Evaluating Automated Redistricting Algorithms Using Measures of Compactness & Partisan Fairness: A Case Study of 2021 Congressional Redistricting in Virginia

Madeleine Goertz

International Community School

AP Research

Randall S. Huberman

May 20th, 2021

using input not include so that starts on same page

Evaluating Automated Redistricting Algorithms Using Measures of Compactness & Partisan Fairness: A Case Study of 2021 Congressional Redistricting in Virginia

Legislative redistricting is the process of drawing district boundaries for elections. In the United States, the redrawing of congressional district boundaries occurs every 10 years after the census. This process of redistricting is generally performed by state legislatures, hybrid commissions, or independent commissions (Project, n.d.).

When redistricting congressional districts, the two most fundamental requirements are that the districts must be "reasonably compact" and of equal population, meaning that the population of each district must be within

$$\pm$$

1% of the average district population ("Wesberry v. Sanders," 1964).

Gerrymandering is drawing the district boundaries in such a way as to unfairly benefit a particular interest group. There are two primary methods of gerrymandering: "packing" and "cracking." The US Supreme Court defines these as creating districts where the opposing party holds far above the majority or is spread out amongst so many districts as to become ineffectual, respectively ("Davis v. Bandemer," 1986).

Automated redistricting algorithms are one proposed solution to prevent gerrymandering. As early as 1961, such "automatic and impersonal procedure[s]" were thought to "produce results not markedly inferior to those which would be arrived at by a genuinely disinterested commission" (Vickrey, 1961, p. 110).

Broadly speaking, the goal of such algorithms is to create a map (or a selection of maps) that is "fair." Some algorithms obey only equal population and compactness constraints (e.g. Altman & Mcdonald, 2009; Chen & Rodden, 2013; Hu et al., 1995), while more-modern algorithms have evolved to address real-world considerations, such as

majority-minority districts, incumbency, and competitiveness (e.g Fifield, Higgins, et al., 2020; Lara-Caballero et al., 2019; McCartan & Imai, 2020).

In parallel to the development of these algorithms, measures of compactness and of partisan fairness were also developed. These compactness measures include geometric measures (e.g. Boyce & Clark, 1964; Harris, 1964; Maceachren, 1985; Polsby & Popper, 1991; Reock, 1961; Schwartzberg, 1966) and measures from graph theory (e.g. Dube & Clark, 2016). Proposed measures of partisan fairness generally try to characterize the relationship between the number of votes and the number of seats won by a political party. Some measures seek to simulate the power balance in the legislature at all possible party votes shares (e.g. Tufte, 1973), while others aim to compute a single metric (e.g. Katz et al., 2020; McDonald & Best, 2015; Stephanopoulos & McGhee, 2014; Wang, 2016; Warrington, 2018). Neither the scholarly literature nor the courts agree on a single correct compactness or partisan fairness measure.

While the scholarship is focused on the development of new automated redistricting algorithms (Fifield, Higgins, et al., 2020), on the evaluation of the mathematical rigor of proposed metrics (see Katz et al., 2020), and on the detection of gerrymandering in existing redistricting plans (e.g. Duchin, 2018; Herschlag et al., 2017), research into the empirical performance of these algorithms in a real-world redistricting scenario is lacking (Fifield, 2020).

Identifying this gap led me to my research question, *How do the hypothetical district maps for the Virginia Congressional delegation for the 2020s generated by different automated redistricting algorithms compare based on partisan fairness and compactness measures?*

I chose Virginia as my case state because (a) it has had to redistrict mid-decade due to gerrymandering (see "Wittman v. Personhuballah," 2016), (b) it is a competitive state from an electoral perspective, (c) it recently approved a constitutional amendment that institutes a redistricting process driven by a hybrid commission (Staff, 2020), and (d) it has

a complex geography.

With the goal of comparing the performance of different automated redistricting algorithms, I chose two algorithms of different levels of complexity that could both meet the real-world requirements of equal population and compactness. The Compact Random Seed Growth algorithm (CRSG) (Chen & Rodden, 2013) and the Sequential Monte Carlo algorithm (SMC) (McCartan & Imai, 2020) were chosen as they both satisfy the aforementioned requirement, are implemented by an open-source software package (Fifield, Kenny, et al., 2020), and represent the evolution in the scholarship on such algorithms.

Since there is no single measure of partisan fairness nor compactness agreed upon by the scholarship, I selected representative samples of the different perspectives on each measure. For compactness, one geometry-based measure (the Polsby-Popper score (Polsby & Popper, 1991)) and one graph-theory-based measure (the edge-cut compactness score (Dube & Clark, 2016)) were chosen. For partisan fairness, the simulated election measure (the seats-votes curve (Tufte, 1973)) and four single-valued fairness measures (partisan bias (Katz et al., 2020), the efficiency gap (Stephanopoulos & McGhee, 2014), declination (Warrington, 2018), and the mean-median difference (McDonald & Best, 2015)) were chosen. Together, these five partisan fairness measures represent differing and disputed views on how "fairness" should be quantified (Katz et al., 2020).

I have chosen to conduct a case-study using the experimental research method, as it allows me to infer the causality of the different redistricting algorithms on the redistricting plan outcomes.

This paper begins with a literature review of the scholarship on automated redistricting algorithms, compactness measures, and partisan fairness measures. Further details of my chosen research method are explained in the Method section. The results of my analysis are presented, interpreted, and reflected upon in the Results, Discussion, and Conclusion sections, respectively.

Literature Review

I will provide an overview of some of the current research into automated redistricting algorithms and compactness and partisan fairness measures.

Automated Redistricting Algorithms

The purpose of automated redistricting algorithms is to generate a set of redistricting plans that are as "impartial" as possible (Chen & Rodden, 2013). While many different algorithms have been proposed (see Altman and Mcdonald (2009), **haas2020**, Lara-Caballero et al. (2019), Macmillan (2001), Weaver and Hess (1963), and Xiao (2008)), I will provide high-level overviews of two algorithms: Sequential Monte Carlo and Compact Random Seed Growth.

Redistricting as Graph Cutting

The following redistricting algorithms all conceptualize redistricting precincts as a graph-cutting problem. A graph is a network of different interconnected points, where the points are "nodes" and the lines connecting them "edges" (Fifield, Higgins, et al., 2020). Every precinct is a node, and geographically-adjacent precincts have their corresponding nodes connected by edges. This is known as the "adjacency graph." Since the goal of redistricting is to assign every precinct a district, the algorithms imagine that edges between nodes are "cut" until "islands" (known as "subgraphs") are formed, where each is disconnected from the others. The disconnected subgraphs then become the districts. See Figure 1 for a visualization (Fifield, Higgins, et al., 2020).

Sequential Monte Carlo

SMC is an implementation of the statistical method Sequential Monte Carlo[1] (McCartan & Imai, 2020). The following overview provides a high-level understanding of the basic SMC algorithm.

SMC views redistricting as a graph-cutting problem, but instead of the adjacency

---

[1] I use SMC to refer to the algorithm, not the statistical method
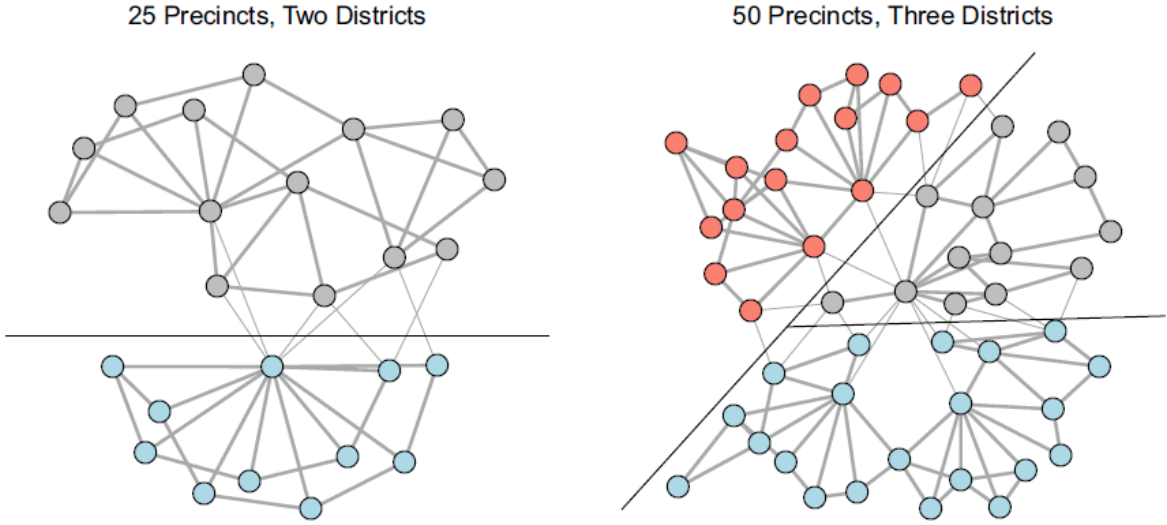
Figure 1

Redistricting as graph cutting. Every node is a precinct, and nodes that share an edge are known to be adjacent precincts. The algorithms "cut" edges between nodes until islands of districts are formed. (Fifield, Higgins, et al., 2020, p. 3)

graph, it begins with a spanning tree, a graph with the minimum number of edges to connect every node. If any edge is removed, 2 subgraphs are formed.

Figure 2 visualizes the iterative splitting procedure used by SMC. First, it compute the deviation from the target precinct population for each subgraph generated by cutting each edge in the spanning tree. Then, it randomly cuts one of the edges, creating two subgraphs. If the smaller subgraph meets the population and compactness requirements, then it's accepted as the first district, and the splitting procedure is repeated with the other subgraph.

This process generates the possible redistricting plans that satisfy the requirements. For more details, please refer to McCartan and Imai (2020).

Compact Random Seed Growth

The following is a high-level overview of the CRSG algorithm. For more details, please refer to Chen and Rodden (2013, pp. 249–50).

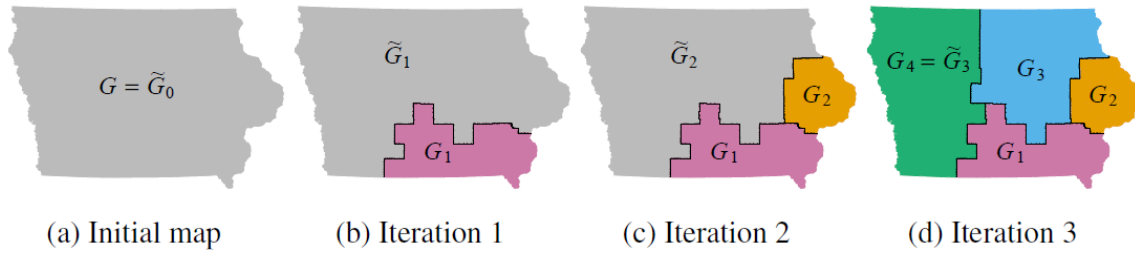| (a) Initial map | (b) Iteration 1 | (c) Iteration 2 | (d) Iteration 3 |

Figure 2

Splitting procedure of SMC. One edge is chosen to be cut probablistically, and the smaller resulting subgraph becomes a district if it meets requirements. The procedure is repeated with the remaining subgraph until the desired number of districts are formed. (McCartan & Imai, 2020, p. 14)

CRSG begins with declaring that every precinct is it's own district. A random precinct is then chosen, and then its geographically-closest neighbor is merged with it, creating one fewer district. This process is repeated until it arrives at the desired number of districts. (Chen & Rodden, 2013, pp. 249–50)

After this procedure, the districts are somewhat compact due to the geographic proximity requirement, but there is no guarantee that the districts are within the required population percentage of each other. To satisfy the population parity requirements, CRSG does the following. First, it identifies the two adjacent districts that have the greatest difference in total population. Then the precinct in the more populous district that is furthest from the center of said district is reassigned to the less populous district, provided that this reassignment doesn't break either district into parts. This process is repeated until all of the districts are within some desired percentage of the mean district population.Chen and Rodden (2013, pp. 249–50).

One run of CRSG will produce one set of districts, and separate runs of CRSG with the same input data may produce slightly different districts.

Compactness Measures

A universal requirement of districts is that they be "reasonably compact." There is no single legal nor scholarly definition of compactness (Katz et al., 2020). The following is a brief overview of several different, commonly used compactness measures that vary in approach, benefits, and weaknesses.

Polsby-Popper Score

Polsby and Popper (1991) introduces the Polsby-Popper score first developed in paleontology to the problem of gerrymandering. It computes the ratio of the area of the district to the area of a circle with the same perimeter as the district (Cox, 1927; Polsby & Popper, 1991). It ranges from 0 to 1, where 1 is most compact (Polsby & Popper, 1991).

Limitations and criticisms of this measure include that it is very sensitive to both geography and map resolution. Particularly near coastlines, even the most compact districts can have Polsby-Popper scores that are lower than gerrymandered districts which don't border coastlines. At finer resolutions, the same district will have a lower Polsby-Popper score as the perimeter increases. (McCartan & Imai, 2020, p. 12).

Edge-Cut Compactness

Edge-Cut Compactness takes a graph theory perspective to district compactness. The edge-cut compactness score is the number of edges that must be "cut" from the original adjacency graph to the subgraphs (districts) (Dube & Clark, 2016). Theoretically, more compact districts will require fewer edges to be cut (Dube & Clark, 2016). When normalized to the number of edges and subtracted from 1, a score of 1 indicates the most compact district.

Benefits of this measure include that it is unaffected by political and natural geography, map resolution, and population density (McCartan & Imai, 2020, p. 11).

Partisan Fairness Measures

The following is an overview of the partisan fairness measures in literature, as well as their merits and disadvantages.

Seats-Votes Curve

Seats-votes curves have been used to assess district plans for more than 40 years, and they are broadly supported by the literature (Katz et al., 2020). Seats-votes curves plot the relationship between the population vote and the power balance in a legislature (or delegation). This plot has V, the average district-level proportion of votes won by Democrats (DVS), on the x-axis, and $S(V)$, the proportion of seats won by Democrats, on the y-axis (Tufte, 1973). Figure 3 illustrates several hypothetical seats-votes curves.
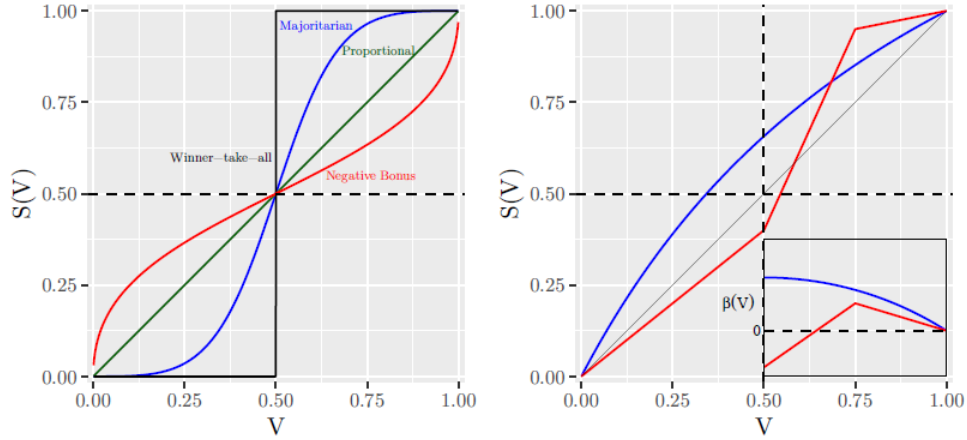

Figure 3

Types of Seats-Votes Curves. Left panel: Symmetric (fair) curves with differing levels of electoral responsiveness Right panel: Asymmetric (biased) curves, including one consistently biased toward the Democrats (blue) and one with biases favoring different parties depending on V (red); the inset graph is for (V) for V 2 [0:5; 1] with the vertical axis scaled to be the same as the main plot, and lines color coded to the seats-votes curves. (Katz et al., 2020, p. 175)

It's very rare to observe sufficient electoral outcomes under the same electoral

system in order to plot the complete curve. (i.e., it's unlikely to observe DVS from 0 to 1 in separate elections occuring in the same districts.) In practice, one can estimate a seats-votes curve using the uniform partisan swing principle (Tufte, 1973).

Uniform partisan swing is the principle that when the DVS changes by some amount between elections, the individual vote share at the district level also changes by the same amount (Tufte, 1973). Katz et al. (2020) empirically verified this to be true in 646 different elections.

Thus, given a list of district-level Democratic vote proportions from one election, one can adjust each vote proportion by an arbitrarily small amount until the entire seats-votes curve is computed (Katz et al., 2020).

The benefit of seats-votes curves is that they allow a complete view of the biases of a district map across all possible average district votes (Gelman & King, 1994). Unfortunately, generating these curves for one electoral map requires the use of statistical assumptions such as uniform partisan swing, as the number of elections occurring under a since district map is usually prohibitively small (Warrington, 2018).

Seats-votes curve visually represent the principles of partisan symmetry and partisan bias.

Partisan Symmetry. A legislature (or delegation) exhibits partisan symmetry if both parties can receive m proportion of the overall votes and therefore have n proportion of the seats (Katz et al., 2020). If Republicans win 60% of the votes but control 65% of the seats, in a symmetrical system, Democrats should also be able to control 65% of the seats by winning 60% of the votes. Seats-votes curves that are rotationally-symmetrical about the center exhibit partisan symmetry.

Partisan Bias. Partisan Bias is the deviation from partisan symmetry in a tied election. If Democrats with 50% of the average district vote but only win 40% of the seats, there is a partisan bias of -0.1. Seats-votes curves that do not intersect the point $(0.5, 0.5)$ (the center) exhibit partisan bias.

Single-Valued Fairness Measures

The following measures compute a single number to assess the partisan fairness of a redistricting plan and are actively disputed by areas of the literature.

Efficiency Gap.   The efficiency gap is the difference between the number of wasted votes for each party, normalized to the total number of votes, where "wasted votes" are all votes for losing candidates and all votes for winning candidates over the 50%-plus-one threshold. Theoretically, wasted votes are a sign of "packing" or "cracking," where a gerrymandered district intentionally grouped voters together into the same district or diluted their political power across several districts. (Stephanopoulos & McGhee, 2014)

Veomett (2018) found that the efficiency gap is not a measure of partisan symmetry, as the measure becomes confused in highly non-competitive elections (e.g. party wins 80% of the vote and 100% of the seats).

Declination.   Declination is another proposed measure of partisan asymmetry that "...relies only on the fraction of seats each party wins in conjunction with the aggregate vote each party uses to win those seats" (Warrington, 2018, p. 3). Broadly, it measures the declination in the line connecting the average vote proportions of one party in districts controlled by the other party. See Figure 4 for a visualization.
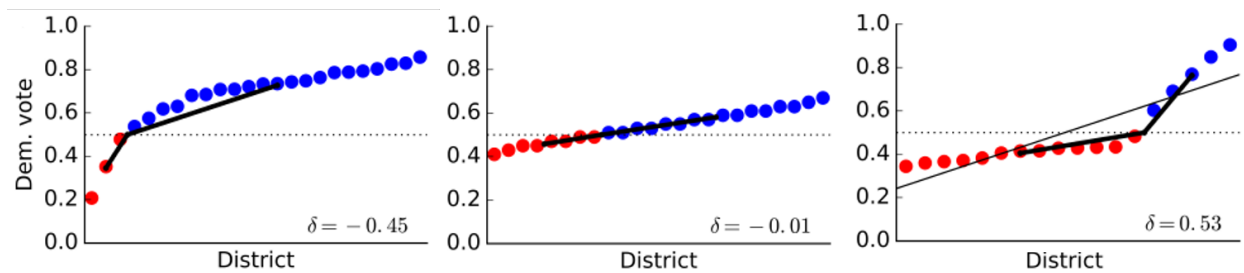

Figure 4

Sample declination plots. Districts plotted by ascending Democratic vote proportion. $\delta \approx 0$ is fair, $\delta < 0$ favors Democrats, and $\delta > 0$ favors Republicans. (Warrington, 2018, p. 6)

Katz et al. (2020) refutes the claim of Warrington (2018) that declination is a measure of partisan symmetry for the same reason that they refute the claim of partisan

symmetry of declination, but acknowledges that it is a useful measure of the skewness of the distribution of district vote proportions.

Mean-Median Difference

The mean-median difference is the difference between the mean and median DVS (first proposed by McDonald and Best (2015)). A mean-median difference of 0 indicates fair districts.

Katz et al. (2020) finds that the mean-median difference is a reliable estimator of partisan bias only in highly competitive elections, when the average DVS is close to 0.5 (Katz et al., 2020, pp. 27–9).
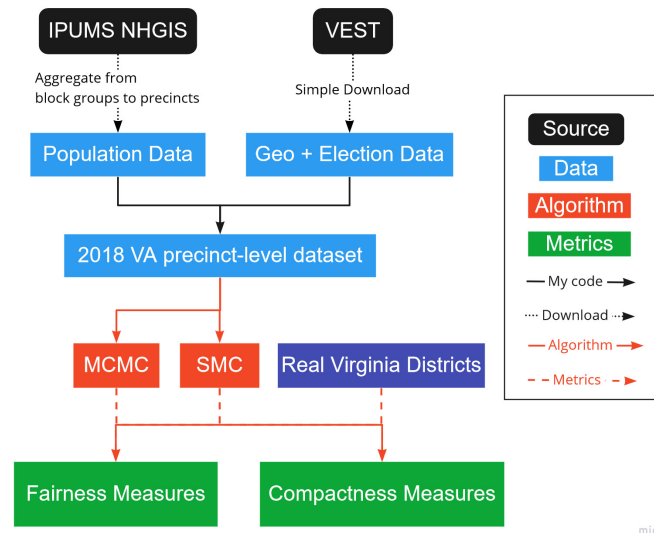
Method

Figure 5

Graphical Overview of project

My research method simulates the 2020 redistricting of the congressional districts in Virginia using two different algorithms supplied with data from 2018. Figure 5 provides and overview of this process.

Choice of Research Method

For this study, I chose to use the experimental design method because it will allow me to isolate the hypothetical impact of the redistricting algorithm from other possible confounding variables. This method also includes the use of a control group, which allows the researcher to establish causation.

Components of Experimental Design

Experimental Units. The experimental units for this study are the complete datasets for each election year in Virginia.[2] Every row in each dataset corresponds to a precinct, the smallest geographical unit by which votes are tabulated in Virginia. For each

---

[2] This corresponds to the blue rectangles in Figure 5.

precinct, I also have the total population and the number of votes cast for the 2018 Democrat, Republican, and other congressional candidates. Additionally, each precinct has a polygon associated with it that represents its geographical shape.

Treatments. The treatments for this study are the two different redistricting algorithm that I'm comparing: Markov chain Monte Carlo (Fifield, Higgins, et al., 2020) and Sequential Monte Carlo (McCartan & Imai, 2020).[3] I'm using the implementations in the R programming language "redist" package (Fifield, Kenny, et al., 2020). See the Literature Review for more information . Broadly, I chose them because they are deterministic. Much of the literature focuses on creating many possible redistricting plans for a commission to choose from, but these three aim to create an "ideal" map. A redistricting plan generated by CRSG serves as the initial map for the MCMC algorithm.[4]

Response Variables. Broadly, the goal will be to evaluate how "fair" and compact each redistricting plan generated by each algorithm for each year is.[5] Refer to the Literature Review for more information on the fairness and compactness measures chosen.

Compactness Measures. I chose to compute the mean Polsby-Popper score, Fryer-Holden score, and the Edge-Cut Compactness measure for the proposed maps from each algorithm. I did not have the necessary computational resources to compute all of the measures that I describe in my Literature Review, so I elected to compute the most popular area-based measure (Polsby-Popper), the only population-based measure (Fryer-Holden), and the primary graph-theory measure (Edge Cut Compactness). While not complete, these measures provide a representative overview of the possible compactness measures.

Fairness Measures. I chose to compute the Partisan Bias, Declination, Efficiency Gap, Equal Population Efficiency Gap, Lopsided Wins, Mean-Median,

---

[3] This corresponds to the red rectangles in Figure 5

[4] MCMC requires a valid map to start with, and the existing map of Virginia was not valid as the population parity was no longer within 1%, as the population has changed since 2016 when the map was drawn.

[5] This corresponds to the green rectangles in Figure 5.

Responsiveness, and Tau Gap measures, all of which are described in the Literature Review, as they represent the fairness measures discussed at present in the literature (Katz et al., 2020). Computational resources were not a barrier for this step, so I was able to calculate all of the measures.

Chamber Power Balance. Since the redistricting that's occurring is hypothetical and I have precinct-level election results for each of these years, I also simulate how many seats the Democratic party would have won in the 2018 General Election under the various redistricting plans.

Control Group. The official VA Congressional district map used in the years 2016-2020[6] will serve as the control group for this experiment. I will compute the same metrics for this map as I will for my hypothetical redistricting plans.[7]

Principles of Experimental Design

The primary principles of experimental research design are randomization, replication, and local control. This is how I plan to address them.

Randomization. Every experimental unit will receive each treatment, and every experimental unit can be replicated many times without issue, so thereâĂŹs no error from a lack of randomization. Think of each treatment operating within a separate parallel universe.

Replication. Each algorithm will generate 100 sample redistricting plans. This is large enough to allow for inferences, but small enough to still be computationally feasible.

Local Control. All of the redistricting will be happening in controlled environments, so there will be no way for lurking variables to creep in and confound my results.

———

[6] Due to racial gerrymandering, VA had to adopt a new congressional district map in the middle of the decade ("Wittman v. Personhuballah," 2016).

[7] This corresponds to the purple rectangle in Figure 5.

Data Cleaning

To create my datasets, I cleaned and compiled three different types of data: demographic data, Geographic Information Systems (GIS) data, and election data.[8]

Demographic Data

One required piece of data in order to redistrict is demographic data at the precinct level. For my purposes, this means the total population of each precinct. In order to run the most accurate redistricting simulations, these data needed to be recent for the year being redistricted. Comprehensive population counts are only conducted by the US Census Bureau every 10 years, so I instead used the 5-year American Community Survey results at the block-group level. This is a sample survey, not a population count, but that is offset by the aggregation of sample data over a 5 year period. I downloaded this data from the IPUMS National Historic GIS project (Manson et al., 2020). Using the "maup" Python Library (Hully, n.d.), I disaggregated the data from the block-group level to the block level, prorating the demographic data based on population. This data was then aggregated up to the precinct level.

GIS Data

In order to redistrict, the algorithms need to know the shape and relative location of each precinct. In practice, this means every precinct has a "polygon" associated with it and a Coordinate Reference System that describes where these polygons fall in space. These data tables with the geometry column are known as "shapefiles." I accessed these shapefiles from the Voting and Election Science Team on their Harvard Dataverse (Voting & Team, 2019). I then merged in my precinct-level demographic data tables, so I now have shapefiles with the necessary demographic data. [9]

---

[8] This is an explanation of the black and blue rectangles in Figure 5 and the transitions between them.

[9] Since election administrators are free to change the precincts between elections, precinct shapefiles are unique to both a place and a time. This is why I couldn't use the 2018 precinct shapefiles with 2020 election results.

Election Data

The last necessary component needed to evaluate redistricting plans is the number of votes won by each party in each precinct in each election.[10] Conveniently, this data was already included in my shapefiles by the Voting and Election Science Team.

---

[10] The algorithms I'm comparing assume a 2 party system, so I only tracked Democratic and Republican votes won in each election.

Results

This section provides an overview of the results from my method.
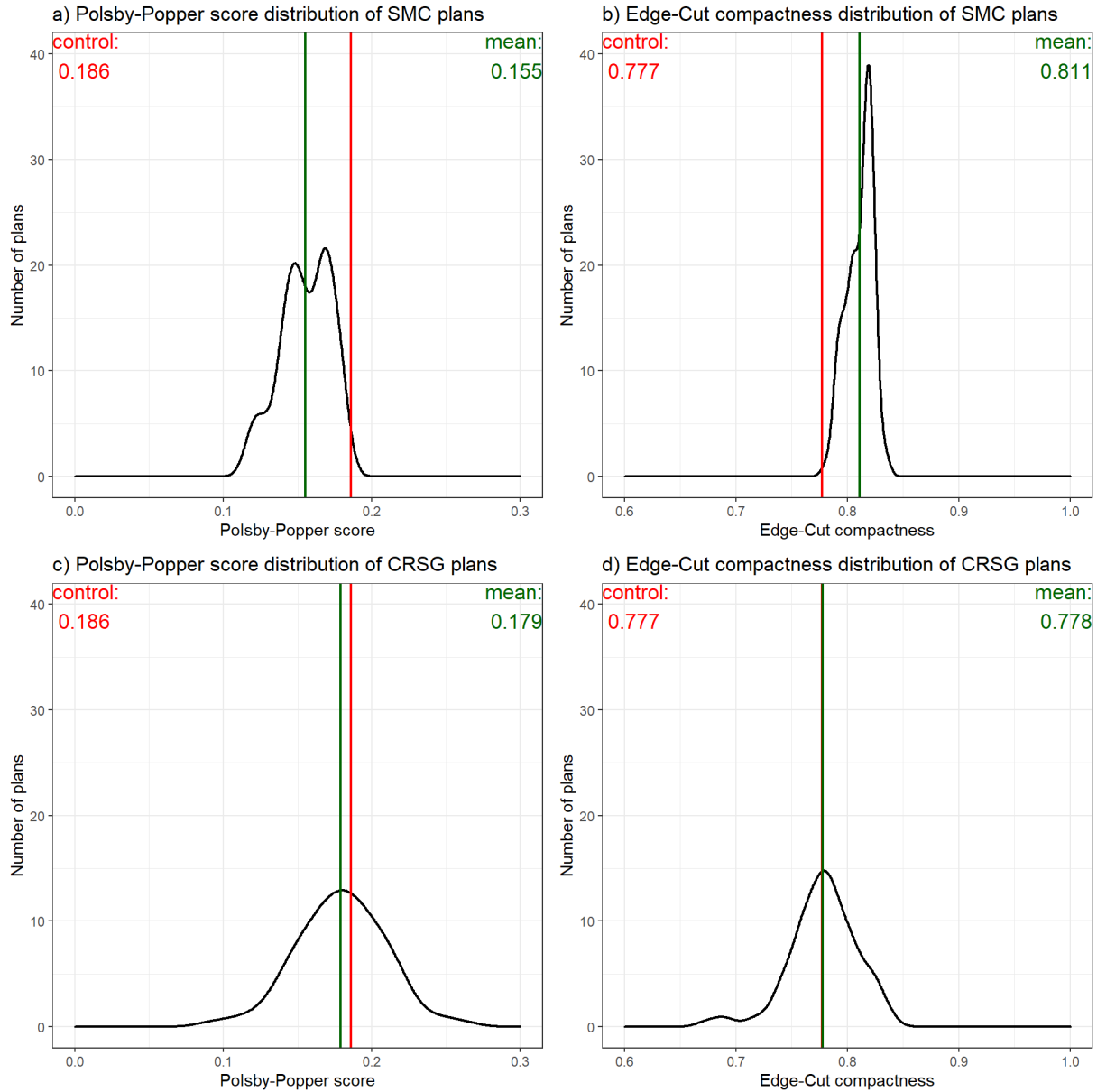
Compactness

Figure 6 visualizes the distribution of two different compactness scores amongst plans generated by both SMC and CRSG. Subfigure (a) shows the distribution of the Polsby-Popper scores of the 100 SMC plans, subfigure (b) shows the distribution of the Edge-Cut compactness measure of the same 100 SMC plans, and subfigures (c) and (d) show the corresponding measure distributions for the plans generated by CRSG. The vertical red lines indicate the value of the measure for the existing district map, and the vertical green lines indicate the mean value of the distribution.

Partisan Fairness

Seats-Votes Curves

Figure 7 shows the seats-votes curves (Katz et al., 2020) for the 2018 General Election under the redistricting plans generated by both algorithms and the existing map. For each plot, the x-axis plots the average of the proportion of votes won by Democrats in each district. The y-axis plots the proportion of seats won by Democrats in the delegation. Both subfigures 7a and 7b have once curve for each redistricting plan (each plot has 100 curves). The seats-votes curve for the real 2018 districts is provided for reference in subfigure 7c.

Single-Valued Partisan Fairness Measures

Figure 6

Distribution of Compactness Measures

Note. The first column of plots shows the Polsby-Popper score distribution; the second column shows the Edge-Cut compactness distribution. The first row corresponds to the plans generated by SMC; the second row corresponds to CRSG. The red lines indicate the measure value for the existing districts; the green lines indicate the mean value of the distribution.

Figure 7

Seats-Votes Curves for SMC, CRSG, and existing plan.
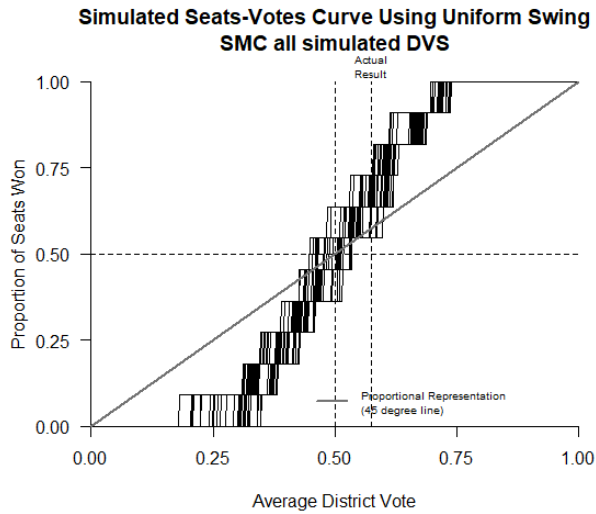
(a) *SMC Seats-Votes Curve*
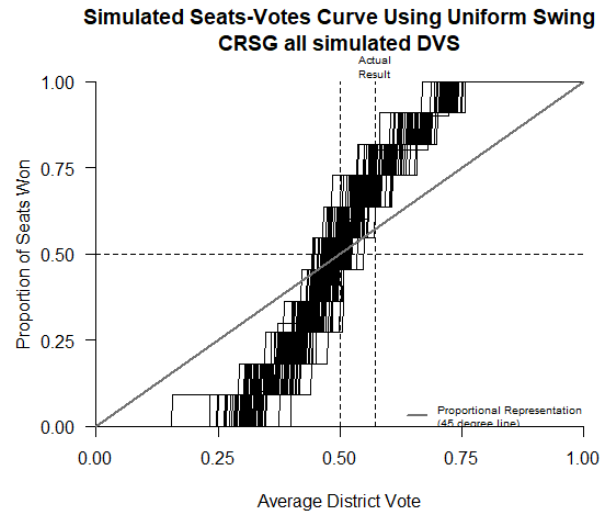
(b) *CRSG Seats-Votes Curve*
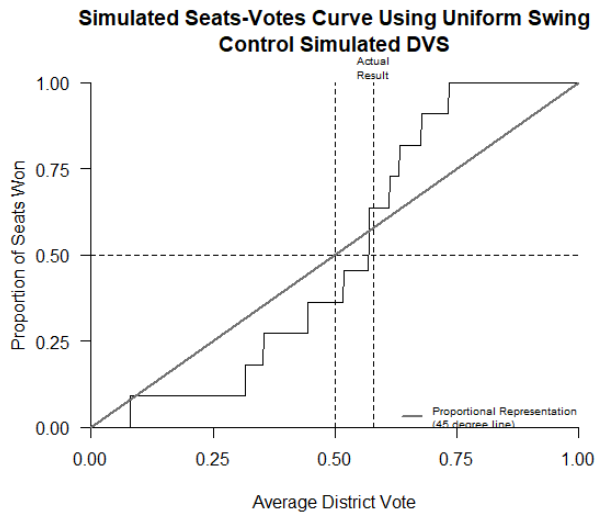
(c) *Real Seats-Votes Curve*

Note. Each plot shows the relationship between average proportion of Democratic vote share by district and the proportion of Democratic seats. Subfigure 7a illustrates this relationship for the 100 plans generated by SMC, subfigure 7b for CRSG, and subfigure 7c is for the existing plan.

Figure 8

Distributions of fairness measures
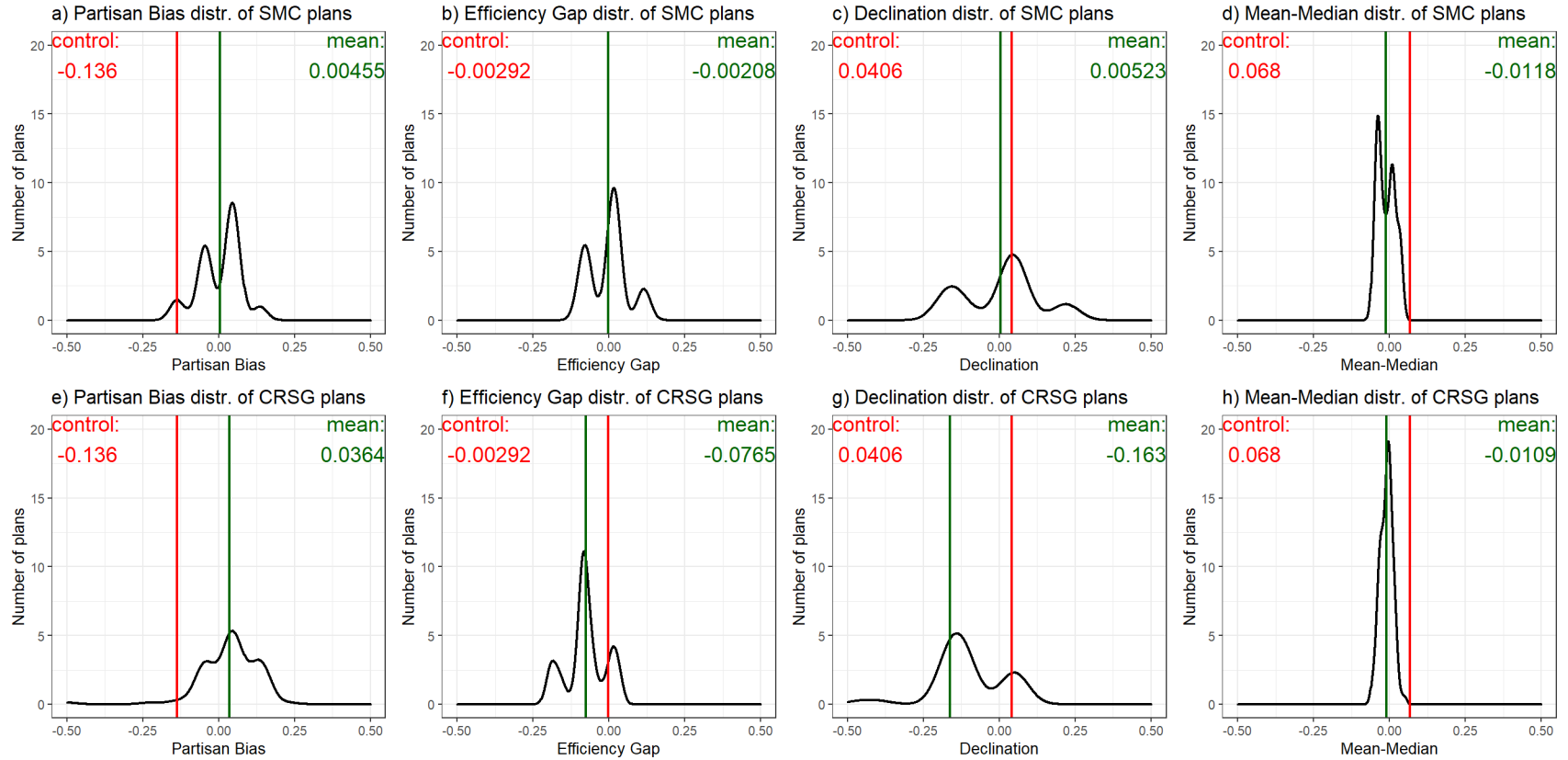
Note. The columns of plots illustrate the distributions of partisan bias, the efficiency gap, declination, and the mean-median difference, respectively. The first row of plots corresponds to SMC plans, the second to CRSG plans. The red lines indicate the corresponding value of the measure in the existing plan; the green lines indicate the mean value of the distribution.

Figure 8 illustrates the distributions of various fairness measures within the plans generated by SMC and CRSG. Columns of plots correspond to different measures, and rows of plots correspond to different algorithms. The red and green vertical lines indicate the control value and mean distribution value, respectively.

Discussion

The goal of my research is to compare two different automated redistricting algorithms in an empirical context and to evaluate the results using a variety of compactness and partisan fairness standards that have been established in the literature.

I will begin by analyzing the compactness metrics and then the partisan fairness metrics. I will end by discussing the limitations of my findings.

Compactness

Beginning with compactness measures, I'm analyzing the results presented in Figure 6.

Polsby-Popper score

The Polsby-Popper score compares the area of the district to the area of a circle with the same perimeter as the district (Polsby & Popper, 1991). It ranges from 0 to 1, where higher values indicate a compacter district.

The SMC plans averaged a Polsby-Popper score of 0.155, which is slightly less compact than the real district (0.186). The CRSG plans also averaged a score just slightly less than the control plan (0.179). Through the lens of this measure, CRSG produced plans as compact as the control, and SMC's plans were slightly less compact.

Edge-Cut compactness

The second compactness metric measured was the edge-cut compactness score, a measure grounded in graph theory. It measure the proportion of edges that had to be cut from the initial precinct graph to form the district, and has been normalized to the total number of districts[11] (Dube & Clark, 2016). This measure isn't affected by geographic features such as coastlines or mountain ranges, which makes it a more reliable measure of compactness than Polsby-Popper to some (see McCartan and Imai (2020)).

Through the lends of edge-cut compactness, SMC plans were more compact than

---

[11] For an explanation of the graph of precincts, please see the Redistricting as Graph Cutting section.

the control ($0.811 > 0.777$), and CRSG plans were just as compact as CRSG plans ($0.778 \approx 0.777$).

Conclusion

Both compactness measures did not reach the same conclusions. Polsby-Popper found the SMC plans to be less compact on average than the control, while edge-cut compactness found SMC to be more compact. This is likely due to the fact that Polsby-Popper scores are very sensitive to resolution and geography (McCartan & Imai, 2020). Virginia is a states with high-perimeter borders, defined by the Blue Ridge Mountains and the Chesapeake Bay. Additionally, Accomack County includes the Virginia Barrier Islands, which are separately from the mainland of the state ({United States Geological Survey}, 2021). All this is to say that the very uneven border of Virginia due to geography likely interfere with the ability of the Polsby-Popper score to accurately quantify compactness.

With that in mind, if we only consider edge-cut compactness, SMC was able to generate compacter plans the CRSG or the control. However, both distributions were reasonably compact, and there are no worrying differences between compactness between the three proposals.

Partisan Fairness

Moving on to the partisan fairness results, I'll first analyze the seats-votes curves, and then the single-value measures outlined in the Literature Review.

Seats-Votes Curves

Figure 7 includes the seats-votes curves that I'm going to analyze. When viewed as a whole, the 100 curves for SMC in a) form a Majoritarian seats-votes curve (see Figure 3). Partisan symmetry is the symmetry about the point $(0.5, 0.5)$ on this graph (Katz et al., 2020), and the SMC plan curves display partisan symmetry. The same is true for the 100 curves from CRSG in b). This indicates that both algorithms generated plans that, when taken together, do not advantage one party over the other.

Compared to the curve for the existing district map in 2018 (shown in c)), the two algorithms generated more-symmetrical plans. The control plan is still Majoritarian, but it favors Republicans from DVS $[0, 0.53]$ and Democrats from $[0.53, 1]$. To illustrate this, we can see that if Democrats won 50% of the average district vote, they'd win about 38% of the seats, not 50%.

In summary, both algorithms generated more symmetrical plans on average when compared to the existing plan.

Single-Valued Measures

Next I will analyze the partisan bias, efficiency gap, declination, and mean-median difference score distributions for the two algorithms. These results are visualized in Figure 8.

Partisan Bias. Partisan bias is the distance from partisan symmetry at an average DVS of 0.5. On a seats-votes curve (see Figure 7), this is the difference between 0.5 and the y-coordinate of the curve at $x = 0.5$. It ranges from -0.5 to 0.5, with values of 0 being least biased. (Katz et al., 2020)

As shown in a), the SMC plans on average were very fair ($0.0045 \approx 0$). e) shows this to be true for the CRSG plans as well, though they do slightly favor Democrats ($0.0364 > 0$). Both algorithms generated less biases plans on average than the control, which significantly favored Republicans ($-0.136 < 0$).

These results align with the the conclusions drawn from the seats-votes curve. Interestingly, the control plan is biased towards Republicans, but this is not evident in the election map, as the actual DVS is greater than 0.5. Nevertheless, both algorithms generated less-biased plans on average than the control plan.

Efficiency Gap. The next measure is the efficiency gap, which measure the difference in "wasted votes" between majority-Democrat and majority-Republican districts, normalized to the total vote count. It ranges from -1 to 1. (Stephanopoulos & McGhee, 2014)

All three sets of plans; SMC, CRSG, and the control; have an efficiency gap within the acceptable margin (± 0.07).

The efficiency gap doesn't find evidence of significant differences in the number of wasted votes between Democratic and Republican districts between the algorithms and the control. This aligns with the literature, as most criticisms of the efficiency gap point out that is mischaracterizes "wasted votes" in highly non-competitive elections (see Veomett (2018) and Katz et al. (2020)). However, Virginia is a very competitive state, so this does not appear to apply in this situation.

Declination. The next measure is declination, which quantifies the concavity of the the line on the plot of districts by ascending DVS; see Figure 4 for a visualization. It ranges from -1 to 1, with values of 0 indicating fairness. (Warrington, 2018)

The SMC plans are on average fair ($0.00523 \approx 0$), as is the control plan ($0.0406 \approx 0$). However, the declination of the CRSG is biased towards Democrats ($-0.163 < 0$). Seeing as the seats-votes curves for SMC and CRSG were found to be symmetrical, these results agree with the findings of Katz et al. (2020) that declination is not a measure of symmetry, but rather of the skewness of the district vote proportion distribution.

This district-level vote proportion distribution appears to be skewed towards Democrats on average in the CRSG plans, and centered in the average SMC plans and the control plan.

Mean-Median Difference. The next measure is the mean-median difference, which is the difference between the mean district DVS and the median district DVS. It ranges in value from -1 to 1, with values of 0 indicating a non-skewed distribution (McDonald & Best, 2015). Katz et al. (2020) finds that at an averaged DVS of 0.5, the mean-median difference measures partisan bias.

The result that the control plan slightly favors Republicans ($0.068 > 0$) while the average SMC plans and CRSG plans are basically fair ($-0.0118 \approx 0$, $-0.0109 \approx 0$), aligns with the results from the partisan bias measure, which supports the finding of Katz et al.

(2020).

Conclusion

When viewed together, the partisan fairness measures illustrate the following conclusions. SMC generated fair and symmetrical districts from the perspective of all four measures. CRSG also generated on average fair districts, with only a skew in the district-level Democratic vote proportion. Both algorithms produced on average fairer districts than the control plan, which was found to be biased towards Republicans by all fairness measures.

Limitations

The limitations of the compactness measures largely are due to their sensitivity to geography. Additionally, the seats-votes curves are very sensitive to the idiosyncrasies of the particular election since the number of seats is so small. Each algorithm generated 100 districts, so the possibility of randomly generated biased districts could be reduced by having a larger sample size. Finally, every district vote proportion was calculated using only the votes for the Democratic and Republican parties, ignoring third-party candidates, though this is unlikely to have had a significant effect as the United States has first-past-the-post elections.

Conclusions

There were not significant differences in the average compactness of plans generated by the algorithms and the control plan. The compactness measures provided additional evidence that the Polsby-Popper score cannot accurately measure compactness when faced with districts that have lengthy perimeters due to their geography.

SMC was found to generate fair, symmetrical districts. CRSG districts were mostly fair, though they showed a slight Democratic bias in some cases. The control districts were found to be biased towards Republicans by all measures of fairness.

Conclusion

The objective of this research was fill the gap in the scholarship w.r.t. comparative, empirical case studies of the performance of automated redistricting algorithms. I therefore set out to answer the question, *How do the hypothetical district maps for the Virginia Congressional delegation for the 2020s generated by different automated redistricting algorithms compare based on partisan fairness and compactness measures?*

In my experimental case study, I simulated the 2021 redistricting of the congressional districts in Virginia using the CRSG and SMC algorithms, and evaluated the resulting maps and the existing map using various measures of compactness (the Polsby-Popper score and the edge-cut compactness score) and of partisan fairness (the seats-votes curve, partisan bias, the efficiency gap, declination, and the mean-median difference).

The key conclusions from my research was that SMC reliably generated the fairest districts, that CRSG generated fair districts less reliably, and that the existing district plan significantly favored Republicans. This suggests that SMC is the favored choice of algorithm for such a redistricting task.

Regarding compactness measures, I confirmed the findings of McCartan and Imai (2020) that the Polsby-Popper measure fails to assess compactness fairly when physical geography necessitates a certain degree of incompactness. I find the edge-cut compactness measure to be robust against variations in physical geography, which confirms the findings of Dube and Clark (2016).

Regarding measures of partisan fairness, the seats-votes curve provided the clearest picture of irregularities in the seats-votes relationship, which agrees with the conclusions of Katz et al. (2020). I also find the mean-median difference to align with the partisan bias measure, as theoretically verified by Katz et al. (2020). My results also confirmed the findings of Veomett (2018) that the efficiency gap can measure partisan symmetry in competitive elections. Lastly, I also find declination to measure the skewness in the district

vote distribution, not partisan fairness, confirming the theory of Katz et al. (2020).

Taken in concert, my findings present a significant evolution in the capabilities of automated redistricting algorithms, as well as initial evidence that the SMC algorithm could successfully be used by a redistricting commission tasked with redrawing district boundaries. Tangentially, I also provided limited empirical evidence supporting the findings of various scholars that certain measures measure quantities other than partisan symmetry and that others are not robust against empirical requirements.

As the SMC algorithm continues to evolve and improve, we could be approaching a reality where a number of "fair" redistricting plans are generated by algorithms, and then the final map is selected by members of a committee that represents various stakeholders.

However, further research is needed to continue evaluating redistricting algorithms as they continue to improve. My research only focused on a single redistricting cycle of a single delegation of a single state. Furthermore, access to greater computing power would enable larger sets of maps to be generated. Additionally, empirical evaluations of these algorithms will need to include further real-world stipulations, such as obeying county boundaries, preserving majority-minority districts, and generating districts that vary minimally from the existing districts.

Such research will enable the move of automated redistricting algorithms from the scholarly literature into practice in independent commissions, which is one necessary step towards the goal of increasing fairness in our district-based democracy.

References

Altman, M., & Mcdonald, M. P. (2009). Bard: Better Automated Redistricting. *Journal of Statistical Software.*

Boyce, R. R., & Clark, W. A. V. (1964). The Concept of Shape in Geography. *Geographical Review, 54*(4), 561. https://doi.org/10.2307/212982

Chen, J., & Rodden, J. (2013). Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quarterly Journal of Political Science, 8*(3), 239–269. https://doi.org/10.1561/100.00012033

Cox, E. P. (1927). A Method of Assigning Numerical and Percentage Values to the Degree of Roundness of Sand Grains. *1*(3), 179–183.

Davis v. Bandemer. (1986). Retrieved April 11, 2021, from https://www.oyez.org/cases/1985/84-1244

Dube, M., & Clark, J. (2016). Beyond the Circle: Measuring District Compactness Using Graph Theory.

Duchin, M. (2018). *Outlier Analysis for Pennsylvania Congressional Redistricting* (tech. rep.). Metric Geometry and Gerrymandering Group.

Fifield, B., Kenny, C. T., McCartan, C., Tarr, A., Higgins, M., Kawahara, J., & Imai, K. (2020). Redist: Simulation Methods for Legislative Redistricting. Retrieved January 29, 2021, from https://CRAN.R-project.org/package=redist

Fifield, B. (2020). The Essential Role of Empirical Validation in Legislative Redistricting Simulation*. *STATISTICS AND PUBLIC POLICY*, 17.

Fifield, B., Higgins, M., Imai, K., & Tarr, A. (2020). Automated Redistricting Simulation Using Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics, 0*(0), 1–14. https://doi.org/10.1080/10618600.2020.1739532

Gelman, A., & King, G. (1994). Enhancing Democracy Through Legislative Redistricting. *American Political Science Review, 88*(3), 541–559. https://gking.harvard.edu/files/abs/red-abs.shtml

Harris, C. C. (1964). A Scientific Method of Districting. *Behavioral Science*, *9*(3), 219–225. https://doi.org/10.1002/bs.3830090303

Herschlag, G., Ravier, R., & Mattingly, J. C. (2017). Evaluating Partisan Gerrymandering in Wisconsin. *arXiv:1709.01596 [physics, stat]*. Retrieved April 11, 2021, from http://arxiv.org/abs/1709.01596

Hu, T. C., Kahng, A. B., & Tsao, C.-W. A. (1995). Old Bachelor Acceptance: A New Class of Non-Monotone Threshold Accepting Methods. *ORSA Journal on Computing*, *7*(4), 417–425. https://doi.org/10.1287/ijoc.7.4.417

Hully, M. (n.d.). Maup: The Geospatial Toolkit for Redistricting Data. Retrieved February 18, 2021, from https://github.com/mggg/maup

Katz, J. N., King, G., & Rosenblatt, E. (2020). Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies. *American Political Science Review*, *114*(1), 164–178. https://doi.org/10.1017/S000305541900056X

Lara-Caballero, A., de-los-Cobos-Silva, S. G., Mora-Gutiérrez, R. A., Rincón-García, E. A., Gutiérrez-Andrade, M. Á., & Lara-Velázquez, P. (2019). Multiobjective Genetic Algorithms for Reinforcing Equal Population in Congressional Districts. *Mathematical Problems in Engineering*, 1–14. https://doi.org/10.1155/2019/2825854

Maceachren, A. M. (1985). Compactness of Geographic Shape: Comparison and Evaluation of Measures. *Geografiska Annaler: Series B, Human Geography*, *67*(1), 53–67. https://doi.org/10.1080/04353684.1985.11879515 _eprint: https://doi.org/10.1080/04353684.1985.11879515

Macmillan, W. (2001). Redistricting in a GIS Environment: An Optimisation Algorithm Using Switching-Points. *Journal of Geographical Systems*, *3*(2), 167. https://doi.org/10.1007/PL00011473

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., & Ruggles, S. (2020). National Historical Geographic Information System: Version 15.0. https://doi.org/10.18128/D050.V15.0

McCartan, C., & Imai, K. (2020). Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans. *arXiv:2008.06131 [cs, math, stat]*. Retrieved January 18, 2021, from http://arxiv.org/abs/2008.06131

McDonald, M. D., & Best, R. E. (2015). Unfair Partisan Gerrymanders in Politics and Law: A Diagnostic Applied to Six Cases. *Election Law Journal: Rules, Politics, and Policy, 14*(4), 312–330. https://doi.org/10.1089/elj.2015.0358

Polsby, D. D., & Popper, R. D. (1991). The Third Criterion: Compactness as a Procedural Safeguard against Partisan Gerrymandering. *Yale Law & Policy Review, 9*(2), 301–353. Retrieved March 12, 2021, from http://www.jstor.org/stable/40239359

Project, P. G. (n.d.). Home. Retrieved April 11, 2021, from https://gerrymander.princeton.edu/

Reock, E. C. (1961). A Note: Measuring Compactness as a Requirement of Legislative Apportionment. *Midwest Journal of Political Science, 5*(1), 70–74. https://doi.org/10.2307/2109043

Schwartzberg, J. (1966). Reapportionment, Gerrymanders, and the Notion of Compactness. *Minnesota Law Review, 50*, 443–452. https://scholarship.law.umn.edu/mlr/1701

Staff, B. (2020). Virginia Question 1, Redistricting Commission Amendment (2020). Retrieved April 11, 2021, from https://ballotpedia.org/Virginia_Question_1,_Redistricting_Commission_Amendment_(2020)

Stephanopoulos, N., & McGhee, E. (2014). *Partisan Gerrymandering and the Efficiency Gap* (SSRN Scholarly Paper No. ID 2457468). Social Science Research Network. Rochester, NY. Retrieved March 26, 2021, from https://papers.ssrn.com/abstract=2457468

Tufte, E. R. (1973). The Relationship between Seats and Votes in Two-Party Systems. *American Political Science Review, 67*(2), 540–554. https://doi.org/10.2307/1958782

{United States Geological Survey}. (2021). The National Map Services. Retrieved March 31, 2021, from https://apps.nationalmap.gov/services/

Veomett, E. (2018). Efficiency Gap, Voter Turnout, and the Efficiency Principle. *Election Law Journal: Rules, Politics, and Policy*, *17*(4), 249–263. https://doi.org/10.1089/elj.2018.0488

Vickrey, W. (1961). On the Prevention of Gerrymandering. *Political Science Quarterly*, *76*(1), 105. https://doi.org/10.2307/2145973

Voting, & Team, E. S. (2019). 2018 Precinct-Level Election Results. https://doi.org/10.7910/DVN/UBKYRU

Wang, S. S.-H. (2016). Three Tests for Practical Evaluation of Partisan Gerrymandering. *Stanford Law Review*, *68*, 1263–1321.

Warrington, G. S. (2018). Quantifying Gerrymandering Using the Vote Distribution. *Election Law Journal: Rules, Politics, and Policy*, *17*(1), 39–57. https://doi.org/10.1089/elj.2017.0447

Weaver, J. B., & Hess, S. W. (1963). A Procedure for Nonpartisan Districting: Development of Computer Techniques. *The Yale Law Journal*, *73*(2), 288–308. https://doi.org/10.2307/794769

Wesberry v. Sanders. (1964). Retrieved April 11, 2021, from https://www.oyez.org/cases/1963/22

Wittman v. Personhuballah. (2016).

Xiao, N. (2008). A Unified Conceptual Framework for Geographical Optimization Using Evolutionary Algorithms. *Annals of the Association of American Geographers*, *98*(4), 795–817. https://doi.org/10.1080/00045600802232458