

# Auditing for Gerrymandering by Identifying Disenfranchised Individuals

Jerry Lin  
jerry.lin@duke.edu  
Duke University  
Durham, North Carolina, USA

Carolyn Chen  
carolyn.chen047@duke.edu  
Duke University  
Durham, North Carolina, USA

Marc Chmielewski  
marc.chmielewski@duke.edu  
Duke University  
Durham, North Carolina, USA

Samia Zaman  
samia.zaman@duke.edu  
Duke University  
Durham, North Carolina, USA

Brandon Fain  
btfain@cs.duke.edu  
Duke University  
Durham, North Carolina, USA

## ABSTRACT

Gerrymandering is the practice of drawing congressional districts to advantage or disadvantage particular electoral outcomes or population groups. We study the problem of computationally auditing a districting for evidence of gerrymandering. Our approach is novel in its emphasis on identifying individual voters disenfranchised by packing and cracking in local fine-grained geographic regions. We define a local score based on comparison with a representative sample of alternative districtings and use simulated annealing to algorithmically generate a witness districting to show that the score can be substantially reduced by simple local alterations. Unlike commonly studied metrics for gerrymandering such as proportionality and compactness, our framework is inspired by the legal context for voting rights in the United States. We demonstrate the use of our framework to analyze the congressional districting of the state of North Carolina in 2016. We identify a substantial number of geographically localized disenfranchised individuals, mostly Democrats in the central and north-eastern parts of the state. Our simulated annealing algorithm is able to generate a witness districting with a roughly 50% reduction in the number of disenfranchised individuals, suggesting that the 2016 districting was not predetermined by North Carolina’s spatial structure.

## CCS CONCEPTS

• **Theory of computation** → *Random search heuristics*; • **Applied computing** → *Law*.

## KEYWORDS

gerrymandering, fairness, audit, simulated annealing

### ACM Reference Format:

Jerry Lin, Carolyn Chen, Marc Chmielewski, Samia Zaman, and Brandon Fain. 2022. Auditing for Gerrymandering by Identifying Disenfranchised

Individuals. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’22)*, June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3531146.3533174>

## 1 INTRODUCTION

Representative democracy in the United States operates by defining winner-take-all districts. The representative of the individuals in that district is typically elected by plurality vote. The district boundaries can thus impact the representation of individuals. Gerrymandering is the act of drawing districts to favor a particular political outcome or population group. It has the potential to cause social harm in the form of disenfranchisement to political and demographic groups [30]. Many studies in the quantitative sciences have attempted to characterize gerrymandering in terms of optimization criteria such as geographic compactness and proportional representation with respect to political parties [3, 10, 29, 31], and many algorithms have been proposed to generate districts that optimize one or more of these criteria [4, 5, 11, 13].

Less attention has been paid to the question of auditing: How to provide evidence that a particular district plan is in fact gerrymandered. Compactness and proportional representation are not required by law, so we begin by asking what constitutes illegal gerrymandering. Our goal is not to engage in the legal interpretative discourse directly, but rather to use that context to inform our technical contribution in the form of a quantitative framework for auditing gerrymandering.

### 1.1 Historical and Legal Background

There have been several attempts at reining in gerrymandering through the passage and enforcement of legislation. Specifically, the Equal Protection Clause of the Fourteenth Amendment to the United States Constitution enforces the doctrine of “one person, one vote” [20], and the Voting Rights Act of 1965 prohibits gerrymandering on the basis of race [1]. Most legal challenges to gerrymandering are based on interpretations of these provisions [21, 23, 24]. Most straightforwardly, it is understood that districts must have very nearly exactly the same populations, and the court has interpreted the Voting Rights Act to mean that states with significant concentrations of minority populations must guarantee some form of representation to those populations, often by means of so-called “majority-minority” districts.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

FAccT ’22, June 21–24, 2022, Seoul, Republic of Korea

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9352-2/22/06...\$15.00

<https://doi.org/10.1145/3531146.3533174>

There is less consensus about what constitutes illegal *partisan* gerrymandering, other than that proportional representation (that the proportion of representatives of a party should approximate the proportion of voters of that party) is not required [22]. Two commonly discussed notions of partisan gerrymandering are vote concentration (or *packing*) and vote dilution (or *cracking*). Packing attempts draw as many voters of the opposing party into one district as possible, thereby decreasing the opposing party’s chance of winning in neighboring districts. Cracking operates in an inverse fashion. If a large number of voters who would constitute a majority in one district can be divided into several districts, they can be deprived of a representative while setting up a series of “easy wins” for the other party.

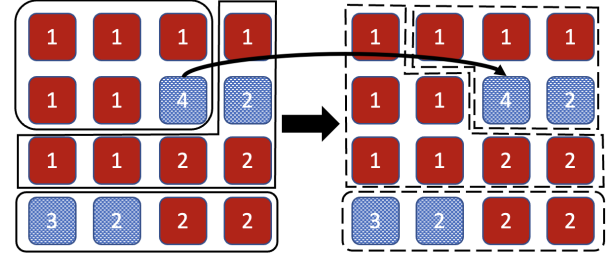
*Gill v. Whitford* was the first major partisan gerrymandering case to be reviewed by the US Supreme Court in the last decade. The plaintiffs attempted to use the idea of the “efficiency gap,” which compares each party’s “wasted” votes—votes cast for a losing candidate or votes for a winning candidate in excess of 50 percent of total votes—to demonstrate their Fourteenth Amendment right to equal protection was violated. But the Supreme Court found that the efficiency gap was merely a “single statewide measure of partisan advantage,” and thus the plaintiffs’ case was about “group political interests, not individual legal rights,” which the Court is not responsible for “vindicating.” Rather, “the Court’s constitutionally prescribed role is to vindicate the individual rights of the people appearing before it” [25].

In a concurring opinion joined by Justice Ginsburg, Justice Breyer, and Justice Sotomayor, Justice Kagan provided more details for how an individual voter may be able to show that he or she has been gerrymandered. As Justice Kagan puts it: “Among other ways of proving packing or cracking, a plaintiff could produce an alternative map (or set of alternative maps)—comparably consistent with traditional districting principles—under which her vote would carry more weight.” Once such justification is met, plaintiffs may then wish to use statewide evidence to demonstrate their charges of statewide packing and cracking in order to seek a statewide remedy from the Court [25].

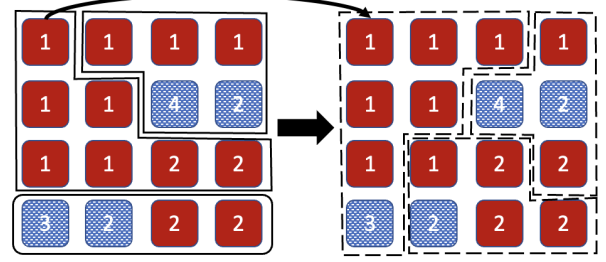
## 1.2 Characterizing Local Gerrymandering

Drawing from Justice Kagan’s statement in *Gill v. Whitford* on the shortcomings of previous justifications of gerrymandering, we take a local—rather than statewide—approach. We devise a metric for how gerrymandered a certain districting is by measuring the number of *people* disenfranchised through packing or cracking in each *voting tabulation district* (VTD). A VTD is a generic term for describing precincts or wards [7], and they are the smallest geographic region for which aggregate voting records are available. In this section we introduce our framework conceptually; formal definitions are deferred to Section 3.

Informally, a districting is a partition of VTDs that satisfies certain constraints such as approximate population balance and contiguity. Any claim that a VTD is packed or cracked in a given districting must refer to a counterfactual districting that shows how the voting power of the VTD could be increased by dispersion or concentration respectively. We call such counterfactual districtings *comparators*. A comparator is not necessarily derived from an



(a) The central blue VTD with 4 voters in the left districting is cracked relative to the comparator at the right. The blue VTD in question receives blue representation under the comparator but not in the original districting.



(b) The top-left red VTD with 1 voter in the left districting is packed relative to the comparator at the right. In both the left districting and the comparator, the red VTD is located in a district in which red wins, but the win in the comparator is achieved with a much lower margin of victory.

**Figure 1: Examples of local gerrymandering in the form of packing and cracking.** There are 16 VTDs arranged in a 4 by 4 grid. There are two parties, one colored solid red and one speckled blue. For simplicity, each VTD has only voters for one party and the number of voters is labeled for each VTD. The total population is 27, 14 of whom support red and 13 of whom support blue. We want to compute three districts, so we place 9 individuals in each district to maintain equal population among districts.

input districting, and there is not necessarily any correspondence between districts in the original map and in the comparator. Instead, a comparator is simply another valid districting.

A claim of packing or cracking must also refer to individuals of a particular partisan preference within that VTD. We say that members of a given party in a VTD are *cracked* in a given districting relative to some comparator if they do not win representation under the districting but do win representation under the comparator. Similarly, we say that they are *packed* if they win representation in both the districting and the comparator, but with a much larger margin of victory in the districting than in the comparator. Note that the two are disjoint in that members of a given party cannot be simultaneously cracked and packed in the same VTD.

Our notions thus take the form of *justified complaints*. Individuals who are cracked may complain that they do not win representation under their preferred partisan affiliation because of an arbitrary selection between valid geographic partitions. Complaints of packing

are inherently less local—that some members of a packed district could perhaps help other neighboring districts to win representation. In either case, the complaint is justified by reference to an alternative valid districting known as a comparator. Note that a comparator need not necessarily claim to be “better” overall than the districting to which it is being compared.

We give a toy example identifying justified complaints of cracking and packing in Figures 1a and 1b respectively (we formally quantify these notions in Section 3.2). Note that these are local complaints of gerrymandering, referencing the individuals of a particular VTD rather than parties writ large across the entire districting. For example, in Figure 1a, the original districting is both proportional and has no efficiency gap (each party has the same number of wasted votes), but nonetheless there is a justified complaint of cracking for some (but not all) of the blue voters.

In Figure 1, there is no districting that will be free from any justified complaints of cracking or packing (one can exhaustively list out all the possible contiguous and equal-population districtings to see this). Indeed, for any given region, there may not exist a districting that is free from any complaints of cracking or packing. The provision of a single comparator is not by itself very compelling evidence that a given VTD has been gerrymandered. How then can we evaluate the magnitude of a justified complaint?

Our basic idea is to simply count the number of comparators for a given partisan affiliation and VTD that demonstrate cracking or packing. More correctly, as the number of comparators may vary, we count the fraction of comparators that demonstrate cracking or packing. This fraction gives the magnitude of a justified complaint: If members of a given party in a given VTD are cracked relative to all comparators then they surely have a valid complaint. A VTD that is cracked relative only to some comparators may still have a legitimate complaint, but the justification is not as strong.

Our measure is of course dependent on the nature of the sample of valid districtings accepted as comparators. This is defined by whatever explicit constraints are placed on the districting process such as approximate population balance and contiguity, but possibly including additional constraints such as not splitting municipal or county borders, ensuring a minimum number of majority-minority districts, or maintaining at least a certain level of compactness or “simplicity” of shapes. Given any such explicit characterization, one can obtain a large representative sample of comparators and our notions of cracking and packing are well-defined.

When auditing for gerrymandering by measuring the local cracking and packing scores, it may be insufficient to simply state the scores. Since there is no single districting that is free from all local cracking and packing in general, how should one interpret what appears to be high levels of gerrymandering in a region? Those wishing to defend the current districting may claim that there would be as many justified complaints no matter what they did in terms of drawing districts—i.e., that it is a necessary result of the distribution of voters in the state. To examine this defense, we introduce the idea of a *witness* districting that can be computed by small perturbations of a given districting. Specifically, we compute a witness by a simulated annealing VTD swapping algorithm to find an alternative districting that reduces the number of gerrymandered individuals locally while maintaining global constraints such as population balance and compactness at approximately the same

level as the districting under audit. If there exists such a witness that dramatically decreases the number of gerrymandered individuals, it constitutes evidence against such a defense of a gerrymandered districting.

### 1.3 Contributions and Results

Our primary conceptual contribution is a framework for auditing gerrymandering. This framework involves a precise quantitative score of local packing and cracking in Section 3 along with the concept of a witness against those levels of packing and cracking. Our score is motivated by the legal framework of partisan gerrymandering and can be computed on real data as part of an audit for gerrymandering. Our method of measurement allows us to identify individual impacted voters and gerrymandered regions of voters with “heat maps,” as shown in Figure 2a.

Technically, we show how to compute our scores by reference to a sample of comparators in Section 3. These can be generated by Markov Chain Monte Carlo techniques as shown in prior work [14]. Given a scoring for a particular districting, we show how to compute a witness against that districting with a lower score, demonstrating how lower levels of gerrymandering are possible with small perturbations. Our simulated annealing algorithm is described in Section 4. Our framework is highly flexible and allows for mixing-and-matching of sampling and witness generation methods. One could use a different method to sample comparators or a different procedure to generate a witness districting but still use our general auditing framework.

Empirically, in Section 5, we demonstrate the use of our framework to audit the congressional districtings of the state of North Carolina in 2012 and 2016. We find a total score of over 1.5 million (recall that each individual voter is cracked or packed—but not both—at a value between 0 and 1) for the 2012 districting and over 1.1 million for the 2016 districting. Figure 2a plots the gerrymandering score by capita at the VTD level from 2016. The impact is not equal across parties or geographic regions. The score for Democrats is more than 2.9 times as high in the 2012 map as for Republicans and more than 5 times as high in the 2016 map. The effect is heavily concentrated in the central and north-eastern parts of the state. Democratic voters in the Charlotte and Raleigh metro areas are heavily packed, and Democratic voters in Greensboro and suburban areas of Raleigh are heavily cracked in the official 2016 districting.

We show that the level of gerrymandering we observe in the 2016 districting of North Carolina is not a necessary result of the distribution of voters. Our simulated annealing algorithm produces a witness districting with a gerrymandering score of less than 0.6 million based on small perturbations to the original 2016 districting and maintaining comparable levels of population balance and compactness—as illustrated in Figure 2b, where we plot the gerrymandering score by capita at the VTD level with respect to the resulting districts of our best run of our algorithm. We also consider two additional witnesses, one *nonpartisan* and the other *bipartisan*. The former is produced by another algorithm for districting that does not take partisan affiliation into account, only geometry and population [19]. This districting achieves a score of less than 1 million. The latter was generated by a bipartisan panel of 10 retired

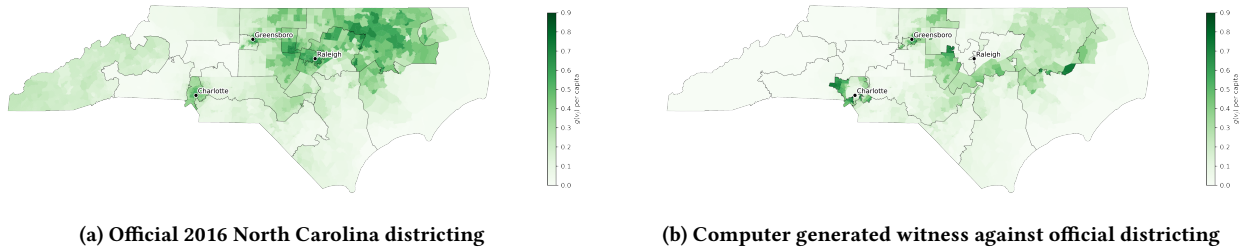


Figure 2: Total per capita gerrymandering score in each VTD

North Carolina judges. This districting achieves a gerrymandering score of just over 1 million, but the gerrymandering scores for Democrats and Republicans are within a factor of 1.2, as compared to the factor of 5 difference for the real 2016 districting.

Our approach offers an interpretable, informative metric for gerrymandering in terms of people impacted, and it allows one to precisely pinpoint which populations are gerrymandered and to what extent, satisfying the stipulations for Article III standing from *Gill v. Whitford*.

## 2 RELATED WORK

The study of districting and gerrymandering is highly interdisciplinary, with contributions being made from mathematics, computer science, political science, law, and other fields. Here, we briefly survey some of the recent work more directly relevant to our contribution relating to the specific issues of algorithmic districting and quantitative metrics.

### 2.1 Redistricting Algorithms

One branch of algorithms has focused solely on creating districts that are compact and have valid population deviations, ignoring electoral and demographic data. Altman [2] shows that the problem of computing legal districtings in this sense is NP-complete. Thus, various algorithms based on heuristics have been proposed.

Levin and Friedler [19] devise a divide and conquer algorithm that achieves a population deviation that is less than 0.005% for 42 out of 43 multi-district states, one of the best in the literature. A simulated annealing algorithm by Brian Olson [27] produces population deviations under 1% in all U.S. states but produces better compactness scores compared to the Levin and Friedler algorithm.

Another branch of algorithms does incorporate electoral data into the algorithmic problem of redistricting. Gurnee and Shmoys [13] create a two-part linear programming algorithm that allows users to pick and choose a fairness metric to optimize. Garg et al. [11] uses Gurnee and Shmoys' algorithm to create maps that optimize for either partisan benefit or proportionality in the case of multi-member districts (MMD). They find that MMDs using single-transferable vote can curtail legislators' ability to perform partisan gerrymandering and can allow for redistricting commissions to potentially create consistently proportionally representative maps.

Becker et al. [5] specifically focus on creating maps that satisfy the Voting Rights Act (VRA). They use a Markov chain procedure to propose successive modifications to districting plans, combining

that with an ecological-inference procedure (inferring information about the individual from ecological, or aggregate, data) and a benchmark plan to identify districtings that give minority candidates effective opportunities to elect their preferred candidates. While they show a proof of concept of their method in Texas, they also emphasize their contribution to be the overarching protocol rather than any single design choice.

### 2.2 Quantifying Gerrymandering via Metrics

Another direction of work has focused on quantifying gerrymandering. This is more in line with our work on auditing. One immediately thinks of proportionality, where the number of seats held by each party reflects the partisan lean of the state, but the courts do not require nor mandate proportionality in a map [18].

Stephanopoulos and McGhee define a metric called the *efficiency gap*, which is calculated as the difference between two parties' wasted votes (votes that are either in excess of the majority needed to win or that go towards a losing candidate) in an election, divided by the total number of votes cast [31]. They argue this metric improves upon the simple idea of proportionality and the widely-used metric of *partisan bias*, which measures the difference in seats won by each party in the counterfactual case where each party received equal vote share. But Duchin and Bernstein criticize the efficiency gap for penalizing proportionality (instead preferring double proportionality) and fetishizing three-to-one landslide districts [6, 32]. The efficiency gap was used in *Gill v. Whitford* and was ultimately unsuccessful at striking down Wisconsin's legislature districts [25].

In response to the double-proportionality nature of the efficiency gap, Warrington created a metric called *declination* [34]. A line drawn through a plot of districts ordered by increasing percent of Democratic (or Republican) voters should make a relatively straight line if districts were not gerrymandered. However, if districts were gerrymandered, one might see a sharp kink at the 50% line on the y-axis. The angle of this kink is thus the declination.

Chen and Rodden use a simulation procedure to generate random, partisan-blind maps that only consider equal population, contiguity, and compactness constraints [8]. Herschlag et al. generate an ensemble of valid comparator maps to analyze the likelihood of the outcome of a given state map compared to the outcomes in the comparators [14]. Using Markov Chain Monte Carlo (MCMC) methods, they are able to show that the electoral outcomes of North Carolina's districtings used in the 2012 and 2016 elections are outliers relative to the ensemble of comparators, while a bipartisan map drawn by retired judges is more representative (in probability space

and in the partisan sense). Their analysis was used in *Rucho v. Common Cause*, where the courts ruled that partisan gerrymandering is not a matter for the federal courts.

### 2.3 Limitations of Existing Approaches

Advances in algorithms have allowed for the cheap and scalable generation of maps that can optimize for fairness. However, what metric of fairness to use is still an open question. A common issue with these statewide, aggregate approaches like outlier analysis and declination is that they are unable to identify which people in the state are most gerrymandered and to what degree; they simply identify how badly a party has been affected by gerrymandering.

This is a problem because courts care about disenfranchisement of people, not parties. For example, *Gill v. Whitford* exemplified that in order to advance vote dilution arguments, individual voters must show that they have been packed or cracked [35]. Instead of focusing on metrics at the statewide level, we take a local approach and devise a metric for how gerrymandered a certain districting is by measuring the number of *people* gerrymandered.

## 3 GERRYMANDERING SCORE FUNCTION

In this section we define our gerrymandering score function and discuss its calculation from a sample of comparators.

### 3.1 Legal Districting Problem Definition

We assume a two-party, winner-takes-all electoral system, like that found in the United States. Following Levin and Friedler [19], we consider the redistricting problem to be a set partition problem. Let  $G = (V, E)$  be a connected graph that represents the geography of a state. Pick a particular historical election from which to draw voter data. Each vertex  $v \in V$  represents a voting tabulation district (VTD), the smallest local level at which aggregate voting records are available (typically a voting precinct). We define functions  $d(v)$ ,  $r(v)$ , and  $t(v)$  to return the number of Democratic votes, Republican votes, and total votes cast in VTD  $v$  in election  $h$ , respectively, and we define  $w(v)$  to be the population of the VTD. Note that  $t(v) \neq w(v)$  because only citizens age 18 and above can vote, and not all voting age citizens participate in every election. Due to third party candidates,  $t(v)$  also need not equal  $r(v) + d(v)$ . There is an edge between vertices  $v, w \in V$  if the associated population units' boundaries are adjacent to each other.

A districting  $D = \{G_1, G_2, \dots, G_k\}$  is a partition of  $G$  into  $k$  connected subgraphs, each representing a district. The districts are denoted by  $G_i = (V_i, E_i)$  where  $V_i \subset V, E_i \subset E$ . The total population of a state is  $n = \sum_{v \in V} w(v)$ , and the total population of each district  $G_i$ , denoted  $population(G_i)$ , equals  $\sum_{v \in V_i} w(v)$ . The load capacity  $C$  is the number of people per district if each district has equal population, so  $C = n/k$ .

We can define several relevant normative properties for a districting. At a minimum, it is typically the case that each subgraph  $G_i \in D$  must be connected; this constraint ensures that all districts in  $D$  are *contiguous*. Furthermore, the *population deviation* from  $C$ , defined as  $\epsilon = (\max_i \{|population(G_i) - C|\})/C$ , is typically constrained to be some small constant (e.g., 0.01). This ensures that each district has approximately the same population. We can characterize how "compact" a district  $G_i \in D$  is based on how "simple" the shape is

geometrically. We use the *Polsby-Popper* compactness score: The ratio of the area of the district to the area of a circle that has the same perimeter as the district, i.e.  $4\pi \times (area(G_i)/perimeter(G_i))^2$  [28]. Scores closer to zero are considered less compact. Our methods are not specific to this particular measure of compactness, and alternatives could be substituted into our framework.

We conduct an audit against a particular districting  $D$  using a sample  $\mathcal{F}$  of valid alternative districtings. For intuition, you can imagine that  $\mathcal{F}$  is defined to be the set of all contiguous districtings with population deviation at most 0.015 and each district constrained to have Polsby-Popper compactness of at least some constant lower bound. There could be additional constraints imposed on  $\mathcal{F}$  such as not splitting counties, including a minimum number of districts with large minority populations, etc. In general,  $\mathcal{F}$  simply defines a large representative sample of valid districtings.

### 3.2 Gerrymandering Score

We build on Warrington's definition of cracking and packing to precisely define a gerrymandering score [35]. Cracking and packing can only be quantified by comparing a candidate districting (i.e., the one under audit) to a comparator districting. We define a voter to be *cracked* if they are in a district where their preferred party loses, but in the comparator they are in a district where their preferred party wins.

**Definition 3.1 (Cracking).** Let  $D = \{G_1, G_2, \dots, G_k\}$  be the current districting and  $D_0 = \{H_1, H_2, \dots, H_k\}$  be a comparator districting. A VTD  $u \in V_i, W_j$  where  $G_i = (V_i, E_i)$  and  $H_j = (W_j, F_j)$  is **cracking Democrats relative to  $D_0$**  if

$$\sum_{v \in V_i} d(v) < \sum_{v \in V_i} r(v) \text{ and } \sum_{v \in W_j} d(v) > \sum_{v \in W_j} r(v)$$

and **cracking Republicans relative to  $D_0$**  if

$$\sum_{v \in V_i} r(v) < \sum_{v \in V_i} d(v) \text{ and } \sum_{v \in W_j} r(v) > \sum_{v \in W_j} d(v)$$

Note that this definition reflects how the same VTD  $v$  may be experiencing different outcomes under  $D$  and  $D_0$  because  $v$  is assigned to different districts  $G_i$  and  $H_j$  respectively under the two districtings.

Similarly, we define a voter to be *p-packed* if their preferred party wins in both maps, but with a margin of victory that decreased by  $p$  or more in the comparator relative to the candidate. For example, if  $p = 0.05$ , the margin of victory would have to decrease by 5 or more percentage points to signal packing. For our experimental analysis in Section 5, we use a threshold of  $p = 0.05$  to moderate the sensitivity of the packing criteria. The particular value of  $p$  is not important to our general framework and can be changed to whatever is deemed sufficient evidence of packing.

**Definition 3.2 (Packing).** Let  $D = \{G_1, G_2, \dots, G_k\}$  be the current districting and  $D_0 = \{H_1, H_2, \dots, H_k\}$  be a comparator districting. Let  $p \in (0, 0.5)$ . Given  $D_0$ , a VTD  $u \in V_i, W_j$  where  $G_i = (V_i, E_i)$  and  $H_j = (W_j, F_j)$  is **p-packing Democrats relative to  $D_0$**  if

$$\sum_{v \in W_j} d(v) > \sum_{v \in W_j} r(v) \text{ and } \frac{\sum_{v \in V_i} d(v)}{\sum_{v \in V_i} t(v)} - \frac{\sum_{v \in W_j} d(v)}{\sum_{v \in W_j} t(v)} \geq p$$

and **p-packing Republicans relative to  $D_0$**  if

$$\sum_{v \in W_j} r(v) > \sum_{v \in W_j} d(v) \text{ and } \frac{\sum_{v \in V_i} r(v)}{\sum_{v \in V_i} t(v)} - \frac{\sum_{v \in W_j} r(v)}{\sum_{v \in W_j} t(v)} \geq p$$

For a given comparator map in  $\mathcal{F}$ , we compare it against the districting under audit and sum up the number of people cracked and packed for each of the districting's VTDs. For each VTD, if it is cracking or p-packing Republicans, we include the number of Republican voters in the sum, and if it is cracking or p-packing Democrats, we include the number of Democratic voters in the sum. We average over  $\mathcal{F}$  to obtain the overall gerrymandering score for each VTD. The gerrymandering score of the entire state is then just the sum of the score over all VTDs.

**Definition 3.3 (Gerrymandering Score).** We are given the current districting  $D$ , a sample of comparators  $\mathcal{F}$ , and a graph of the state  $G = (V, E)$ . For each comparator  $D' \in \mathcal{F}$  and VTD  $v \in V$ , let  $e(D', v) = 1$  if  $v$  is either cracking or p-packing Democrats relative to  $D'$  and 0 otherwise. Similarly, let  $s(D', v) = 1$  if  $v$  is either cracking or p-packing Republicans relative to  $D'$  and 0 otherwise.

Define the **gerrymandering score** of a VTD  $v$  with respect to comparators  $\mathcal{F}$  to be

$$g(v, \mathcal{F}) = \frac{1}{|\mathcal{F}|} \sum_{D' \in \mathcal{F}} (e(D', v)d(v) + s(D', v)r(v)).$$

The **gerrymandering score** of the entire districting with respect to comparators  $\mathcal{F}$  is simply the sum of gerrymandering scores of each VTD

$$\sum_{v \in V} g(v, \mathcal{F}).$$

In Figure 3 we walk through a simple example of our process. We can compute the gerrymandering score for each VTD in this toy state. Consider the specific VTD  $v$  in the upper right corner of the original districting, consisting of one voter who supports the blue party. Let the district containing  $v$  in the original districting be  $d$ . In  $d$ , the red party is winning with a  $\frac{4}{7} - \frac{3}{7} = \frac{1}{7}$  margin of victory (in terms of fractions of the total votes casted).

To compute the gerrymandering score for VTD  $v$ , we then iterate through each comparator districting in  $\mathcal{F}$ —there are seven in this example—and sum up  $e(D_i, v)d(v) + s(D_i, v)r(v)$  for the  $i$ 'th comparator. Consider Comparator 1 and let the district that contains VTD  $v$  be district  $d_1$ . In district  $d_1$ , the blue party is winning with a  $\frac{6}{7} - \frac{1}{7} = \frac{5}{7}$  margin of victory. Since the blue party was losing in  $d$ , but is winning in the Comparator 1 district  $d_1$ , the blue party is cracked in VTD  $v$  relative to Comparator 1. On the other hand, there is no evidence of cracking or packing against the red party in this case. Thus, the gerrymandering score for VTD  $v$  with respect to Comparator 1 is just the number of blue voters, 1.

One can repeat this calculation for the other six comparators. Ultimately, the sum of gerrymandering score for  $v$  relative to each comparator in turn from 1 to 7 is just  $1 + 0 + 0 + 1 + 0 + 1 + 1 = 4$ . The final gerrymandering score for the VTD  $v$  is just the average of the gerrymandering score against each comparator, yielding  $\frac{4}{7}$ . The total gerrymandering score for the whole state is the sum of the gerrymandering score for each VTD. In this example, the gerrymandering score for the state would be  $\frac{6}{7} + \frac{4}{7} + \frac{12}{7} + 3 = 6\frac{1}{7}$ .

Subfigure 3b also demonstrates that individuals in the dark-green VTD (with 3 individuals gerrymandered) have the strongest justification to complain that they are victims of partisan gerrymandering. Out of all 8 possible districting plans (the original districting plus 7 comparators) their preferred party (the blue party) wins in every scenario *except* in the districting under audit.

### 3.3 Obtaining Comparator Maps

In general, the set of all feasible districtings—subject only to minimal nonpartisan constraints such as contiguity, compactness, and equal population—is a large combinatorial object of VTDs. We cannot expect it to be enumerated exactly and so cannot expect  $\mathcal{F}$  to be exhaustive of all possibilities or given explicitly up front. Instead, we follow recent work on sampling from some nonpartisan distribution on  $\mathcal{F}$  [14]. In principle the distribution could be uniform, though in practice one may wish to put additional weight on “nice” districtings that are highly compact or exhibit other desirable nonpartisan properties. Given such a distribution, [14] shows that Markov Chain Monte Carlo techniques can be employed to efficiently generate a sample of comparators.

We argue that it is sufficient for the purpose of approximating our gerrymandering score defined on some large number of feasible districtings  $\mathcal{F}$  to have access to a random sample  $\hat{\mathcal{F}}$ .

**THEOREM 3.4.** *Let  $\hat{\mathcal{F}}$  be an i.i.d. random sample of districtings from  $\mathcal{F}$ . Let  $v \in V$  be a VTD. Then*

$$\Pr\left(|g(v, \hat{\mathcal{F}}) - g(v, \mathcal{F})| \geq p \cdot t(v)\right) \leq 2e^{-2|\hat{\mathcal{F}}|p^2}.$$

**PROOF.** The statement follows from a simple application of Hoeffding's inequality, see Theorem 2 in [15]. Note that  $g(v, \mathcal{F}) = \mathbb{E}[g(v, \hat{\mathcal{F}})]$ . Furthermore, note that  $g(v, \hat{\mathcal{F}})$  is the sum of  $|\hat{\mathcal{F}}|$  independent random variables—one for each  $D' \in \hat{\mathcal{F}}$ —of the form

$$\frac{1}{|\hat{\mathcal{F}}|} (e(D', v)d(v) + s(D', v)r(v)).$$

Each such random variable is bound between 0 and  $t(v)/|\hat{\mathcal{F}}|$ . The theorem statement follows directly.  $\square$

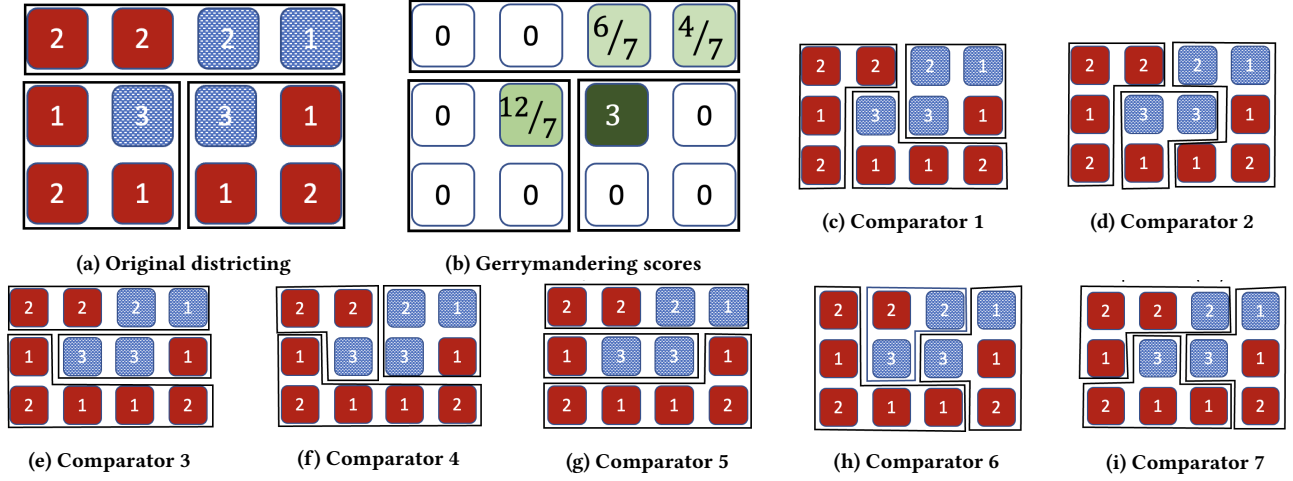
This allows us to derive a confidence interval for an estimate of the gerrymandering score based on a random sample.

**COROLLARY 3.5.** *Let  $\gamma(v, \hat{\mathcal{F}}) = 1.358 \frac{t(v)}{\sqrt{|\hat{\mathcal{F}}|}}$ . With probability at*

*least 95% over the random draw of  $\hat{\mathcal{F}}$ ,  $g(v, \mathcal{F})$  lies in the range  $g(v, \hat{\mathcal{F}}) \pm \gamma(v, \hat{\mathcal{F}})$ .*

In practice, we obtain a sample of 1055 comparator maps from a study by Herschlag et. al [14]. So, for our data empirical results,  $\gamma(v, \hat{\mathcal{F}}) \approx 0.00129t(v)$ . This implies, in particular, that we have reasonable confidence in our gerrymandering scores plus or minus one for a typically sized VTD of hundreds of individuals.

To sample from the space of congressional redistricting plans in NC, Herschlag et. al [14] construct a probability distribution that is weighted toward districting plans adhering to non-partisan design criteria. There are four non-partisan design criteria: requiring equal population among NC's 13 districts, having compact districts, minimizing split counties, and ensuring African-American voters



**Figure 3: Example calculation of gerrymandering score.** There are 12 VTDs arranged in a 4 by 3 grid. There are two parties, one colored solid red and one speckled blue. For simplicity, each VTD only has voters for one party and the number of voters is labeled for each VTD. The total population is 21, 12 of whom support red and 9 of whom support blue. We want to compute  $k = 3$  districts, so 7 individuals comprise a population equal district. 7 alternative equal population contiguous comparators are shown.

are sufficiently concentrated in two districts to affect the winner. They capture these criteria in a score function and use this to define a Gibbs distribution. Using a Monte Carlo Markov chain (MCMC) algorithm combined with a simulated annealing procedure, they generate compliant redistricting plans by sampling from this probability distribution. Critically, partisan voting data or demographic data is not used when sampling redistricting plans. Analysis based on this sampling method has been presented as evidence in numerous court cases, including before the Supreme Court in the 2019 case *Rucho v. Common Cause* [14]. However, we note that one can substitute any sampling procedure of their choosing into our framework and calculate a gerrymandering score.

#### 4 GENERATING A WITNESS

Suppose one computes a gerrymandering score for each VTD as outlined in the previous section. The results may be concerning in many ways including high scores overall, regional variation, or very different levels of impact when broken down by partisan affiliation or demographics. However, the designer of the districting under audit may respond to such concerns by claiming that it is a natural consequence of the distribution of voters in the state. There is no guarantee that there exists a districting with a gerrymandering score of 0, so this defense may seem plausible.

In this section, we describe an algorithm to compute a *witness* against a districting being so defended. Computing such a witness is necessary to contextualize the gerrymandering score by showing how small perturbations to the districting under audit can substantially reduce the gerrymandering score. This witness is not necessarily intended to be used as an alternative “improved” map. Rather, the existence of this witness simply weakens the argument that partisan gerrymandering in the original districting was “unavoidable.”

We propose a *simulated annealing* swapping algorithm (Algorithm 1) that, given an initial districting, continually swaps VTDs that border two districts to reduce the gerrymandering score.

Simulated annealing is a metaheuristic, defined by a *temperature* variable. As the algorithm runs its course, the temperature decreases geometrically by a *tempFactor* multiplier. While the temperature is high (i.e., during the initial steps of the algorithm), it allows swaps that are non-beneficial to be made. While the temperature is low (i.e., nearing the end), it will greatly prefer beneficial swaps. We use a generic simulated annealing framework such as in [17] and modify it to fit our purpose. We use this simulated annealing approach instead of a simpler greedy local search algorithm because the problem is highly constrained and there are many local optima.

We begin the algorithm with the initial districting. At the beginning of each iteration of the algorithm, on line 3, we calculate candidate VTDs for swaps as all VTDs that border any district in the districting. If the candidate VTD is currently in district  $G_i$  and borders a district  $G_j$ , a *swap* adds the candidate vertex to district  $G_j$  and removes the candidate vertex from district  $G_i$ . We randomly iterate through these candidates looking for a swap that is *valid* (line 4).

A valid swap is a swap that satisfies three constraints. First, we do not allow the average or minimum Polsby-Popper compactness score of all the state’s districts to decrease from the original districting’s average and minimum Polsby-Popper scores by more than a percentage which we call the compactness deviation, *compDev*. Second, we limit the population deviation  $\epsilon$  between districts to no more than a parameter we call *popDev*. Finally, we require contiguity. One district cannot fully enclose part of another district inside of it, nor can they consist of two disjoint, unconnected elements, except in special cases (e.g. when the district contains an island). If a swap means the resulting districts satisfy these three conditions—compactness, population deviation, and contiguity—then it

**Algorithm 1:** Simulated Annealing Algorithm

---

```

1 frozenCount  $\leftarrow$  0, iterations  $\leftarrow$  0,  $T \leftarrow$  startTemp,
  champScore  $\leftarrow$   $\infty$ ;
2 repeat
3   candidates  $\leftarrow$  VTDs that border any district in current
    districtShapes;
4   for swap in candidates do           // Iterate through
    candidates in random order
5     if iterations equals  $L$  then
6        $T \leftarrow$  tempFactor *  $T$ ;
7       if number of accepted swaps/iterations <
        minPercent then
8         frozenCount  $\leftarrow$  frozenCount + 1;
9       if current gerrymandering score < champScore
        then
10        frozenCount  $\leftarrow$  0;
11        champScore  $\leftarrow$  current gerrymandering
          score;
12        if frozenCount is 5 then
13          return districtShapes;
14      if the swap satisfies population deviation,
        compactness, and contiguity then
15        iterations += 1;
16        if the swap improves the gerrymandering score
          then
17          Accept the swap and update districtShapes;
18        else
19           $\Delta \leftarrow$  post-swap score – pre-swap score;
20          Accept the swap and update districtShapes
            with probability  $e^{-\Delta/T}$ ;
21        if the swap was accepted then
22          break;           // Recalculate candidates

```

---

is a valid swap. Note that additional desiderata (such as maintaining county boundaries or majority-minority districts) could easily be incorporated into this notion of a valid swap within the algorithm.

Out of these valid swaps, we always accept swaps which lower the gerrymandering score (line 16). To avoid local optima, we also accept swaps that increase the gerrymandering score with probability  $e^{-\Delta/T}$ , where  $\Delta$  is the increase in gerrymandering score and  $T$  is the current temperature (line 20). Upon accepting a swap, we recalculate candidate VTDs for swaps and begin another iteration of the algorithm.

We initially set  $T$  to be an input variable called *startTemp*. We also define input variables *tempFactor*, the temperature length  $L$ , and *minPercent*. After  $L$  valid swaps have been proposed (these valid swaps do not necessarily have to be accepted and actually swapped), we set  $T = T * \text{tempFactor}$ . Internally, the algorithm defines a *frozenCount* variable that is initially set to 0. The algorithm halts on line 12 if *frozenCount* reaches some threshold (we use 5 in our implementation), which ends the algorithm once no progress is being made. To determine when to halt, we define a variable called *champScore*. After every  $L$  swaps at the specified temperature level,

on line 11, we set *champScore* to be the current districting’s gerrymandering score if it is better than the current *champScore*. After every  $L$  swaps, we also increment *frozenCount* by 1 on line 8 if both the number of accepted swaps divided by  $L$  is less than *minPercent* and *champScore* does not improve (meaning the algorithm is not making much progress). If *champScore* does improve, then we reset *frozenCount* to 0 on line 10.

## 5 EMPIRICAL RESULTS

We performed all our coding in Python 3, relying on the *geopandas* package for spatial data analysis and operations. Our code is publicly available on GitHub.<sup>1</sup>

### 5.1 Gerrymandering Score

We calculated the overall gerrymandering score of four different districting plans for North Carolina: the official map used in the 2012 election (the 2012 map), the official map used in the 2016 election (the 2016 map), the map generated using the Divide and Conquer Redistricting Algorithm developed by Levin and Friedler (the ACR map) [19], and the map drawn by a bipartisan panel of 10 retired North Carolina judges (the Judges map). For readability, displayed scores are truncated to the nearest whole number. When drawing the Judges map, the judges focused “only on keeping each relatively equal in population and in compliance with the federal Voting Rights act,” without “considering political party registration or voting history.” [9] Each of these four maps’ district numberings and partisan leans are shown in Figure 4. The geographic distribution of the gerrymandering score in the 2016 map can be seen in Figure 2a.

To generate the 2012, 2016, and Judges map districtings, we used files compiled by Herschlag et al. [14]. We generated the ACR map through our own Python implementation of Levin and Friedler’s algorithm [19]. Population and demographic data was drawn from the 2010 census [12]. We used 2016 presidential election data as the underlying electoral data to simulate which party would hypothetically win each district. This historical data was drawn from that compiled by Herschlag et al. [14]. In the 2016 North Carolina presidential election, there were 2,362,631 Republican votes (49.8% of total) and 2,189,316 Democratic votes (46.2% of total) [33].

Table 1 shows various characteristics of these four maps, including how they fare on our gerrymandering metric. The Judges map and the ACR map, which did not take party data into account in the districting process, performed better overall, with lower gerrymandering scores than the 2012 and 2016 maps. Although both the ACR and Judges maps do not take party data in account, the Judges map can be thought of as a bipartisan gold standard (drafted by experts from both parties), and our metric is able to capture this by showing a better balance of the gerrymandering score between Democrats and Republicans. On the other hand, the ACR map is simply *nonpartisan*, so although it is able to achieve the lowest score among all 4 maps, it is less balanced by party.

The 2012 map performed the worst, with a gerrymandering score of 1,503,519. The score is 197% higher for Democrats than Republicans. Since this districting plan was struck down by the Supreme

<sup>1</sup><https://github.com/auditing-gerrymandering/gerrymandering>

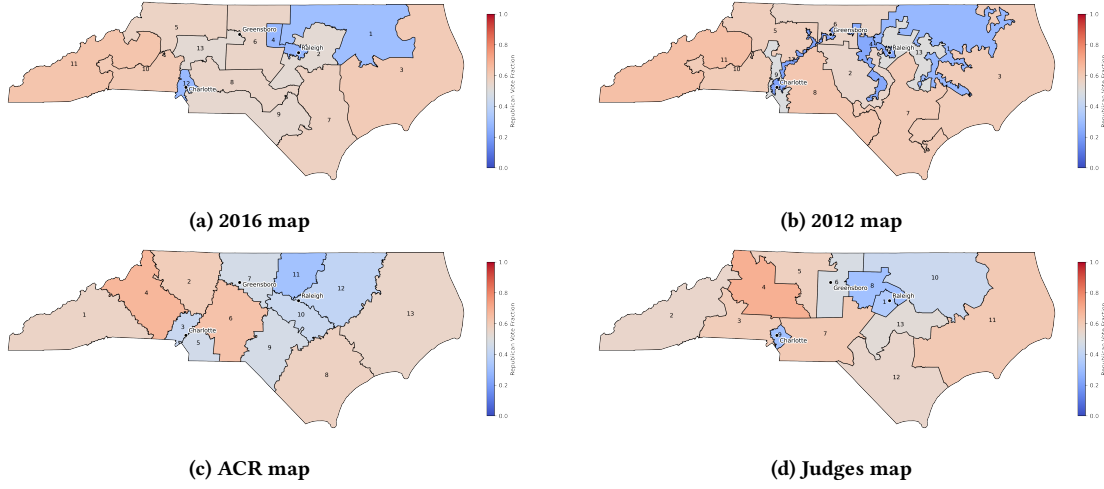


Figure 4: District numberings and partisan leans for each of the four initial maps

Table 1: Gerrymandering scores (broken down by Democrats and Republicans), number of seats won by each party, efficiency gap, and average compactness (using Polsby-Popper metric) for the 2012, 2016, ACR, and Judges maps prior to any swapping.

Districting Plan	Gerrymandering Score for			Rep. Seats	Dem. Seats	Eff. Gap	Avg. Polsby-Popper
	Total	Dem.	Rep.				
2012 map	1,503,519	1,124,669	378,850	10	3	R+24%	0.11867
2016 map	1,151,027	959,777	191,250	10	3	R+22%	0.24774
ACR map	931,975	250,247	681,727	7	6	D+9%	0.30971
Judges map	1,026,842	550,404	476,437	8	5	R+6%	0.33094

Court due to racial gerrymandering [16], this is not surprising, and in general, our score measurements for these districtings are consistent with our expectations.

## 5.2 Visualizing the Gerrymandering Score

Recall the definition of a VTD being cracked or packed in Definitions 3.1 and 3.2. Using the method described in Warrington [35], when iterating over all comparator districtings, we can calculate for each VTD what percent of the time that VTD is cracked or packed against Democrats, and what percent of the time the VTD is cracked or packed against Republicans. Then, we can shade each VTD with color based on what percent of the time it is cracked or packed, thus creating a “heat map,” depicting which regions of the state are more gerrymandered than others from different partisan perspectives. This local approach allows us to clearly identify regions of voters who have been impacted by cracking or packing.

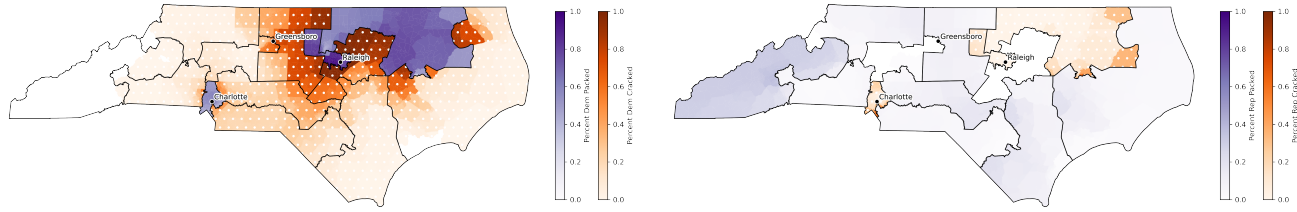
Figure 5 is a heat map generated on the 2016 map. It demonstrates how Democratic voters in the Charlotte and Raleigh-Durham metro areas were heavily packed in the 2016 map, while Democratic voters in Greensboro and suburban areas of Raleigh-Durham were heavily cracked.

## 5.3 Algorithmically Generated Witness

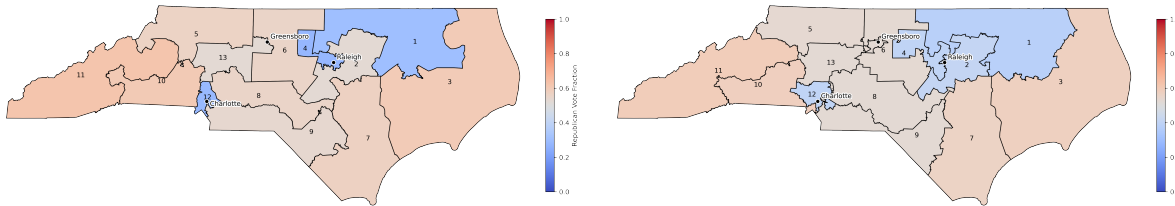
We generated witnesses against multiple initial districtings, including all those shown in Figure 4. Here, we report the witness against

the 2016 map, as it was in real use in North Carolina and its gerrymandered status has been disputed [26]. In one of our best runs on the 2016 districting, our simulated annealing algorithm generates a witness that decreases the gerrymandering score from 1,151,027 to 571,611, comprising a 50% reduction. We report results from this one run for simplicity, but in 50 random trials we obtain a gerrymandering score less than 627,050 at least half of the time. We conclude that the high gerrymandering score of the 2016 congressional districting of North Carolina is not a necessary feature of the distribution of voters in the state. Though it was not an explicit objective of our algorithm, we note that the resulting map has 4 instead of 3 Democratic districts, and multiple districts appear more competitive, with parties winning by smaller margins of victory, as can be seen in Figure 6. In this run, we used the following parameters:  $popDev = 0.02$ ,  $compDev = 0.1$ ,  $startTemp = 800$ ,  $tempFactor = 0.9747$ ,  $L = 272$ ,  $minPercent = 0.02$ , and the districting plan = 2016 map.

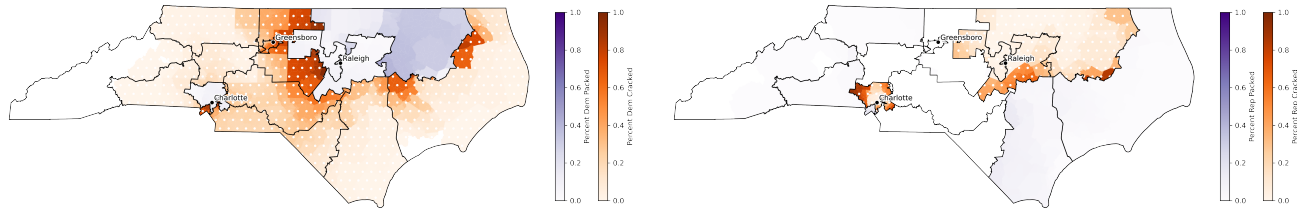
In the original 2016 heat map shown in Figure 5, Charlotte, Raleigh, and the rural first district were heavily packed while the Greensboro and suburban Raleigh areas experienced much cracking. The heat map in Figure 7 demonstrates how Democrats are much less packed after swapping and cracking is reduced, though some persists in suburban areas especially. Packing of Republicans has decreased in far western North Carolina, but cracking has increased slightly in the Charlotte and Raleigh areas.



**Figure 5: Percent of Democrats (left) and Republicans (right) packed and cracked in the original 2016 map relative to 1055 comparators. Packed individuals are shaded purple and cracked individuals are shaded orange (with hatches).**



**Figure 6: Partisan vote lean for the original 2016 map (left) and the new swapped map (right)**



**Figure 7: Percent of Democrats (left) and Republicans (right) packed and cracked in the computer generated witness relative to 1055 comparators. Packed individuals are shaded purple and cracked individuals are shaded orange (with hatches).**

Overall, there is a major net decrease in the total number of gerrymandered voters, demonstrating that the level of gerrymandering we observe in the 2016 districting of North Carolina is not a necessary result of the distribution of voters in the state. Our simulated annealing algorithm is able to produce witness districtings with total scores of less than 0.6 million based on only small perturbations to the original 2016 districting and maintaining comparable levels of population balance and compactness.

## 6 CONCLUSION AND FUTURE DIRECTIONS

Our paper provides an interpretable score for gerrymandering in terms of people impacted that allows one to precisely pinpoint which populations are gerrymandered and to what extent. We use this score along with an algorithmically generated witness to audit districtings for evidence of gerrymandering. We use our framework in an audit demonstrating substantial levels of partisan gerrymandering in the 2016 congressional districting of North Carolina.

In the policy sphere, we hope that others can use our approach to audit their own states' districtings. We also hope this can encourage the development of a justiciable standard for partisan gerrymandering legal cases, since it fulfills the courts' desire outlined in *Gill*

v. *Whitford* to have individual voters show standing by showing how they have been packed or cracked. This is a transition away from statewide metrics like the efficiency gap in the gerrymandering literature and towards an analysis of localized effects of gerrymandering.

With respect to the quantitative literature, this paper focuses on the auditing problem: Given a set of comparators, how can we quantify the level of gerrymandering observed and argue—via an algorithmically generated witness—that these levels are not necessary? Future work could focus on the algorithmic question of optimizing for a map with the *minimum* gerrymandering score using different approaches than simulated annealing.

We hope this work provides a new and more local way to think about gerrymandering that allows lawmakers, litigators, and academics to work toward a more fair redistricting process.

## REFERENCES

- [1] 1965. Voting Rights Act of 1965, Pub. L. 89-110, 79 Stat. 437.
- [2] Micah Altman. 1997. The Computational Complexity of Automated Redistricting: Is Automation the Answer. *Rutgers Computer & Technology Law Journal* 23, 1 (1997), 81–142. <https://heinonline.org/HOL/P?h=hein.journals/rutcomt23&i=87>

- [3] Theodore S. Arrington. 2016. A Practical Procedure for Detecting a Partisan Gerrymander. *Election Law Journal: Rules, Politics, and Policy* 15, 4 (Dec. 2016), 385–402. <http://doi.org/10.1089/elj.2016.0383>
- [4] Kevin Baas. 2016. *Auto-Redistrict*. Retrieved December 17, 2021 from <http://autoredistrict.org/contact.php>
- [5] 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 (Dec. 2021), 407–441. <http://doi.org/10.1089/elj.2020.0704>
- [6] Mira Bernstein and Moon Duchin. 2017. A Formula Goes to Court: Partisan Gerrymandering and the Efficiency Gap. *Notices of the American Mathematical Society* 64, 9 (May 2017), 1020–1024. <https://doi.org/10.1090/noti1573>
- [7] U.S. Census Bureau. 1994. *Voting Districts*. Retrieved May 1, 2022 from <https://www2.census.gov/geo/pdfs/reference/GARM/Ch14GARM.pdf>
- [8] Jowei Chen and Jonathan Rodden. 2013. Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quarterly Journal of Political Science* 8, 3 (2013), 239–269. <https://doi.org/10.1561/100.00012033>
- [9] Capital Broadcasting Company. 2016. *Editorial: Elect legislators who will enact non-partisan redistricting; revive judicial election reforms*. Retrieved December 17, 2021 from <https://www.wral.com/editorial-elect-legislators-who-will-enact-non-partisan-redistricting-revive-judicial-election-reforms/16108426/>
- [10] Chao Fan, Wenwen Li, Levi J. Wolf, and Soe W. Myint. 2015. A Spatiotemporal Compactness Pattern Analysis of Congressional Districts to Assess Partisan Gerrymandering: A Case Study with California and North Carolina. *Annals of the Association of American Geographers* 105, 4 (2015), 736–753. <https://doi.org/10.1080/00045608.2015.1039109>
- [11] Nikhil Garg, Wes Gurnee, David Rothschild, and David B. Shmoys. 2021. Combating Gerrymandering with Social Choice: the Design of Multi-member Districts. (2021). [arXiv:2107.07083](https://arxiv.org/abs/2107.07083)
- [12] Metric Geometry and Gerrymandering Group. 2021. *North Carolina Election Shapefile*. Retrieved December 17, 2021 from <https://github.com/mggg-states/NC-shapefiles>
- [13] Wes Gurnee and David B. Shmoys. 2021. Fairmandering: A column generation heuristic for fairness-optimized political districting. (2021). [arXiv:2103.11469](https://arxiv.org/abs/2103.11469)
- [14] 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. <https://doi.org/10.1080/2330443X.2020.1796400>
- [15] Wassily Hoeffding. 1963. Probability Inequalities for Sums of Bounded Random Variables. *J. Amer. Statist. Assoc.* 58, 301 (1963), 13–30. <https://doi.org/10.2307/2282952>
- [16] Lawrence Hurley. 2017. Supreme Court tosses Republican-drawn North Carolina voting districts. (May 2017). Retrieved December 17, 2021 from <https://www.reuters.com/article/us-usa-court-voters-idUSKBN1811SG>
- [17] David S. Johnson, Cecilia R. Aragon, Lyle A. McGeoch, and Catherine Schevon. 1989. Optimization by Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning. *Operations Research* 37, 6 (1989), 865–892. <https://doi.org/10.1287/opre.37.6.865>
- [18] J. Clark Kelso. 2004. *Vieth v. Jubelirer*: Judicial Review of Political Gerrymanders. *Election Law Journal: Rules, Politics, and Policy* 3, 1 (2004), 47–51. <https://doi.org/10.1089/153312904322739925>
- [19] Harry A. Levin and Sorelle A. Friedler. 2019. Automated Congressional Redistricting. *ACM J. Exp. Algorithmics* 24, 1, Article 1.10 (April 2019), 24 pages. <https://doi.org/10.1145/3316513>
- [20] The Supreme Court of the United States. 1962. *Baker v. Carr*, vol. 369. U.S.
- [21] The Supreme Court of the United States. 1964. *Wesberry v. Sanders*, vol. 376. U.S.
- [22] The Supreme Court of the United States. 2004. *Vieth v. Jubelirer*, vol. 541. U.S.
- [23] The Supreme Court of the United States. 2012. *Tennant v. Jefferson County Commission*, vol. 576. U.S.
- [24] The Supreme Court of the United States. 2017. *Cooper v. Harris*, vol. 581. U.S.
- [25] The Supreme Court of the United States. 2018. *Gill v. Whitford*, vol. 585. U.S.
- [26] The Supreme Court of the United States. 2019. *Rucho v. Common Cause*, vol. 588. U.S.
- [27] Brian Olson. 2013. *Redistricter*. Retrieved December 17, 2021 from <https://github.com/brianolson/redistricter>
- [28] Daniel D. Polsby and Robert D. Popper. 1991. The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering. *Yale Law & Policy Review* 9, 2, Article 6 (1991). <http://digitalcommons.law.yale.edu/ylpr/vol9/iss2/6>
- [29] Princeton Gerrymandering Project. 2021. *Redistricting Report Card Methodology*. Retrieved December 17, 2021 from <https://gerrymander.princeton.edu/redistricting-report-card-methodology>
- [30] Laura Royden, Michael Li, and Yuriy Rudensky. 2018. *Extreme Gerrymandering & the 2018 Midterm*. Retrieved November 5, 2021 from [https://www.brennancenter.org/sites/default/files/publications/extreme%20gerrymandering\\_2.pdf](https://www.brennancenter.org/sites/default/files/publications/extreme%20gerrymandering_2.pdf)
- [31] Nicholas Stephanopoulos and Eric McGhee. 2015. Partisan Gerrymandering and the Efficiency Gap. *University of Chicago Law Review* 82, 2, Article 4 (2015). <https://chicagounbound.uchicago.edu/uclrev/vol82/iss2/4>
- [32] Kristopher Tapp. 2019. Measuring Political Gerrymandering. *The American Mathematical Monthly* 126, 7 (2019), 593–609. <https://doi.org/10.1080/00029890.2019.1609324>
- [33] The New York Times. 2017. 2016 North Carolina Results. Retrieved January 8, 2022 from <https://www.nytimes.com/elections/2016/results/north-carolina>
- [34] Gregory Warrington. 2018. Quantifying Gerrymandering Using the Vote Distribution. *Election Law Journal: Rules, Politics, and Policy* 17, 1 (March 2018), 39–57. <http://doi.org/10.1089/elj.2017.0447>
- [35] Gregory S. Warrington. 2018. Packed voters and cracked voters. [arXiv:1806.11074](https://arxiv.org/abs/1806.11074) [physics.soc-ph]