that will typically elect a Democrat; any box that lies below the 50% line corresponds to a (ranked) district that will typically elect a Republican (e.g., in Figure 4 (top), districts ranked 11-14 reliably elect Democrats, and districts ranked 1-6 reliably elect Republicans).

We evaluate the enacted plan with each set of votes and plot the ordered district results over the box plots. If results of particular districts lie either far above or far below the ensemble at the same

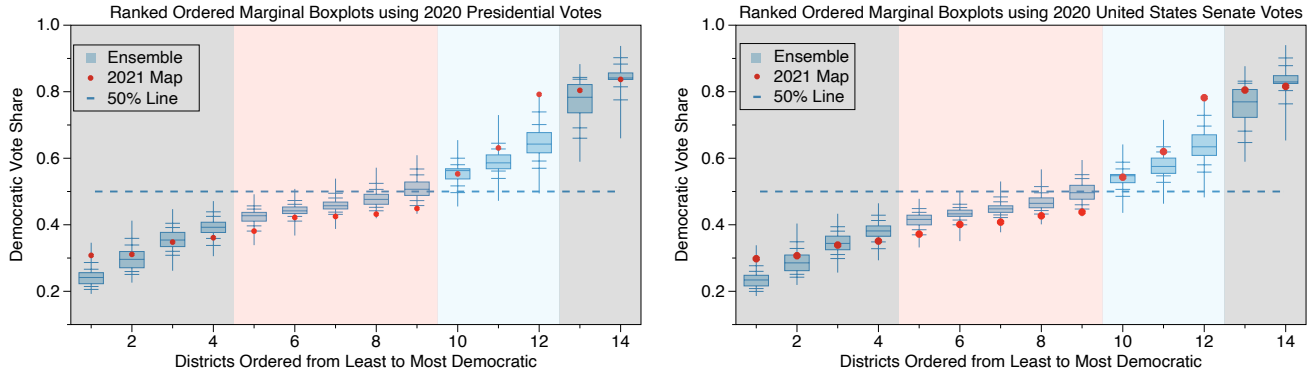

Figure 4: The red dots display the ordered Democratic vote share for each of the 14 districts in the enacted 2021 plan based on 2020 presidential (top) and 2020 United States Senate (bottom) statewide elections. The box plots display the range of Democratic vote share observed for plans in the ensemble, with the 14 vote shares for each plan ranked from least to most Democratic.

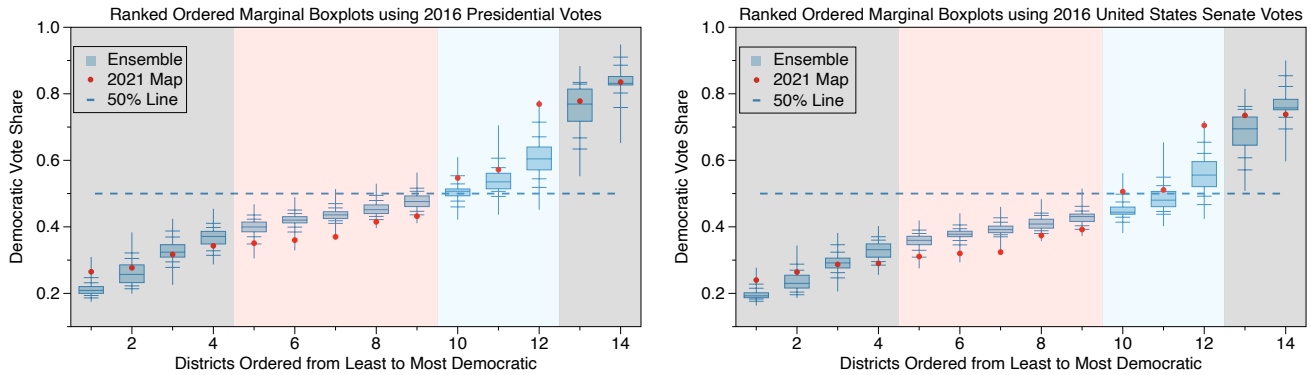

Figure 5: The red dots display the ordered Democratic vote share for each of the 14 districts in the enacted 2021 plan based on various 2016 statewide elections. The box plots display the range of Democratic vote share observed for plans in the ensemble, with the 14 vote shares for each plan ranked from least to most Democratic.

ranking, this can indicate that the district was drawn to increase or decrease one party's representation within it.

In Figures 4-5, we examine a variety of elections from Figure 2 across 2016 to 2020 and consistently find that:

- the 5th-9th most Republican districts of the enacted 2021 plan have significantly fewer Democratic votes than the corresponding 5th-9th most Republican districts of plans in the ensemble (e.g., see the pink highlighted regions),
- the 10th-12th most Republican districts of the enacted 2021 plan have significantly more Democratic votes than the corresponding 10th-12th most Republican districts of plans in the ensemble (e.g., see the blue regions).

We consider the total Democratic votes in the 5th-9th most Republican districts from each plan in the ensemble and compare them to the total Democratic votes in the 5th-9th most Republican districts from the enacted plan. Of the 159,997 plans in our ensemble, across 17 elections from 2016 to 2020, no more than 0.017% of the plans (27 out of 159,997) have the same or fewer Democratic votes than the enacted plan, which suggests that the 2021 plan polarizes voters across the 5th-9th most Republican districts to make those districts more Republican.

On the other hand, when we consider the total Democratic votes in the 10th-12th most Republican districts from each plan in the ensemble and compare them to the sum of the Democratic votes in the 10th-12th most Republican districts from the enacted plan, we observe that, across all elections, no more than 0.2% of the plans (320 out of 159,997) would have the same or more Democratic votes than the enacted plan. This suggests that the 2021 plan polarizes voters across the 10th-12th most Republican districts to make those districts more Democratic. These statistics are summarized in Table 1 of SI. Consequently, districts that could have been more responsive to voters have been solidified in their partisan lean, and the enacted plan is stable and non-responsive to voting pattern shifts.

5 A SPATIAL ANALYSIS OF PACKING AND CRACKING IN THE 2021 PLAN

In general, the ranked ordered marginal distributions do not correspond to geographic regions. However, in Georgia, the most Republican district and the three most Democratic districts in the ensemble share geographical consistencies across most plans in the ensemble and the enacted plan. We outline the enacted districts along with a heat plan capturing typical locations of the corresponding districts in the ranked-marginals in the ensemble (see Figure 6).

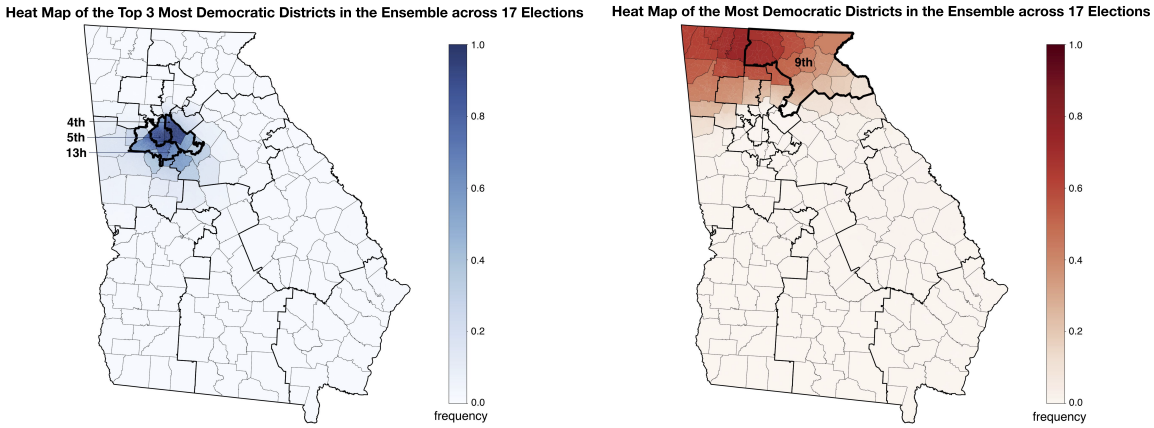

Figure 6: We display the frequency map, over the 17 statewide elections, of the proportion of plans in the ensemble in each precinct is one of the three most Democratic districts (left) and the most Republican district (right).

The geographic locations highlighted by the heat map correspond well to the locations of the enacted districts.[7] In particular, the most Republican district in the ensemble consistently corresponds to the 9th, and the three most Democratic districts correspond to the 4th, 5th and 13th. Using this geographic similarity, we can identify localized differences in partisan behavior of those districts in the 2021 plan and the ensemble. We find that the 4th, 5th and 13th Districts have been *packed*[8] with Democrats, while the 9th District has been used to *crack*[9] the Democratic vote.

5.1 Packing in the three most Democratic districts

In all 17 historic elections across both the old 2011 and new 2021 enacted plans, we find that the three most Democratic districts are the 4th, 5th, and 13th Districts. We also examine where the three most Democratic districts occur in the ensemble by examining the frequency with which each precinct exists within the three most Democratic districts over each of the 17 elections. We plot the resulting heat map in the ensemble and highlight the three most Democratic districts in the 2021 plan in Figure 6 (left). We find substantial similarity between the geographical location of the three most Democratic districts in the ensemble and those of the 2021 plan.

We then compare the fraction of Democratic voters in the three most Democratic districts of the ensemble and the 2021 enacted plan. Despite the geographical similarities, we find that the 2021 plan has the same or more Democrats than 99.41%-99.96% of the corresponding districts in the ensemble of plans. This demonstrates that these three districts, which correspond to a specific region of the state, have been artificially *packed* with Democrats.

[7] In Georgia, there is close alignment between the overarching structures of the 2011 and 2021 district plans; this is to say that the 1st District in 2011 "looks like" 1st District in 2021. We illustrate the similarity between the two plans in Figure 11 of SI.

[8] *Packing* refers to concentrating atypically many voters of one type into a district to reduce their influence in other districts [35].

[9] *Cracking* is dispersing voters of one type into many districts in order to deny them a dominant voting bloc in any particular district [35].

5.2 Cracking in the most Republican district

In the 2021 enacted plan, the 9th District, in the northeastern part of the state, is consistently the most Republican.[10] We find that in the ensemble, the most Republican district is most often located in the northern part of the state and may either encompass precincts to the northwest or northeast (see Figure 6, right). As above, this analysis suggests a similar geographic location of the most Republican district across plans in the ensemble and 2021 enacted plan.

We find the most Republican district in the 2021 enacted plan contains **fewer** Republican voters than over 98.4% of the most Republican district from plans in the ensemble across all elections. This suggests that Democrats have been atypically introduced into this District and, correspondingly, Republicans have been removed. A direct consequence is that the newly included Democrats are removed from surrounding districts. In our analysis below, we show that the Republican voters removed from the 9th District now dilute the voting power of Democrats in 6th and 10th Districts.

To investigate *where* the additional Republican voters have been moved, we begin with the following observations:

(1) According to the 2020 census, the 9th District as drawn in the 2011 plan had only a 1% population deviation from being perfectly balanced in 2020. Other criteria, such as county boundaries, incumbent locations, and seats assigned to Georgia, have not changed. Therefore, this district *did not* have to be substantially redrawn.
(2) The 2011 9th District is consistently the most Republican in the 2011 plan. However, its Democratic vote fraction is typical across all elections when compared to the most Republican district in each plan in the ensemble.

In short, the 9th District could have been almost entirely unchanged, and if so would have had a typical vote fraction in the context of the ensemble. Instead, it has been modified and is now an outlier, relative to our ensemble, with an atypically small number of Republican voters. To determine where these Republican voters are removed,

[10] The 9th District is the most Republican across all but one (the 2020 presidential election, where it was the second most Republican district) of the 17 elections between 2016 and 2020 studied in this report.

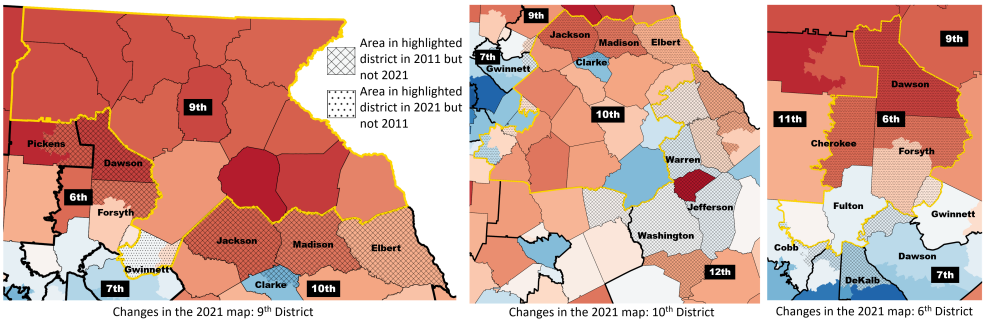

Figure 7: We display how the 2021 plan has been updated from the 2011 plan in the 9th (left), 10th (middle) and 6th (right) Districts, outlined in gold. In each plot, dots show area gained by the district from 2011 to 2021 and crosshatching shows area lost by the district from 2011 to 2021. The vote fraction of the 9th District in the 2011 plan is typical. However, in the 2021 plan, the 9th District absorbs Democratic voters (blue areas of the map) away from more competitive districts and shed Republican voters (red areas of the map) toward these districts. This move has a cascading effect that directly contributes to the changes in the 10th and 6th Districts.

we contrast the 2011 plan with the 2021 plan in Figure 7 (left). We shade the regions that are no longer part of the 9th District with crosshatching and regions where the 9th District has expanded with dots. We also color the counties (and parts of counties) based on the Democratic vote fraction in the 2020 presidential election. The 9th District has been changed to shed Republican leaning regions to the west and south, and it has expanded to the southwest to pick up Democratic voters in northern Gwinnett County.

The exchange, and consequent removal of more Republican voters from the 9th District, has a *cascading effect*. First, it adds Republican voters to the northern part of the 10th District. The 10th District also recedes from the south as shown in Figure 7 (middle). This motion causes the 10th District to gain Republican voters in Jackson, Madison, Elbert, and Clarke counties (once in the 9th District) and give up Democratic voters in Gwinnett County, to the 7th District, and majority-African American Warren, Washington, and Jefferson counties, to the 12th District. The 10th District is within the cluster of plans that have been depleted of Democrats, as presented in the polarization analysis in Section 4.

Similarly, Republican voters removed from the 9th District are added to the 6th District. This causes the 6th to move northward picking up Republican voters in Dawson, northern Forsyth, and eastern Pickens counties (once in the 9th District) and shedding Democratic voters from part of Cobb, Fulton, DeKalb and Gwinnett counties (see Figure 7, right). These shifts dramatically increase the Republican vote fraction in the 6th District. In the 2011 plan, the 6th District was consistently either the fourth- or fifth-most Democratic district; in the 2021 plan it is now one of the districts with atypically few Democrats and would stably elect a Republican representative, according to historic voting trends. Localized analysis of gerrymandering is still a developing field. A new tool has recently emerged to match districts spatially between the ensemble and enacted plans [32]. We plan to utilize this tool in a follow-up study.

6 CONCLUSION

This report shows that the 2021 enacted congressional district plan in Georgia is likely to be highly non-responsive to the changing opinions of the electorate. Moreover, there is mathematical evidence of polarization of competitive districts, which has been caused in part by the redrawing of the 6th, 9th, and 10th Districts. The non-responsiveness of Georgia's congressional plan is highly improbable, even when considering effects of the Voting Rights Act in ensuring 4 of 13 districts can elect an African-American representative (i.e., a proportional number). We note we have not yet tested the effect of enforcing 4 (near) majority-minority districts, so it is possible that this enforcement is what lead to the lack of responsiveness. As found in the Supreme Court Ruling in *Cooper v. Harris*, it is questionable whether such an extreme concentration is legal, so we have omitted such an analysis. See more details in Section 9 of SI.

We implemented tempering techniques on existing multi-scale tree-based methods. To the best of our knowledge, this is the first time tempering has been explicitly used to sample with these techniques. These modifications have allowed us to sample from measures that were previously inaccessible. Specifically, they have allowed us to tune the Polsby-Popper compactness to match that of the enacted plan. We remark that our ensemble *is* still weighted toward plans with higher numbers of associated spanning trees and that further work is still needed to overcome this limitation; however, to our knowledge, this work has successfully sampled from the most flexible target measure on a complicated redistricting problem. Furthermore, these methods should be fully portable to other states. Although beyond the scope of the current work, we have also used these techniques to analyze Georgia's General Assembly plans. We save a full analysis of these plans for future work, but note that these plans are also significantly less responsive than the ensemble of plans. We display some of our results in Section 10 of SI.

REFERENCES

[1] Eric Autry, Daniel Carter, Gregory Herschlag, Zach Hunter, and Jonathan C. Mattingly. 2021. Metropolized Forest Recombination for Monte Carlo Sampling of Graph Partitions. arXiv:1911.01503 [cs.DS]

[2] Eric A. Autry, Daniel Carter, Gregory J. Herschlag, Zach Hunter, and Jonathan C. Mattingly. 2021. Metropolized multiscale forest recombination for redistricting. *Multiscale Modeling & Simulation* 19, 4 (2021), 1885–1914.

[3] Richard Barnes and Justin Solomon. 2021. Gerrymandering and compactness: Implementation flexibility and abuse. *Political Analysis* 29, 4 (2021), 448–466.

[4] Amariah Becker, Moon Duchin, Dara Gold, and Sam Hirsch. 2021. Computational redistricting and the voting rights act. *Election Law Journal: Rules, Politics, and Policy* 20, 4 (2021), 407–441.

[5] Jowei Chen and Jonathan Rodden. 2013. Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quarterly Journal of Political*

Science 8 (2013), 239–269.

[6] Jowei Chen and Jonathan Rodden. 2015. Cutting through the Thicket: Redistricting Simulations and the Detection of Partisan Gerrymanders. *Election Law Journal* 14, 4 (2015), 331–345.

[7] Maria Chikina, Alan Frieze, Jonathan C. Mattingly, and Wesley Pegden. 2020. Separating effect from significance in markov chain tests. *Statistics and Public Policy* 7, 1 (2020), 101–114.

[8] Maria Chikina, Alan Frieze, and Wesley Pegden. 2017. Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences* 114, 11 (Mar 2017), 2860–2864. https://doi.org/10.1073/pnas.1617540114

[9] Wendy K Tam Cho and Yan Y Liu. 2016. Toward a Talismanic Redistricting Tool: A Computational Method for Identifying Extreme Redistricting Plans. *Election Law Journal* 15, 4 (2016), 351–366.

[10] Daryl DeFord and Moon Duchin. 2019. Redistricting reform in Virginia: Districting criteria in context. *Virginia Policy Review* 12, 2 (2019), 120–146.

[11] Daryl DeFord, Moon Duchin, and Justin Solomon. 2021. Recombination: A Family of Markov Chains for Redistricting. *Harvard Data Science Review* (31 3 2021). https://doi.org/10.1162/99608f92.eb30390f https://hdsr.mitpress.mit.edu/pub/1ds8ptxu.

[12] Benjamin Fifield, Michael Higgins, Kosuke Imai, and Alexander Tarr. 2020. Automated redistricting simulation using Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics* 29, 4 (2020), 715–728.

[13] Benjamin Fifield, Michael Higgins, Kosuke Imai, and Alexander Tarr. 2020. Automated Redistricting Simulation Using Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics* 29, 4 (Oct 2020), 715–728. https://doi.org/10.1080/10618600.2020.1739532

[14] Georgia [n.d.]. 2021-2022 Guidelines for the House Legislative and Congressional Reapportionment Committee. https://www.house.ga.gov/Documents/CommitteeDocuments/2021/Legislative_and_Congressional_Reapportionment/2021-2022%20House%20Reapportionment%20Committee%20Guidelines.pdf.

[15] Harper 2019. Harper v. Lewis. NC: District Court; No. 5:19-CV-452-FL.

[16] Harper20 2022. Harper/League of Conservation Voters/Common Cause v. Hall. NC: Supreme Court; No. 413PA21.

[17] Gregory Herschlag, Han Sung Kang, Justin Luo, Sachet Bangia, Christy Vaughn Graves, Robert Ravier, and Jonathan C. Mattingly. 2020. Quantifying Gerrymandering in North Carolina. *Statistics and Public Policy* Under review (2020).

[18] Gregory Herschlag, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier, and Jonathan C. Mattingly. 2020. Quantifying Gerrymandering in North Carolina. *Statistics and Public Policy* 7, 1 (2020), 30–38.

[19] Gregory Herschlag, Jonathan C. Mattingly, Matthias Sachs, and Evan Wyse. 2020. Non-reversible Markov chain Monte Carlo for sampling of districting maps. https://doi.org/10.48550/ARXIV.2008.07843

[20] Gregory Herschlag, Robert Ravier, and Jonathan C. Mattingly. 2017. Evaluating partisan gerrymandering in Wisconsin. *arXiv preprint arXiv:1709.01596* (2017).

[21] Simon Jackman. 2014. The Predictive Power of Uniform Swing. *PS: Political Science & Politics* 47, 2 (2014), 317–321. https://doi.org/10.1017/S1049096514000109

[22] Mark Jerrum and Alistair Sinclair. 1996. The Markov chain Monte Carlo method: an approach to approximate counting and integration. *Approximation Algorithms for NP-hard problems, PWS Publishing* (1996).

[23] Aaron R Kaufman, Gary King, and Mayya Komisarchik. 2021. How to measure legislative district compactness if you only know it when you see it. *American Journal of Political Science* 65, 3 (2021), 533–550.

[24] Lisa Lebovici, Samuel Eure, Rahul Ramesh, Gregory Herschlag, and Jonathan Mattingly. 2018. Gerrymandering and the extent of democracy in America. https://bigdata.duke.edu/projects/gerrymandering-and-extent-democracy-america.

[25] Lewis 2019. Common Cause v. Lewis. NC: Supreme Court; No. 417P19.

[26] Yan Y Liu, Wendy K Tam Cho, and Shaowen Wang. 2016. PEAR: a massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation* 30 (2016), 78–92.

[27] Jonathan C. Mattingly. 2019. Expert Report for Common Cause v. Lewis. *Common Cause v. Lewis* (2019).

[28] Jonathan C. Mattingly. 2019. Rebuttal Of Defendant's Expert Reports for Common Cause v. Lewis. *Common Cause v. Lewis* (2019).

[29] Jonathan C. Mattingly. 2021. Expert Report for Common Cause v. Hall and Harper v. Hall. *NORTH CAROLINA LEAGUE OF CONSERVATION VOTERS v. Hall* (2021).

[30] Jonathan C. Mattingly and Christy Vaughn. 2014. Redistricting and the Will of the People. arXiv:1410.8796 [physics.soc-ph]

[31] Cory McCartan and Kosuke Imai. 2020. Sequential Monte Carlo for sampling balanced and compact redistricting plans. *arXiv preprint arXiv:2008.06131* (2020).

[32] Tom Needham and Thomas Weighill. 2021. Geometric averages of partitioned datasets. *arXiv preprint arXiv:2107.03460* (2021).

[33] Ohio 2019. Ohio A. Philip Randolph Institute v. Householder. In *F. Supp. 3d.* Number 373.1: 18-cv-357. Dist. Court, SD Ohio, 978.

[34] Pennsylvania 2018. League of Women Voters v. Commonwealth. In *A. 3d.* Vol. 178. PA: Supreme Court, 737.

[35] Nicholas O Stephanopoulos. 2017. The causes and consequences of gerrymandering. *Wm. & Mary L. Rev.* 59 (2017), 2115.

[36] Saifuddin Syed, Alexandre Bouchard-Côté, George Deligiannidis, and Arnaud Doucet. 2019. Non-reversible parallel tempering: a scalable highly parallel MCMC scheme. *arXiv preprint arXiv:1905.02939* (2019).

[37] Saifuddin Syed, Vittorio Romaniello, Trevor Campbell, and Alexandre Bouchard-Côté. 2021. Parallel tempering on optimized paths. In *International Conference on Machine Learning*. PMLR, 10033–10042.