Using Gerrymandering For School Redistricting

Joshua Park, Amy Suo, Jonathan Sy, Henry Xie

August 13, 2020

Counselor: Mirilla Zhu

1 Abstract

With the 2020 Census currently being conducted across the United States, redistricting is at the forefront of many political discussions, as each party does not want to have a disadvantage going into elections for the next decade. However, we will investigate the possible applications of these concepts to school redistricting. In addition to writing a computer program to score school districting plans, we also propose some modifications to the current methodology used to generate the plans. We provide a couple of algorithms to make the process more efficient and derive a formula for the probability that a plan can even be made for an initial partition under a set of circumstances.

2 Background

2.1 Context of Gerrymandering

Gerrymandering is a practice in which groups manipulate political boundaries to win unfair political gain. There are two primary tactics used in gerrymandering: "cracking," diluting an opposition party's base of support among many districts, and "packing," concentrating an opposition party's support in relatively few districts to reduce its power in other districts.

Gerrymandering is a practice as old as modern democracy. The term is named after Elbridge Gerry, a Governor of Massachusetts who, in 1812, infamously drew the boundaries of a partisan political district shaped like a salamander. Gerrymandering has been a consistent aspect of American politics to this day and remains a hot-button issue that is surrounded by much debate.

Critics argue that gerrymandering is inherently undemocratic because it makes elections less competitive, reduces transparency, and harms effective representation. In 2019, the Supreme Court ruled in Rucho v. Common Cause that issues relating to partisan gerrymandering are beyond the political reach of the federal courts, thereby refusing to settle the question of gerrymandering. Thus, the debate regarding gerrymandering rages on. While gerrymandering has long been considered primarily a political issue, districting in other situations is also prone to being manipulated much like how political districts are manipulated in gerrymandering. In this project, we analyze school districting by using tools developed for the analysis of partisan gerrymandering.

2.2 Why Massachusetts?

Gerrymandering first emerged during political redistricting in Massachusetts. Gerrymandering is the practice of manipulating the boundaries of a district to favor a certain group. The name gerrymandering came during the redistricting of Massachusetts, where a newly drawn district sliced up Essex county, a political stronghold for the Federalist party, in order to weaken the Federalist majority within the state. This redistricting plan, signed by Elbridge Gerry, looked like a strange salamander monster, thus giving the name “Gerrymandering”.

Many argue that gerrymandering is undemocratic, because it creates an uneven representation of each state. Gerrymandering has been often discussed within court, but always fails to be resolved.

Many see gerrymandering as a political problem; however, it has also been seen to apply to school districts and race as well. Racial segregation within Boston schools has occurred many times within history. Schools for Africans Americans were created in the 18^{th} century, when African Americans did not want to face discrimination and racism within public schools with a white majority population. This started segregation in Boston public schools.

African American students were refused entrance into public schools that were near their homes and instead were forced to go to schools where African Americans were the majority. These schools had less funding and were seen as “inferior.” African American families became angry about this discrimination and protested against such.

The Racial Imbalance Act made the segregation of public schools illegal in Massachusetts. This act was often broken by Boston’s school committee, when they refused to follow through such plans. This was then followed by the boycotting of schools by African American families. Boston embarked on a campaign of desegregation in the 70s and 80s, in which they bused white children to primarily African American schools and African American children to primarily white schools.

However, the issue of de facto segregation persists in Massachusetts, which has some of the most racially imbalanced schools in the country. In 2019, two out of three students of color attended an intensely segregated school, one with more than 90% of students being students of color, while in 1980, this was only 2$. More than half of Boston public schools are heavily segregated today.

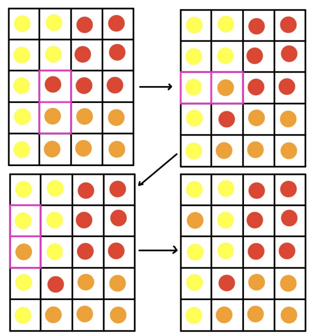
2.3 Gerrychain

Our work analyzing school districts makes extensive use of Gerrychain, which is a project developed by Metric Geometry and Gerrymandering Group, a research group actively exploring models of gerrymandering. While Gerrychain is usually used to analyze political districting, we have applied it analogously to analyzing school districting.

We used Gerrychain to generate an ensemble of many school districting plans and score them based on the distribution of race and household income in each plan. In particular, we used two methods, the flip method and the ReCom method, to generate ensembles.

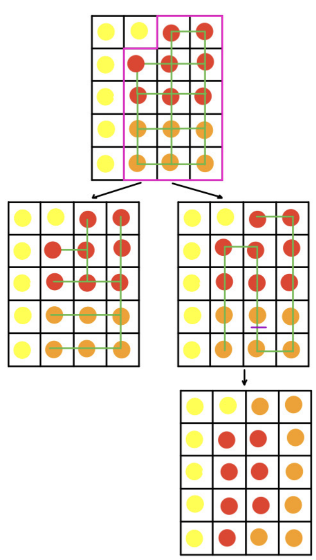
2.4 Flip Method

The Flip method involves continually repeating a process in which two neighboring districts are selected, two census units that border each other within the two districts are selected, and finally, the units are flipped between the two districts. In each step, we check to ensure that the resulting districts remain contiguous and have roughly equal populations as they started with. By repeating this process a large number of times, we generate an ensemble of districting plans. However, even after numerous steps, we can still see that the plan retains some of its original characteristics.



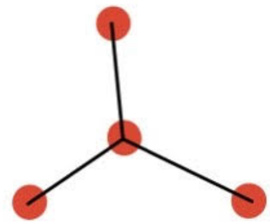
2.5 ReCom Method

We use this method to go on a random walk on the space of districting plans. We begin with a districting plan and merge two neighboring districts. We then take its adjacency graph and identify a spanning tree. Then, we remove a cut edge to make two trees, each of which forms a new district. We evaluate these new districts to ensure that they satisfy enrollment and school conditions. Comapred to the Flip Method, even with one step it is hard to identify the initial partition.



3 Splitting Trees

When using the ReCom method, we must choose an edge to cut such that the two resulting components have equal populations. However, there is not always such a edge.



If a tree does have such an edge, let that edge be e and the tree be called splittable. This begs the question of what the probability is for any given tree on $2n$ vertices to be splittable. For this section, we will assume that the weight of each vertex is 1. We consider two cases: trees on labeled vertices and non-labeled vertices.

3.1 On Labeled Vertices

Theorem 3.1.1. *For a tree on 2n labeled vertices, the probability the tree is splittable is* $\frac{(2n-1)\binom{2n-2}{n-1}}{n \cdot 2^{2n-2}}$.

Proof. When the edge with this property is removed, we want two trees of n vertices. Since they are both trees, the size should be one less than the order. Thus, each tree has $n-1$ edges, leading to a total of $2n-1$ edges when we include both trees and the edge we removed. The resulting construction must be connected. Thus, each tree on $2n$ vertices that has an edge with this property can be constructed by taking two trees on n vertices and adding an edge incident to a vertex of each tree. Since this is on labeled vertices, we first select two vertices that are incident to the edge with this property from the 2n total vertices. Next, we select the $n-1$ vertices needed to complete one of the trees (the rest goes to the other tree). Finally, by Cayley's Tree Formula [1], there are n^{n-2} trees on n labeled vertices. Each of the two separate initial components has n^{n-2} different possibilities. Thus, the total number of trees that are splittable is equal to $\binom{2n}{2}(n^{n-2})^2\binom{2n-2}{n-1}$.

Again, by Cayley's Tree Formula, the total number of trees on 2n vertices is $(2n)^{2n-2}$. Thus, our probability is $\frac{\binom{2n}{2}(n^{n-2})^2\binom{2n-2}{n-1}}{(2n)^{2n-2}}$, or $\frac{(2n-1)\binom{2n-2}{n-1}}{n \cdot 2^{2n-2}}$. □

3.2 On Non-Labeled Vertices

The probability for trees on non-labeled vertices is much more complicated. While we were not able to obtain a closed formula, we did run Python simulations to find the probability for $n \leq 6$. Below, we have listed the results of our experiments. As seen in OEIS entry A000055, the number of trees on non-labeled vertices grows exponentially, not allowing much brute analysis for larger n.

n	Splittable Trees	Total Trees	Probability
1	1	1	1/1
2	1	2	1/2
3	3	6	3/6
4	10	23	10/23
5	45	106	45/106
6	211	551	211/551

Table 1: Probability for Trees on $2n$ Non-Labeled Vertices

Conjecture 3.2.1. *The probability that a given tree up to isomorphism with* $2n$ *vertices has an edge that you can cut to yield two subtrees each with* n *vertices is no greater than* $\frac{1}{2}$ *for* $n > 1$.

Conjecture 3.2.2. *The probability that a given tree up to isomorphism with* $2n$ *vertices has an edge that you can cut to yield two subtrees each with* n *vertices approaches* 0 *as* n *approaches infinity.*

4 Proposed Algorithms

When given a tree, we want to know if it is splittable and if it is where the splitting edge is located. Recall a tree is splittable if there is an edge such that when removed it yields two subtrees of equal order.

Currently, when running the ReCom method, edges are selected at random to be cut. Afterwards, the two components are compared to see if their populations are within 1% of each other. As the number of vertices runs into the thousands, this method will become incredibly inefficient, since it will need to test all edges if there is no such cut. We have listed below two algorithms that we believe are more efficient alternatives.

Algorithm 1: Proposed Algorithm 1

Result: Is There An Edge That Is Cut-able?

1. If all vertices are less than half the total weight, continue to 2. If a vertex has weight equal to half the total weight and the vertex is a leaf, cut the edge incident to it. END. If a vertex has weight equal to half the total weight and the vertex is an internal vertex, there is no such edge. If a vertex has weight greater than half the total weight, there is no such edge. END.
2. Take the vertex of largest degree. Sum the weights of that vertex and any leaves that are adjacent to that vertex. If sum is greater than half the total weight, there is no such edge. END. If sum is equal to half the total weight, continue to 3. If sum is less than half the total weight, continue to 4.
3. If the vertex of highest degree is adjacent to more than one vertex that is not a leaf, there is no such edge. If the vertex of highest degree is adjacent to exactly one vertex that is not a leaf, cut the edge incident to the vertex of highest degree and the non-leaf. If the vertex of highest degree is adjacent to zero vertices that are not leaves, there is no such edge.
4. Choose a vertex adjacent to the vertex with highest degree that is not a leaf. There must be at least one or else the graph is a star graph and there is no such cut, since we know no individual vertex is equal to half the total weight. Take the vertex-induced subgraph using vertices that are descendants of this non-leaf vertex, excluding the initial vertex of highest degree. Repeat for all such internal vertices adjacent to the initial vertex.
5. If the largest sum is greater than half the weight of the tree, continue to 6. If the largest is less than half the weight of the tree, continue to 7. If any sum is equal to half the weight of the tree, cut the edge incident to both that internal vertex and the initial vertex.
6. Repeat steps 1-5 on the vertex-induced subgraph. The ideal population still remains as half the total weight of the tree (even though not necessarily all vertices in the tree are in the subgraph).

If the largest sum is greater than half the weight of the tree, it could be possible that the a cut edge is contained within that subgraph. It would not make sense for it to be outside, since that subgraph already weighs more than half the total weight.

7. If the sum of the sums from step 2 and step 5 is equal to half the weight of the entire tree, continue to 8. If the sum of the sums from step 2 and step 5 is greater than half the total weight, take the component and the group from step 2 and repeat steps 1-5. If the sum of the sums from step 2 and step 5 is less than half the total weight, there is no such edge. END.
8. If there is more than one other internal vertex incident to the initial vertex, there is no such edge. END. If there is one internal vertex incident to the initial vertex, cut the edge incident to both.

Any possible edge cut will be on one of the subgraphs. If more than one other internal vertex exists, there are at least 3 subgraphs. Any cut must contain two of the subgraphs.

Another possible approach we can take to this issue is starting at the leaves and working our way inwards. Both are advantageous over the current partition, since they can bail on trees much quicker. In addition, they can group vertices that must go together in the same component. Thus, we avoid needing to poking around in the dark and checking our guess.

5 Applying Gerrymandering to School Districting

Although gerrymandering is typically associated with politics, we can attempt to apply the same principles to school districting plans to achieve equal representation for various statistics, including race and income. Using Gerrychain, we can create millions of plans using the Flip or ReCom method and evaluate these plans using numerous indicators. We have attached all our Gerrychain projects and the data used in this GitHub repository: `https://github.com/Park2101/Gerrymandering-Final.git`.

5.1 Data Collection

Census data is provided on the National Historical Geographic Information System (NHGIS) website. We were able to obtain data on race, income, and total population for the state of Massachusetts. We were also able to download the shape files that were later used to create the adjacency graphs and run the Gerrychain code.

Data for the enrollment of all public schools in Massachusetts and their addresses were found on the Massachusetts State Department of Education website. In order to sort the schools into the blocks they belong to, a Geocoder found on the United States Census Bureau website was used.

5.1.1 Assumptions

Since data was pulled from various sources, there were bound to be discrepancies with the two data sets. There are some assumptions that were made in order to make our evaluation possible. Since the population and age data is based on the 2010 Census and the child population has since declined, the number of students to a school will need to exceed its capacity. Thus, we need to scale the populations down to create a hypothetical situation. For our overall block group data, we multiplied the "Age 15-19" by 0.625 and multiplied the "Age 5-14" column by 0.80 to bring the data into a range that was reasonable in comparison to the enrollment data.

5.2 Scores

We wanted to measure the partition on three criteria. The first was in relation to income. We used the metric provided by NHGIS that gives the number of people in a population with a certain ratio between their income and the poverty line. We wanted to have the metrics be as close to the overall state's as possible. Thus, the score for each district is the square of the difference between the district and state percentages. The overall score for a plan is the sum of all district scores within it.

The race score was defined in a similar way and considered all races that had data from the NHGIS website.

Finally, we wanted to calculate a Polsby-Popper test score, in order to avoid extreme travel times to school.

To factor any combination of these, we can scale these by any factor of our choosing, based on their importance.

5.3 Suffolk County

Suffolk County, the county which Boston belongs to, was of particular interest. Based on the school district data we found, there does not seem to be any laws limiting the population of school districts or prohibiting districts that span more than one district. Suffolk County was quite intriguing, since more than a hundred schools were packed into a single district. We wanted to break down this district ourselves and look into the distribution of people to a smaller set of schools. However, this became very problematic, since it hindered our ability to set appropriate bounds.

5.4 Massachusetts State

Traditionally, school districts are created by group census blocks (explanation of census blocks). However, since the amount of data for the census blocks was too large to run in our program, we used census block groups instead, to model the entirety of Massachusetts. For comparison, there are 157,509 blocks in Massachusetts and 4,986 block groups in Massachusetts. Although using only block groups will severely limit the range of options, we wondered if it still can provide a decent representation of the possible plans that could be implemented.

6 Future Research

We were only able to present a closed formula for the probability for trees on $2n$ labelled vertices to have a edge that can be cut to yield two subtrees with n vertices. However, we could not answer the question about trees on non-labeled vertices. Further research could investigate the property for trees on non-labeled vertices, as well as just finding a formula for the number of trees on n non-labeled vertices. As we mentioned earlier, trees in the ReCom algorithm will not all weigh 1 unit, but have weights corresponding to the populations of those precincts or blocks. Thus, we could ask the same two questions, but now seeking to find two subtrees with the same weights.

Further testing on our algorithms should be done with larger amounts of vertices to model more realistic applications, where there are thousands of vertices that need to be split. In addition, all trees should be included to test the efficiency on not only a particular set of trees with a specific property. That will enable us to also test if there is a practical difference in computational times between the current and proposed algorithms. We have coded our algorithms and are currently conducting some of these tests ourselves.

As we mentioned above, the main problem we ran into was the fact that there are no clear-cut regulations on the number of schools in a district or the number of counties a district can span across. Thus, it makes it hard to create restrictions on that directly. We can try to invent our own partition that limits the number of schools, as well as the population range, so schools are not over-crowded.

Finally, the data from the 2020 Census will be released at the end of this year. With this new data, new census blocks will be created and the analysis can be repeated with much fewer assumptions. Since the 2020 Census data will match the data provided on the Massachusetts Department of Education website, less drastic modifications will need to be made on the data set. In addition, it will provide a more updated version of population distribution that more accurately reflects the current arrangement. Furthermore, this new data can be extracted to identify possible gerrymandering in other areas, such as religion.

7 Acknowledgements

There are many people that we would like to thank for their support. We would like to extend our gratitude to Steve Huang, David Fried, the PROMYS Foundation, and the Clay Mathematics Institute for running this incredible Research program at PROMYS. We also want to express our deepest gratitude to Diana Davis for proposing such an intriguing problem and providing a plethora of resources for us to use throughout our project. In addition, a big thank you to the MGGG group for allowing us to use their packages to code our programs. Finally, we would like to thank our counselor Mirilla Zhu for always being there to provide helpful insight or push us in the right direction.

8 References

[1] Cayley, A., A Theorem on Trees. Quart. J. Math., 23(1889), 376 - 378

[2] Mira Bernstein, Gerrymandering: Why It's More Complicated Than You Might Think, talk at Bryn Mawr College, April 15, 2019.