Appendix

A NP-Hardness Proofs

Theorem 2. LF AUDITING *is NP-complete.*

Proof. We reduce the NP-complete Connected k-Subgraph Problem on Planar Graphs with Binary Weights (CkS-PB) [25] to the LF AUDITING problem. Given a connected planar graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ where $\mathsf{V} = \{\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_{|\mathsf{V}|}\}$, a vertex weight $\omega(\mathsf{v}_i) \in \{0, 1\}$ for each $\mathsf{v}_i \in \mathsf{V}$, a size $M \geq 2$, and a targeted total weight $\Omega \in \mathbb{Z}^+$,[10] the decision version of CkS-PB asks whether there is a subset $\mathsf{W} \subseteq \mathsf{V}$ of M vertices such that its induced subgraph $\mathsf{H} \subseteq \mathsf{G}$ is connected and $\sum_{\mathsf{v}_i \in \mathsf{W}} \omega(\mathsf{v}_i) \geq \Omega$. Here it is assumed that $M \leq |\mathsf{V}|$ and $\omega(\mathsf{v}_i) < \Omega$ for each $\mathsf{v}_i \in \mathsf{V}$; otherwise the problem is trivial.

Given an arbitrary instance of CkS-PB, we construct an instance for LF AUDITING as follows. For each vertex $\mathsf{v}_i \in \mathsf{V}$, we construct two precinct nodes v_i and u_i, and an edge between v_i and u_i. For each edge $(\mathsf{v}_i, \mathsf{v}_j) \in \mathsf{E}$, we construct an edge between v_i and v_j. Hence, the resulting graph $G = (V, E)$ has $2\,|\mathsf{V}|$ precinct nodes and $(|\mathsf{V}| + |\mathsf{E}|)$ edges, and is still planar.

For all $i = 1, 2, \ldots, |\mathsf{V}|$, we let $\rho(v_i) = \tau(v_i) = 2M\Omega$ and $\gamma(v_i) = \left(\frac{M-1}{2M} + \frac{\omega(\mathsf{v}_i)}{2\Omega}\right)$; note that each $\gamma(v_i)$ is in $(0, 1)$ and each $\gamma(v_i) \cdot \tau(v_i)$ is an integer. For all i, we let $\rho(u_i) = \tau(u_i) = 2M(M-1)\Omega$, and $\gamma(u_i) = 0$. We call the precinct nodes v_i *regular* and the precinct nodes u_i *auxiliary*. Let $k = |\mathsf{V}|$, $c = 1/2$, and pick ε such that $0 < \varepsilon < \frac{1}{M}$. Finally, let $\Pi = \{D_1, D_2, \ldots, D_k\}$, where $D_i = \{v_i, u_i\}$ for all $i = 1, 2, \ldots, k$. Observe that we have $\beta(D_i) > \frac{1}{2}$ for all $i = 1, 2, \ldots, k$, i.e., we have $B_\Pi = V$ and $R_\Pi = \varnothing$. Note that since $0 < \varepsilon < \frac{1}{M}$, we have $(2M^2 - 2M)\Omega < (1 - \varepsilon)2M^2\Omega$ and $(2M^2 + 2M)\Omega > (1 + \varepsilon)2M^2\Omega$, and thus every feasible district has total population exactly $2M^2\Omega$.

Suppose the CkS-PB instance is a `Yes` instance, i.e., there is a subset $\mathsf{W} \subseteq \mathsf{V}$ of M vertices such that its induced subgraph $\mathsf{H} \subseteq \mathsf{G}$ is connected, and $\sum_{\mathsf{v}_i \in \mathsf{W}} \omega(\mathsf{v}_i) \geq \Omega$. Let $W = \{v_i \mid \mathsf{v}_i \in \mathsf{W}\}$ be the set of regular precinct nodes corresponding to the vertices in W. Then we have $|W| = M$ and thus $\rho(W) = 2M^2\Omega$. Furthermore, we have

$$\begin{aligned}\sum_{v \in W \cap B_\Pi} \gamma(v)\tau(v) = \sum_{v \in W} \gamma(v)\rho(v) &= \sum_{i:\mathsf{v}_i \in \mathsf{W}} \left(\left(\frac{M-1}{2M} + \frac{\omega(\mathsf{v}_i)}{2\Omega}\right) \cdot 2M\Omega\right) \\ &\geq \left(\frac{M-1}{2} + \frac{\Omega}{2\Omega}\right) \cdot 2M\Omega = M^2\Omega = c \cdot \tau(W).\end{aligned}$$

Since W induces a connected subgraph of G, it is a red c-deviating group of Π.

For the other direction, suppose the LF AUDITING instance is a `Yes` instance, i.e., there is a red c-deviating group W of Π. Suppose W contains at least one auxiliary precinct node u_i. Recall that $\rho(W) = 2M^2\Omega$. Hence, we have

$$\sum_{v \in W \cap B_\Pi} \gamma(v)\tau(v) = \sum_{v \in W} \gamma(v)\rho(v) < \sum_{v \in W \setminus \{u_i\}} \rho(v) = 2M\Omega \leq M^2\Omega = c \cdot \tau(W),$$

a contradiction. Hence W can only contain regular precinct nodes. Then we have $|\mathsf{W}| = M$, and

$$\begin{aligned}\sum_{\mathsf{v}_i \in \mathsf{W}} \omega(\mathsf{v}_i) &= \sum_{v \in W \cap B_\Pi} \left(2\Omega\gamma(v) - \frac{(M-1)\Omega}{M}\right) \\ &= \frac{1}{M} \cdot \sum_{v \in W} \left(\gamma(v)\rho(v) - (M-1)\Omega\right) \\ &\geq \frac{\rho(W)}{M} - (M-1)\Omega = M\Omega - (M-1)\Omega = \Omega.\end{aligned}$$

Since W induces a connected subgraph of G, the corresponding CkS-PB instance is a `Yes` instance.

Combining the above, there is a polynomial-time reduction from CkS-PB to LF AUDITING. Since LF AUDITING is trivially in NP, we conclude that it is NP-complete. □

[10] We use M and Ω here to avoid confusing with the notations k and W in our problem.

We then observe that the same reduction also gives the hardness of LFP GENERATION.

Theorem 3. LFP GENERATION *is NP-complete.*

Proof. Observe that in each LF AUDITING instance constructed in the proof for Theorem 2, the associated plan Π is the only feasible redistricting plan. To see this, notice that every auxiliary node u_i has a single neighbor v_i and does not possess enough population to form a feasible district itself, and thus every u_i must be paired with its corresponding v_i to make a district. Therefore, the LF AUDITING and LFP GENERATION are identical on this family of instances, and the hardness of LFP GENERATION follows from the same reduction. □

Remarks. We note that the proofs above still hold even if we add more edges to the constructed graph. More specifically, for both the proof of Theorems 2 and 3, we can safely add edges as long as (i) planarity is preserved and (ii) at least one of the endpoints of each additional edge is an auxiliary precinct node. To see this, observe that adding additional edges does not impact the feasibility of Π, and since valid deviating groups contain only regular precinct nodes, the set of possible deviating groups for Π remains identical. For the proof of Theorem 3, the additional edges may allow multiple feasible redistricting plans, but each of the feasible redistricting plans must still contain k districts of one regular and one auxiliary precinct node each, and by the same reduction, either all or none of them are locally fair (corresponding to `Yes` and `No` CkS-PB instances).

Note further that in both LF AUDITING and LFP GENERATION, we do not explicitly require the districts and deviating groups to be compact with respect to any specific criteria. Under certain restrictive compactness constraints, the problems may become tractable. For example, if the districts and deviating groups are restricted to be subsets of precincts fully contained in a circle centered around a precinct point, then the set of possible districts and deviating groups has polynomial size, and thus LF AUDITING can be solved by enumeration in polynomial time.

B Speeding up the DP and Sufficiency of Ensemble-based Auditing

Although our dynamic programming algorithm for solving LF AUDITING on trees run in polynomial time, the time complexity of the algorithm is prohibitively high to be efficient in practice. Our goal in this section is to empirically demonstrate that this approach is not needed in practice, that is, it does not find reasonable deviating groups on plans that the ensemble-based method deems locally fair, hence showing that the ensemble-based auditing method is sufficient and obviating the need for the computationally inefficient dynamic programming.

Towards this end, we first show that the dynamic program can be sped up significantly if (among other things) we *interpolate* the voter information to the entire population of a precinct, so that $\tau(v) = \rho(v)$ for each precinct v. After performing such interpolation on the data used in our experiments (Section 4), we first run the ensemble-based auditing method to find fair and unfair plans for $c = 0.5$. Next, for each of these plans, we run the dynamic program to find deviating groups, checking each one for compactness and the value c for which it is a c-deviating group. We show that the dynamic program is unable to find compact deviating groups with $c \geq 0.52$ on the ensemble-audited 0.5-locally fair plans (on the interpolated data). This demonstrates the the sufficiency of ensemble-based auditing if we relax the strength c of the deviating group slightly.

We note that our main experiments in Section 4 use actual voter data, since it is unclear how such data should be interpolated in a principled way to the entire population. In the current section, we perform the interpolation in a simple way only to make the DP run efficiently, which in turn enables us to demonstrate the conceptual point that the ensemble-based method suffices. This provides strong evidence that even without interpolation, the ensemble-based method will suffice.

B.1 Improving Running Time of Dynamic Program

We first describe our approach to speed up the running time of the dynamic program.

Special case of $\tau(v) = \rho(v)$. We first assume that $\tau(v) = \rho(v)$ holds for all $v \in V$, i.e., every individual is labeled red or blue in every precinct. In this case, we can reduce the state space by dropping the state variable p, that is, we let $A[v, i]$ denote the maximum number of unhappy blue

voters in a subtree $W \subseteq T_v$ such that $v \in W$ and $\tau(W) = \rho(W) = i$. In this case, for leaf precincts v we have

$$A[v,i] = \begin{cases} \text{uhp}(v), & i = \tau(v) = \rho(v); \\ 0, & \text{otherwise,} \end{cases}$$

and for general precincts v (with children $\{u_1, \ldots, u_{\deg(v)}\}$) we have

$$A[v,i] = \text{uhp}(v) + B_{v,i}[\deg(v), p - \rho(v)], \tag{3}$$

where $B_{v,i}[1,x] = A[u_1,x]$ for all x, and for all $j \geq 2$ we have

$$B_{v,i}[j,x] = \max_{x' \in [0,x]} \{B_{v,i}[j-1, x-x'] + A[u_j, x']\}. \tag{4}$$

Now, the algorithm computes $O(|V| \cdot \sigma)$ values of $A[v,i]$, each computing $O(\deg(T) \cdot \sigma)$ values of $B_{v,i}[j,x]$, each requiring $O(\sigma)$-time to loop through all values of x'. The overall time complexity thus drops to $O\left(|V| \cdot \sigma^3 \cdot \deg(T)\right)$.

Relaxing size of deviation. We next modify the state $A[v,i]$ to be the maximum number of unhappy blue voters in a subtree $W \subseteq T_v$ such that $v \in W$ and $\tau(W) = \rho(W) \leq i$, i.e., it is now allowed that the subtree W has an aggregate population of less than i. Note that this induces a potential one-sided error in checking for the existence of deviating groups. To see this, consider the case when the algorithm finds some (v,i) such that $i \in [(1-\varepsilon)\sigma, (1+\varepsilon)\sigma]$ and $A[v,i] > i/2$. Now this corresponds to a subtree W rooted at v with a population (or voter count) of *at most* i and a total number of unhappy blue voters of *at least* $i/2$. While this still ensures a majority of voters in W are unhappy voters of the same color, the actual population size may be less than i and thus outside of the acceptable range $[(1-\varepsilon)\sigma, (1+\varepsilon)\sigma]$. However, we observe that this error is one-sided: Suppose there is indeed a deviating group W of the correct population size $i \in [(1-\varepsilon)\sigma, (1+\varepsilon)\sigma]$ with a total number of unhappy blue voters of *at least* $i/2$, our algorithm must either find W, or find a deviating group W' with population at most i and a larger number of unhappy blue voters. Therefore, if the algorithm does not find any deviating group under the relaxed definition, we can still conclude that there is no deviating group (with respect to the current spanning tree T).

Pruning the states. Under the modified semantics of $A[v,i]$, we observe that each $A[v,i]$ is non-decreasing in i and each $B_{v,i}[j,x]$ is non-decreasing in x. Now consider the computation of some fixed $B_{v,i}[j,x]$. We maintain an upper bound ub and a lower bound lb of $B_{v,i}[j,x]$. whenever $lb \geq ub$, we terminate the computation early and return $lb = ub$ as $B_{v,i}[j,x]$. Since the $B_{v,i}[j,x]$ are computed in increasing order of x, we initialize $lb = B_{v,i}[j, x-1]$ and let $lb = 0$ if $x = 0$. We also initialize $ub = B_{v,i}[j-1,x] + A[u_j, x]$.

When the $\max$ function in Eq. (4) is evaluated in increasing order of x', we update:

- $lb \leftarrow \max\{lb, B_{v,i}[j-1, x-x'] + A[u_j, x']\}$;
- $ub \leftarrow \min\{ub, B_{v,i}[j-1, x-x'] + A[u_j, x]\}$.

The second step is because that for any $x'' > x'$, we have

$$B_{v,i}[j-1, x-x''] + A[u_j, x''] \leq B_{v,i}[j-1, x-x'] + A[u_j, x].$$

Therefore, $B_{v,i}[j-1, x-x'] + A[u_j, x]$ is the maximum possible value of $B_{v,i}[j,x]$ if the function is maximized at any $x'' > x'$. If this is matched by lb, then the final maximum value will be exactly lb. In this case, we terminate the computation without examining any $x'' > x'$ in Eq. (4).

The same idea is applied to the computation of $A[v,i]$: We maintain a lower bound lb' for $A[v,i]$ (initialized to $A[v, i-1]$), and whenever

$$lb' \geq ub' = \text{uhp}(v) + \sum_{j=1}^{\deg(v)} A[u_j, i'],$$

we return $A[v,i] = lb'$.

Rounding the population. For a fixed threshold parameter P, we round each $\rho(v)$ down to the largest multiple of P that is smaller than or equal to $\rho(v)$. Formally, let $\rho'(v) = \left\lfloor \frac{\rho(v)}{P} \right\rfloor \cdot P$. For any subtree $W \subseteq T$, we have $\sum_{v \in W} \rho'(v) \leq \sum_{v \in W} \rho(v)$.

Let $A'[v, i]$ denote the output of the algorithm when the rounded population level $\rho'(v)$ is used instead of $\rho(v)$. Then we have $A'[v, i] \geq A[v, i]$. Therefore, running our algorithm with rounding introduces one-sided error: If there exists any deviating group W of Π of population level i, then the algorithm with rounding can also output W since it has the same number of unhappy blue voters with a rounded-down population level. We must then relax the acceptable size range from $[(1-\varepsilon)\sigma, (1+\varepsilon)\sigma]$ to $[(1-\varepsilon)\sigma - P \cdot |V|, (1+\varepsilon)\sigma]$ to accommodate this error and so that W becomes a candidate deviating group.

Final running time. With all these strategies incorporated, the running time of the dynamic program is reduced to $O\left(|V| \cdot (\frac{\sigma}{P})^3 \cdot \deg(T)\right)$ as there are now only $O\left(\frac{\sigma}{P}\right)$ population levels. This is the version of the algorithm that we implement.

B.2 Empirical Results and Sufficiency of Ensemble-based Auditing

We now use both our ensemble approach and the dynamic program to audit the ensemble for NC. We use the same experimental setup as in Section 4, except for one change. Since the DP assumes $\tau(v) = \rho(v)$, we need to interpolate the voter labels to the entire precinct. We do this in the natural way. We keep the same $\gamma(v)$ and $\beta(v)$ values determined from an election, but let $\tau(v) = \rho(v)$. The number of red and blue voters in a precinct v become $\gamma(v) \cdot \rho(v)$ and $\beta(v) \cdot \rho(v)$, respectively. This is equivalent to assuming that in each precinct v, the rate of red/blue preferences of non-voters is identical to that of the voters. Accordingly, a c-deviating group must have a c fraction of the total population being unhappy individuals of the same color.

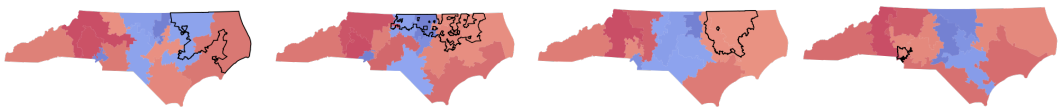
Using the interpolated voter labels on NC data, we first run the ensemble-based auditing method assuming $c = 0.5$. We find that 52 among the 1,000 plans (5.2%) in the ensemble do not have 0.5-deviating groups and are deemed fair by the ensemble approach. We again rank the plans in the ensemble by their unfairness score unf(Π). We then construct two groups of plans: (1) 26 Plans deemed 0.5-fair by the ensemble approach; (2) 10 Plans in the bottom 5% (most unfair, in terms of unf score) in the ranking. We generate 5 random spanning trees of the NC precinct graph. For each plan and each spanning tree, we run the dynamic program where the population rounding parameter is set as $P = 750$. For each group of plans (fair and unfair), we obtain the set of all deviating groups found by the dynamic program on any spanning tree. We measure the Polsby-Popper compactness score of each deviating group on the original graph and the *strength* c for which the group is a c-deviating group in that plan (the largest c for which the group is indeed deviating). Note that a larger value of c implies the deviation is robust to small population changes, and is more significant in terms of unfairness.

(a) Fair plans.

(b) Unfair plans.

Figure 4: Heatmaps of deviating groups found by the dynamic program on fair and unfair NC plans.

In Figures 4a and 4b, we plot the heatmaps of the deviating groups found by the dynamic program for the fair and unfair sets of plans, respectively, where the x-axis and y-axis demonstrate their Polsby-Popper score and their *strength* respectively. As shown, for the plans deemed 0.5-fair by the ensemble approach, most deviating groups are either not compact (having low Polsby-Popper scores



(a) A .512-deviating group with a .152 Polsby-Popper score ($D1$ in Fig. 4a).

(b) A .572-deviating group with a .020 Polsby-Popper score ($D2$ in Fig. 4a).

(c) A .593-deviating group with a .234 Polsby-Popper score ($D3$ in Fig. 4b).

(d) A .730-deviating group with a .161 Polsby-Popper score ($D4$ in Fig. 4b).

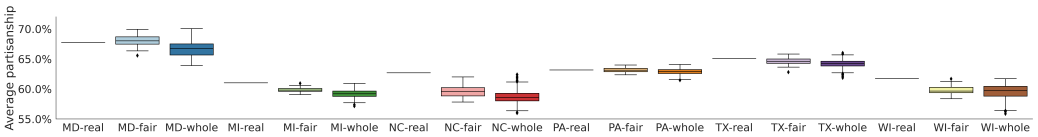
Figure 5: Deviating groups found by the dynamic program for two fair plans (left two maps) and two unfair plans (right two maps).

of < 0.1) or not strong (having strength values close to 0.5). As context, the minimum and average Polsby-Popper scores over all NC districts in the ReCom-generated NC ensemble are 0.053 and 0.177, respectively (corresponding to the two vertical lines in both plots); in other words, deviating groups with Polsby-Popper scores of < 0.053 (to the left of the dashed vertical line) are less compact than *every one* of the 10k districts in the ensemble.

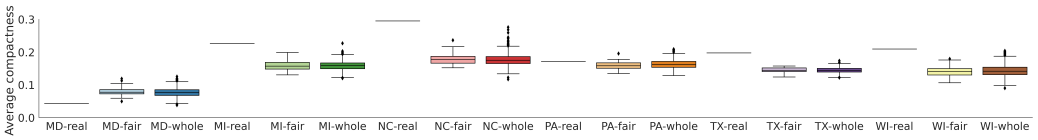
In fact, our dynamic program finds no deviating group with an above-average Polsby-Popper score for any 0.5-fair plan. The closest deviating group, shown as D1 in Figure 4a, has Polsby-Popper score 0.152 and a strength of 0.512; we visualize it in Figure 5a. In Figure 5b, we visualize another deviating group found by the dynamic program (shown as D2 in Figure 4a) with a Polsby-Popper score of 0.020 and a strength of 0.572. As manifested in the visualizations, deviating groups with very low Polsby-Popper scores like 0.02 are spurious with artificial shapes (such as holes) and should not be considered when it comes to determining whether a redistricting plan is fair. All the deviating groups for the 0.5-fair plans with a Polsby-Popper score at least 0.053 have lower strength (< 0.55). In summary, we can reasonably conclude that most of the fair plans found by the ensemble-based auditing approach do not admit strong, contiguous, and reasonably compact deviating groups even when audited by the dynamic program.

In contrast, for the plans ranked in the bottom 5% according to the ensemble-based approach, the dynamic program is able to find both strong and compact deviating groups quite easily. In Figures 5c and 5d we show (a) a .593-deviating group with a .234 Polsby-Popper score (D3 in Figure 4b), and (b) a .730-deviating group with a .161 Polsby-Popper score (D4 in Figure 4b) that the dynamic program find for one of the unfairest plans. These results show that the strengths of deviating groups for fair plans (according to ensemble based auditing) are considerably lower than that for unfair plans. In other words, the results via DP validates that via the ensemble based approach.

C Alternative Fairness and Compactness Metrics



(a) Average partisanship



(b) Average Polsby-Popper score

Figure 6: Distribution of alternative fairness and compactness metrics among subsets of generated plans.

Average partisanship. For each district, let its *partisanship* be the percentage of votes in the majority color. Therefore, a low partisanship (towards 50%) implies better competitiveness in that district. We define the average partisanship of Π to be the average of partisanship values over its districts (ignoring small differences in population) as an alternative to the competitiveness metric used in

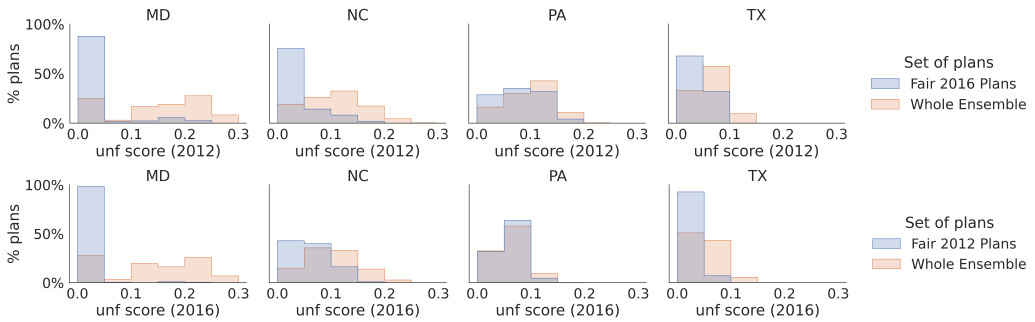


Figure 7: Histograms of unf scores. The top plots compute 0.5-fair maps using 2016 voter data, and the blue bars plot the histograms of unf scores when these plans are audited using 2012 voter data. The orange bars represent the histograms of unf scores of the entire ensemble when audited using 2012 data. The bottom plots switch the roles of 2016 and 2012, so the 0.5-fair plans are generated using 2012 data and audited using 2016 data.

Figure 3b. In Figure 6a, we compare the average partisanship among the three subsets of plans used in Section 4.3 (top-5% fairest plans, whole ensemble, and real enacted plans).

Results show that fair plans generate slightly more partisan districts. However, compared to the whole ensemble, the (roughly) 2-3% of shift in average partisanship is small and comparable to other uncertainties (e.g., voter turnouts or year-to-year election result gaps). Furthermore, the median average partisanship of the fair plans is smaller than that of the real-world redistricting plan for all but one state, showing that local fairness remains compatible with reasonably small partisanship.

Average compactness. We define the average compactness of Π to be the average of Polsby-Popper scores over its districts. In Figure 6b we compare the average compactness among the three subsets of plans. Similar to the results for minimum compactness, the average compactness of the locally fair plans remains comparable to that of the entire ensemble. On the other hand, the enacted plans perform better on average compactness than on minimum compactness, showing that enacted plans have larger variances in the compactness scores than plans in the ensemble.

D Robustness of Local Fairness to Voting Patterns

To test the robustness of the local fairness notion to changes in voting patterns, we repeat the ensemble-based audit process in Section 4 for MD, NC, PA, and TX with $\gamma(v)$, $\tau(v)$, and $\beta(v)$ values replaced by label values obtained from the 2012 presidential election. We do not consider MI (only a few plans are fair, so the sample size is too small) and WI (most plans are fair, and thus the ensemble and fair plans yield similar statistics).

For $c = 0.5$, we obtain the set of locally fair plans (i.e., plans without c-deviating groups) when audited using 2012 (resp. 2016) voter labels. We then compute the unfairness of these plans using the 2016 (resp. 2012) voter data, i.e., from the other election. We repeat this for all the plans in the ensemble, obtaining the unf score for each plan. We plot the histograms of these unf values in Figure 7, where the x-axis is the bucketed unf score, and the y-axis is the percentage of plans among the fair maps (resp. ensemble) that fall in that bucket.

For MD, NC, and TX, the blue bars are skewed significantly towards the left compared to the entire ensemble, showing that the fairest plans identified by auditing with a specific election remains significantly fairer compared to the entire ensemble when measured by another election. This shows the local fairness notion is fairly robust, or insensitive to year-to-year election result fluctuations. For PA, the fair plans are more sensitive to the specific election used, which reflects the role of PA as a swing state across elections.

E Visualization of Fair and Unfair Plans for Other States

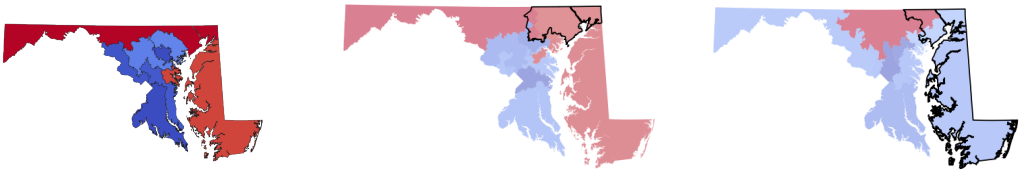
We show additional visualizations of fair plans and deviating groups found using ensemble-based auditing. As before, we show deviating groups with black outline, and the districts are coded by its color and the extent of partisanship: districts with a larger value of γ (resp. β) are colored in darker red (resp. blue). For each state (MD, MI, TX, WI), we show a fair plan (Figures 8a, 9a, 10a, 11a), and an example of a deviating group of each color. We discuss the deviating groups in each state.

In MD, the central blue districts tend to not be competitive, and the geography of the state contributes to the difficulty of forming red deviating groups. Thus MD is the only state where the precincts in the most deviating groups are not densely populated areas (see Figure 2a). Instead, the two "panhandles" (the western and coastal eastern regions of the state) tend to be part of deviating groups, see Figures 8b and 8c.

In Figures 9b and 9c we show a red and a blue deviating group in MI. Michigan had the lowest number of fair plans in the ensemble (see Table 2). The precincts in deviating groups are clustered in one region of MI, where the districting is sensitive to which precincts belong in red or blue districts. Both red and blue deviating groups are concentrated around this area.

In contrast, some deviating groups in TX concentrate around urban areas, while others intersect a large area of less densely populated precincts. The districting shown in Figure 10b shows a blue deviating group in proximity to an urban area. In contrast, Figure 10c shows a red deviating group on a districting plan intersecting a large blue district.

In Figure 11b, a blue deviating group of WI pulls in a portion of an urban area (left) and spans across to another blue district. In the fair plan, the urban counties are contained fully in a blue district, while the middle red precincts that deviate in Figure 11c are in their own red district.

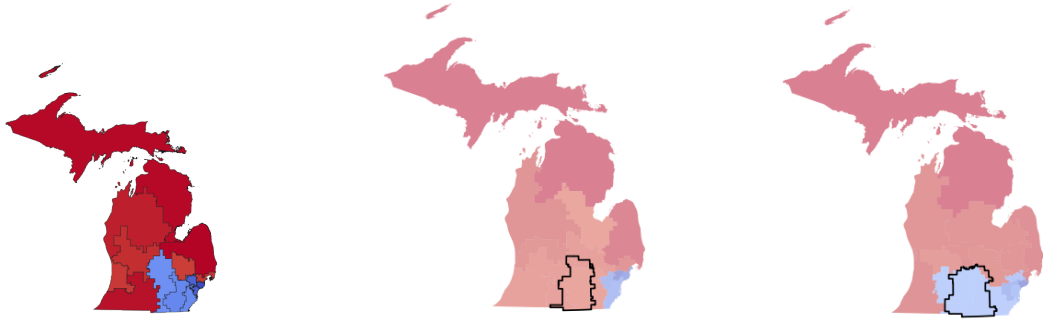


(a) MD fair plan

(b) MD plan with a blue deviating group

(c) MD plan with a red deviating group

Figure 8: Maryland plans without and with deviating groups



(a) MI fair plan

(b) MI plan with a blue deviating group

(c) MI plan with a red deviating group

Figure 9: Michigan plans without and with deviating groups