An Online Automated Redistricting Simulator

Rutvik Patel*†

1 Introduction

The assignment of U.S. Congressional districts has a large impact on the outcomes of House elections, because in general voting patterns can be predicted based on location. With recent increases in computing power, mathematical and computational redistricting has come to replace the assignment of districts by hand. Research on this new form of redistricting, which has been around for over half a century, initially focused on integer programming and graph-theoretic methods. For example, Hess et al. introduced an integer programming formulation based on compartmentalization heuristics [13], Garfinkel and Nemhauser [10] implemented an enumerative method to find an optimal solution to small instances of the problem, Bodin [3] used a graph partitioning heuristic to quickly find non-optimal solutions, and Plane[18] and Birge[2] proposed quadratic programming formulations. However, Altman[1] mentions several redistricting subproblems that are NP-complete, such as finding equal-population or competitive districts, and therefore any redistricting problem based on some of those goals must be computationally hard. Even so, there was a generation of papers [11, 14, 17] that sought to refine the original integer programming and graph-based methods. More

*Department of Computer Science, Rice University, Houston, TX 77005, USA
†Mitch Belmer, Andrew Hadad, and Robbie Foley also worked on an earlier version of this project.

recently, researchers have explored newer approaches that can be classified as local search metaheuristic procedures. These include tabu search [4], simulated annealing [5], Markov chain Monte Carlo or MCMC [8, 12, 16], techniques from statistical physics [6, 7], genetic algorithms [9], and the application of Voronoi diagrams[19]. The third of these methods, MCMC, was used in this project to generate large collections of maps that can be viewed and filtered based on relative priorities of fairness, competitiveness, and compactness constraints, with an online tool located at https://rice-gerry.surge.sh.

2 Methodology

MCMC algorithms are used to approximate probability distributions when exhaustively computing them is infeasible. The idea is to start with a seed state and iteratively perturb it to produce a chain of states that eventually converges to the target distribution. MCMC methods seem well-placed to handle the challenges of automated districting, because the number of valid district plans for even the smallest states is astronomically high, even when constraints like population parity and compactness are enforced. We used an MCMC algorithm to generate an ensemble of district maps for New Hampshire and Texas, as well as developed the infrastructure needed to do the same for any state, and created a web tool to allow for visualizing and filtering the maps in these ensembles.

For a given state, we model it using a graph in which the nodes are VTDs (voting tabulation districts, which represent an idealized notion of a precinct) and the edges signify geographical adjacency. The two most common ways to determine adjacency are rook and queen, which respectively check if two polygons share a side or a point. We prefer to use rook adjacency, as it is more closely aligned with what we, and presumably many judges, intuitively think of as adjacency. Furthermore, we associate to each VTD a district number assignment, and we consider a map contiguous if for each district, the subgraph induced by the vertices associated with that district is connected. Each VTD also has population, geographical, and House election data associated with it.

Our algorithm starts with an existing district map (for New Hampshire, approximately the one used for the 2012 elections; for Texas, a randomly generated one that satisfies contiguity and 1% population parity, except for a single outlier district within 10% of the ideal population) and generates proposals using the ReCom scheme[16]. This algorithm randomly selects two neighboring districts and merges them, generates a spanning tree for the merged unit, and cuts a random tree edge that would leave two new districts within 1% of the ideal population. If the proposed map violates contiguity or a population parity constraint (1% for New Hampshire; 10% for Texas to account for the seed map's outlier, even though most of the districts in most of the generated maps were within 1% of the ideal population), we immediately reject it. Otherwise, we accept it. We ran the New Hampshire chain until 5000 maps were accepted, and the Texas chain until 2000 maps were accepted.

Next, we compute fairness, unfairness, competitiveness, and compactness scores for each generated map. The fairness score is simply $1-|d-d'|$, where d is the proportion of the House seats that would be won by Democrats in a hypothetical election based on the election data associated with each VTD, and d' is the proportion of the voters in the state that vote Democratic. Similarly, the unfairness score (in favor of Democrats) is $(d-d'+1)/2$. The competitiveness score is $1-2\sqrt{\sum_{i=1}^{n}(d_i-0.5)^2}/\sqrt{n}$, where n is the number of House seats in the state and d_i is the proportion of the voters in district i that vote Democratic. Although other metrics for competitiveness exist [see 8], this particular one was chosen for its simplicity, orthogonality to the fairness score, and emphasis on non-competitive districts. The compactness score is $\sum_{i=1}^{n} P(i)/n$, where n is the number of House seats in the state and $P(i)$ is the Polsby-Popper score for district i, i.e. $P(i) = 4\pi A(i)/p(i)^2$ where

$A(i)$ and $p(i)$ are respectively the area and perimeter of district i. Many other compactness measures exist, but the Polsby-Popper score was chosen for its historical significance and adherence to visual intuition. Note that each of these scores has a minimum possible value of 0 and a maximum possible value of 1.

Finally, for a given state, we sort the generated maps for that state based on $0-5$ values for fairness, competitiveness, and compactness constraints provided by the user of the application to indicate his or her relative priorities. The filtering process computes a scalar value for each map by summing normalized versions of the fairness, competitiveness, and compactness scores, weighted by the user-provided parameters. It then performs a sort to find the map with the greatest value for this scalar and shows it to the user, along with some statistics and graphs. The normalization procedure simply linearly transforms the distribution of a score so that 0 and 1 respectively become the minimum and maximum values of the score among the generated maps. The user can also specify that the fairness constraint be interpreted as unfairness in favor of either the Democrats or Republicans, in which case either the unfairness score or 1 minus the unfairness score is used, respectively, instead of the fairness score.

3 Results

Figures 1 and 2 display the seed maps for New Hampshire and Texas, and Figures 3 and 4 show the most compact map generated for each state. The usefulness of the compactness constraint is exemplified in the second New Hampshire map, which contains a roughly circular district. In general, the ReCom proposal tends to produce fairly compact maps, which is useful because it eliminates the need to favor high compactness during generation, and because non-compact maps are unlikely to be instated. Figures 5 and 7

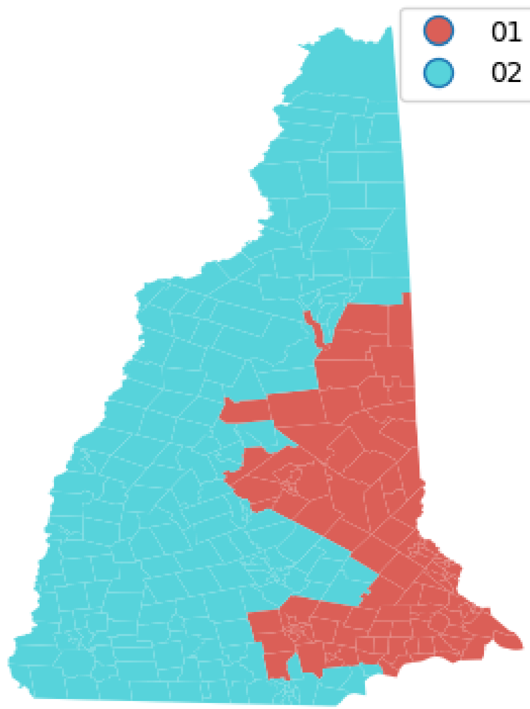

Figure 1: New Hampshire seed map

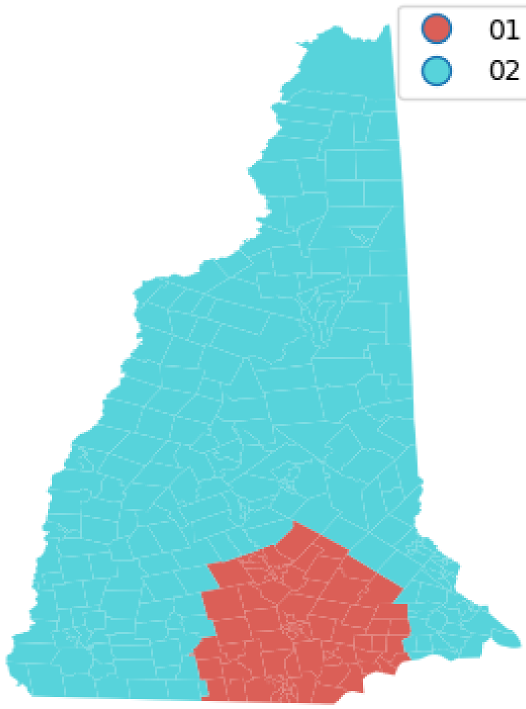

Figure 2: Most compact New Hampshire map generated

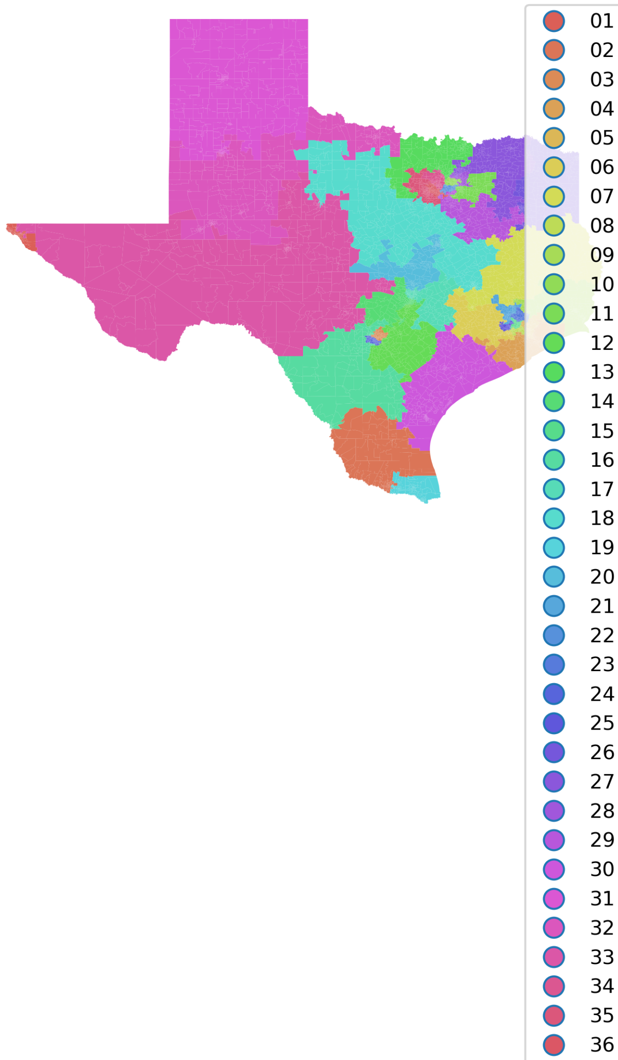

Figure 3: Texas seed map

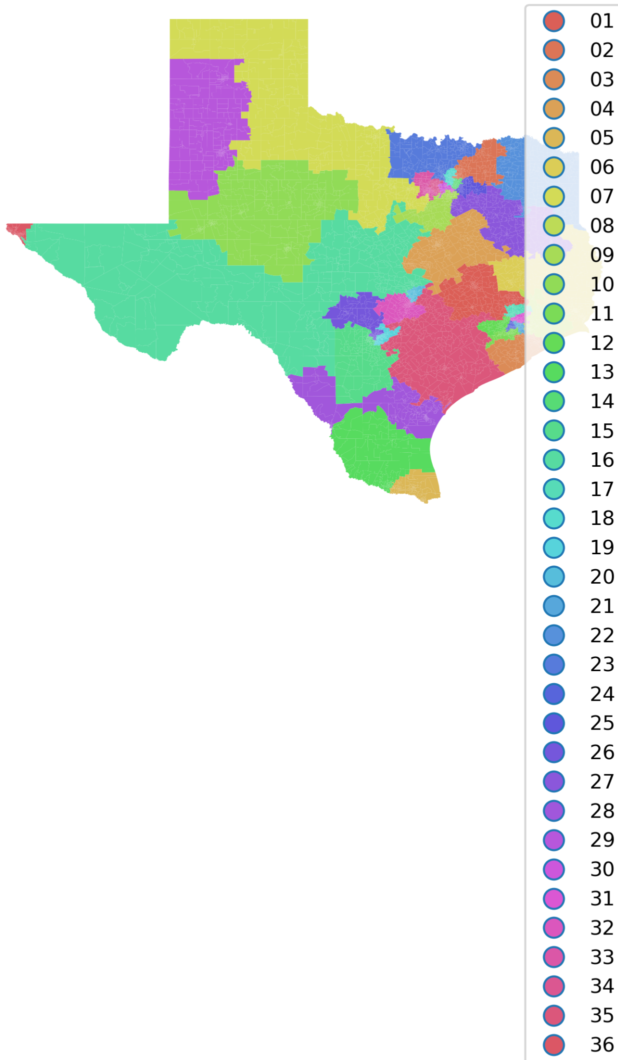

Figure 4: Most compact Texas map generated

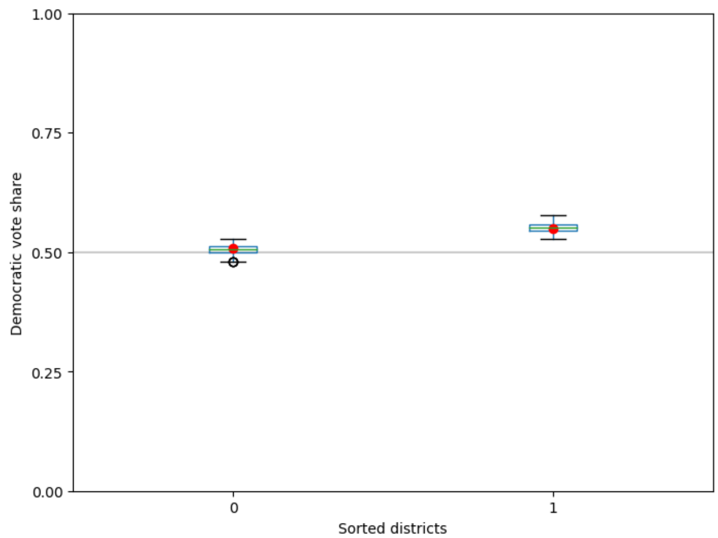

Figure 5: Distribution graph for the New Hampshire seed map

graph the per-district distributions of Democratic vote share across the New Hampshire and Texas ensembles, and indicate in red where the seed maps lie. Curiously, the generated maps for Texas have strong Republican biases, and the actual number of Democratic House seats in the 113th Congress was never encountered. Figures 6 and 8 graph the distributions (kernel density estimation plots) of the fairness, competitiveness, and compactness scores across the New Hampshire and Texas ensembles, and indicate with dotted lines where the fairest maps lie. Interestingly, the fairest maps were not commonly generated.

4 Conclusion

This project used MCMC methods to generate large numbers of valid district maps for New Hampshire and Texas, and introduced a unique web tool allowing users to visualize and filter those maps based on relative priorities of fairness, competitiveness, and compactness constraints. Experiments with

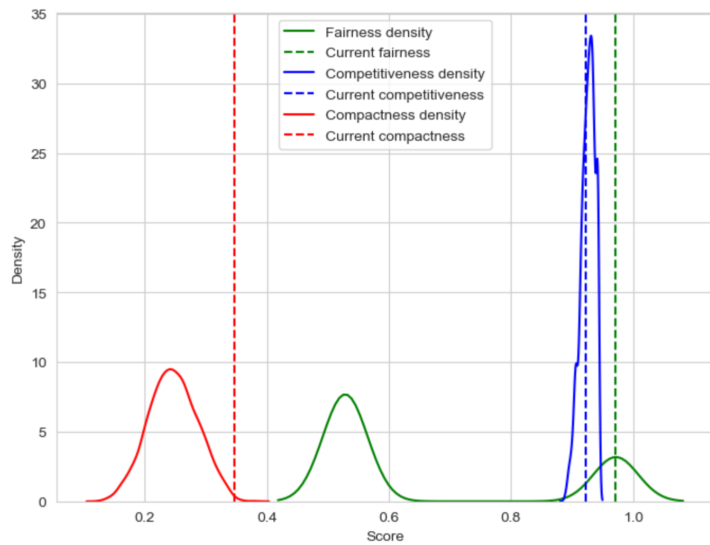

Figure 6: Distribution graph for the fairest New Hampshire map generated

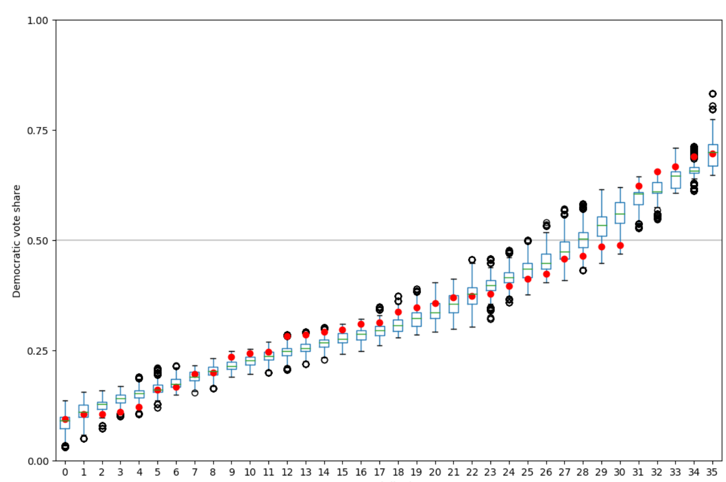

Figure 7: Distribution graph for the Texas seed map

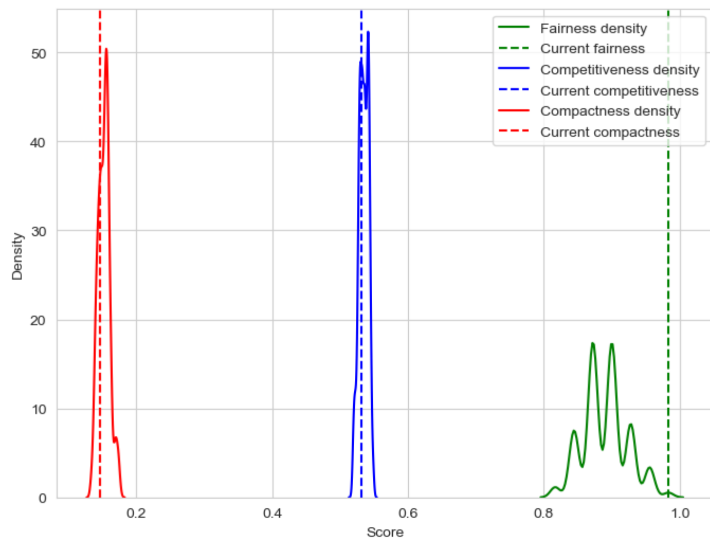

Figure 8: Distribution graph for the fairest Texas map generated

this tool show that there are trade-offs between these constraints (especially between fairness and competitiveness), and that fair maps are not necessarily the most common maps generated. One major direction for future work is to make the web tool more dynamic. In particular, it is conceivable that the ability to run MCMC trials on-demand and compare the results to existing runs could offer confidence in the true distribution of district maps for a given state. Considering that properties like fairness and competitiveness are not easily visible from actual maps, the ability to filter map ensembles and see where the query results lie on approximated distributions for those properties is critical. Another direction for future work involves generating more maps and taking into account factors like minority populations, in the spirit of the Voting Rights Act. Finally, it is of interest to extend the web tool to support more states, ideally all states with more than one House seat.

References

1. M. Altman. The computational complexity of automated redistricting: Is automation the answer? *Rutgers Computer and Law Technology Journal*, 23:81, 1997.

2. J. R. Birge. Redistricting to maximize the preservation of political boundaries. *Social Science Research*, 12(3):205-214, 1983.

3. L. D. Bodin. A districting experiment with a clustering algorithm. *Annals of the New York Academy of Sciences*, 219(1):209-214, 1973.

4. B. Bozkaya, E. Erkut, and G. Laporte. A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, 144(1):12-26, 2003.

5. M. H. Browdy. Simulated annealing: An improved computer model for political redistricting. *Yale Law & Policy Review*, 8(1):163-179, 1990.

6. C.-I. Chou and S.-P. Li. Statistical physics approach to political districting problem. *International Journal of Modern Physics B*, 21(23):3975-3980, 2007.

7. C.-I. Chou and S.-P. Li. Taming the gerrymander - statistical physics approach to political districting problem. *Physica A: Statistical Mechanics and its Applications*, 369(2):799-808, 2006.

8. B. Fifield, M. Higgins, K. Imai, and A. Tarr. A new automated redistricting simulator using Markov chain Monte Carlo. Technical report, available at https://imai.fas.harvard.edu/research/files/redist.pdf, 2018.

9. S. L. Forman and Y. Yue. Congressional districting using a TSP-based genetic algorithm. *Genetic and Evolutionary Computation Conference*, 2072-2083, Springer, 2003.

10. R. S. Garfinkel and G. L. Nemhauser. Optimal political districting by implicit enumeration techniques. *Management Science*, 16(8):B-495, 1970.

11. J. A. George, B. W. Lamar, and C. A. Wallace. Political district determination using large-scale network optimization. *Socio-Economic Planning Sciences*, 31(1):11-28, 1997.

12. G. Herschlag, H. S. Kang, J. Luo, C. V. Graves, S. Bangia, R. Ravier, and J. C. Mattingly. Quantifying Gerrymandering in North Carolina. *arXiv:1801.03783*, 2018.

13. S. Hess, J. Weaver, H. Siegfeldt, J. Whelan, and P. Zitlau. Nonpartisan political redistricting by computer. *Operations Research*, 13(6):998-1006, 1965.

14. M. Hojati. Optimal political districting. *Computers & Operations Research*, 23(12):1147-1161, 1996.

15. Y. Y. Liu, W. K. Tam Cho, and S. Wang. PEAR: A massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation*, 30:78-92, 2016.

16. Metric Geometry and Gerrymandering Group. Comparison of districting plans for the Virginia House of Delegates. Technical report, available at https://mggg.org/VA-report.pdf, 2018.

17. S. K. Norman and J. D. Camm. The Kentucky redistricting problem: Mixed-integer programming model. Technical report, 2003.

18. D. A. Plane. Redistricting reformulated: A maximum interaction/minimum separation objective. *Socio-Economic Planning Sciences*, 16(6):241-244, 1982.

19. L. Svec, S. Burden, and A. Dilley. Applying Voronoi diagrams to the redistricting problem. *UMAP Journal*, 28(3):313-329, 2007.