Jeff McCrea

CSS 600 - Spring 2023

Professor Fukuda

CSS 600 Term Report – Algorithmic Redistricting & Gerrymander Detection

Contents

Overview

In the field of computational political science, one of the more important and burgeoning fields of study is the use of algorithms to create political redistricting plans and evaluate proposed district plans for objective fairness. During my CSS 600 course this quarter, I focused on researching and assessing the feasibility of agent-based algorithmic redistricting and gerrymandering detection in the MASS framework. To evaluate this topic area for potential further research, I conducted a systematic literature review on current research, reviewed specifics related to the topic, reviewed previously implemented frameworks for algorithmic redistricting, and sought to understand algorithmic implementations for generating and evaluating district plans.

In the following sections, I will provide some background information on redistricting and gerrymandering, look at the specifics related to algorithmic redistricting and gerrymandering detection, look at how district plans are evaluated, discuss how gerrymandering can be detected from a computational perspective, and conclude with some potential limitations and future work.

Background

Redistricting Process

Redistricting is the process of redrawing the boundaries of electoral districts in the United States. The process typically follows the decennial census, which updates population data and is used to ensure equal representation based on population shifts. As a state's population changes, the number of seats in the House of Representatives allotted to an individual state can grow or shrink depending on population changes. This process is called reapportionment and serves as the drive behind redistricting.

Redistricting can be a lengthy process comprised of several different steps, but redistricting can generally be distilled down to six key steps:

1.) Census Population Gathering: Every ten years, the United States Census Bureau conducts a nationwide data-gathering operation that asks residents of the United States to report demographic information regarding their family unit. The Census Bureau gathers information regarding racial background, education, health, housing, number of family members, and transportation-based questions. While the amount of information collected is relatively diverse, the pertinent information for redistricting centers around the number of people residing in residence and the racial makeup of the family unit to ensure equal representation.

2.) Reapportionment: Following the tabulation of the census data, the population data is used to determine how many seats each state should have in the House of Representatives. The United States Constitution fixes the number of House seats at 435, and when the population grows in one state but shrinks or stays static in another, this can cause the number of seats allotted to change. For example, in the 2020 reapportionment following the 2020 census, the outflow of people from California caused the state to lose one seat, while the population increase in Texas increased the number of seats by two.

3.) State Redistricting: While the federal government, through the Census Bureau, collects the population data and determines the number of seats allotted to each state, the individual states must redraw the district boundaries within their borders. The specifics of the redistricting process vary from state to state but typically have components regarding population equality between districts, contiguous districts, and compact districts. Some states have different mandates prescribed by their state constitution and Federal mandates, like compliance with the Voting Rights Acts.

Just as the rules for redistricting vary between states, so does the entity that carries out the drawing of districts. In some states, the process is overseen by the partisan state legislature. While this may seem to be a suitable body to carry out the process as the population directly elects them, many have questioned the suitability of having a body with a vested interest in protecting their power drawing districts. In recent years, many states have turned to independent bodies to reduce the amount of partisan influence in the process. Even with independent redistricting commissions, many examples of unfair maps still emerge from those entities.

4.) Public Input: An essential component of the redistricting process is the elicitation of feedback from the public on potential redistricting maps. While many ignore the process due to a lack of interest, many states hold public hearings to allow individual citizens to voice questions or concerns on the potential map. In some states, citizens can even draw and submit their maps. With many citizens lacking the expertise to draw district maps from scratch, some commercial and open-source software packages have emerged to support the drawing of district maps, with some ensuring those maps compile with the redistricting criteria in the state.
5.) Legal Challenges: After a candidate map has passed through the necessary redistricting commission or body, the map is still subject to legal challenges from affected parties. In fact, many of the notable instances of partisan gerrymandering have been discovered through the use of legal challenges. Redistricting maps can be challenged on several grounds, but most challenges boil down to racial and partisan gerrymandering. While we will detail the specifics of gerrymandering in a later section, the process typically involves manipulating district boundaries to favor a specific group or political party. When legal challenges are leveraged against a potential redistricting plan, courts may review the plan to ensure compliance with state and federal redistricting constraints. Many academics involved in research on redistricting have served as expert witnesses during these cases and have used algorithms and statistical analysis to reinforce their position.
6.) Implementation: Once a redistricting plan has cleared all statutory and legal hurdles, the state implements the district bounds specified by the plan for the next election cycle. Voters may find themselves in a newly drawn district with potentially different representation.

The redistricting process can be complex, with the involved parties balancing the need to redraw boundaries with respect to statutory requirements and interest groups with varying objectives. One important element to note is that until relatively recently, the process was ripe for political interference, and the difficulty of the process left little room for citizen oversight. As we will explore in later sections, the rise of computing power and the availability of granular data has opened an avenue to reduce human bias and develop a fairer system for generating redistricting plans.

Gerrymandering

The term "gerrymandering" dates back to the early 19th century when the then-governor of Massachusetts, Elbridge Gerry, approved a redistricting bill that created a district that many thought looked like a salamander. A newspaper mocked the district plan in a political cartoon, shown in figure 1**,** which combined the governor's last name, Gerry, with the word salamander to create the term "gerrymander." While the term was a tongue-in-cheek joke at the time, it has left a lasting legacy on American politics and a stain on the redistricting process.

Today, gerrymandering is commonly defined as the practice of manipulating electoral district boundaries to favor a particular party or group. The corruption of redistricting process involved creating an unfair advantage for one party in elections by concentrating or diluting the voting power of rival groups. The focus is gerrymandering today centers around two main types:

1.) Partisan Gerrymandering: In an ideal form, the redistricting process should produce maps drawn to ensure fair representation for all voting members of society. When maps are drawn to maximize one political party's power intentionally, partisan gerrymandering occurs. This form of gerrymandering is the most commonly identified form when people think about the subject.
2.) Racial Gerrymandering: Unlike partisan gerrymandering, which focuses on maximizing the power of a particular political party, racial gerrymandering seeks to draw districts that dilute the voting strength of racial or ethnic minority groups.

Figure 1 - "The Gerry-Mander" Cartoon [2]

The goal of gerrymandering is to amplify a desired political party's power beyond what they would receive based on the makeup of the voting population. Power amplification typically comes in the form of two complementary methods, packing and cracking, shown in figure 2[20]**.** Packing is the process of diluting the power of one voting block by consolidating a majority of the target voting population into a small number of districts. While this would give the packed party the illusion of proportional representation through overwhelming victories in the consolidated districts, the remaining members would be spread across many districts, known as cracking. In combination, packing and cracking create what are known as gerrymandered districts which insulate the current ruling party from competition and deprive marginalized groups of meaningful representation.

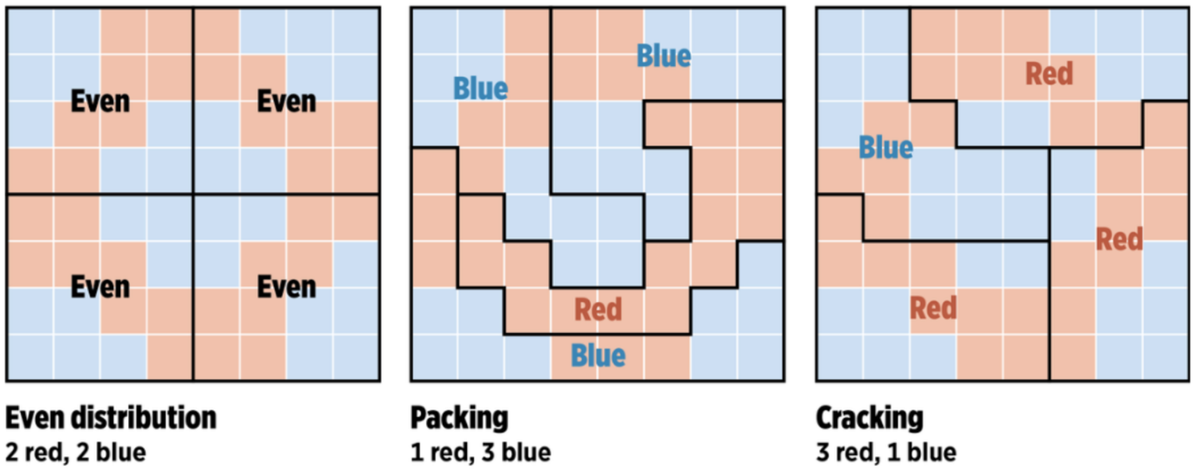

Figure 2 - Packed & Cracked District Maps[8]

Gerrymandering can result in some significant negative consequences. While research has shown that the effects of gerrymandering are largely diluted on the national level, the impacts can be much more present on a state or local level. The effects of gerrymandering can be distilled into a few key impacts. First, it manipulates representation. Electoral results can be distorted by denying the creation of districts that accurately reflect the electorate's overall preferences. Without a legislature that accurately reflects the composition and will of the voters it is supposed to represent, the outcome of elections in gerrymandered districts is inaccurate. Second, gerrymandered districts greatly reduce the competitiveness of the elections. One of the primary motivations in partisan gerrymandering is to create effectively "safe seats" for incumbents where only a particular party can win. Safe seats reduce competition and discourage potential challengers, as there can be an insurmountable partisan barrier to winning the election. Third, and particularly pertinent to racial gerrymandering, is disenfranchisement. By creating districts that dilute the voting power of a group or party, those voters are left without equitable representation and thus are prevented from having their voices heard. Finally, gerrymandered districts serve to amplify political polarization. Suppose a politician or political party is overwhelmingly dominant to the point that competitiveness in the district is nonexistent; there is little incentive for a candidate to appeal to voters outside of their political base.

While states have made efforts to create a more neutral redistricting process through means like independent redistricting commissions, stringent district criteria, and other methods, those efforts still leave the process open to human bias and partisan influence. To address the issue of gerrymandering, researchers have proposed empirical methods to remove partisan influence and generate redistricting plans based on geographical and voting population requirements [9].

Algorithmic Redistricting

The injection of human bias into the redistricting process has long been a point of consternation when drawing and assessing a potential district map. Civic-minded academics have long sought to remove this bias and create neutral districting plans with algorithms serving as unbiased arbiters in redistricting. Even with the overwhelming optimism that algorithms would fix what many regarded as a broken redistricting system, limitations in computational capacity and data granularity stymied much of the development in the early days.

Algorithmic redistricting dates back to the 1960s, with works from people like Vickrey heralding algorithms as the key to "procedural fairness" in the redistricting process [13]. Vickrey felt that by using algorithms, the redistricting process could remove the human element and create a process that was solely machine-driven in nature, thus creating a process that left no room for human choice and political bias. Another scientist named Edward Forrest [6] echoed Vickrey's sentiments, stating, "Since the computer doesn't know how to gerrymander … the electronically generated map can't be anything but unbiased". While the sentiment expressed by the two authors may have been too idealistic, as a computer can clearly be programmed to generate a gerrymandered map, even ones that may be incredibly difficult to detect, the consensus among many in the emerging field is that algorithms can be a crucial component in drastically reducing the amount unfairness in the redistricting process.

In today's modern computing environment, the algorithmic ideas of early academics have evolved into a wealth of different software packages designed to support redistricting efforts across the United States. Redistricting entities have utilized these software packages to design and analyze districting plans. When coupled with increased computing power and data granularity, these software packages can create highly detailed and objectively fair maps.

While the overarching goal of Vickrey and Forrest to create "perfect" districting plans may be technically intractable, recent progress has made algorithmic redistricting an essential element to reforming the redistricting process [15]. This section explores how computing resources have been leveraged in redistricting and surveys the various redistricting algorithms utilized in modern algorithmic redistricting. First, we will look at algorithms used to generate district plans and follow up with methods used to analyze a potential/proposed plan.

Defining the Redistricting Problem

Across many works in the field of algorithmic redistricting, the process of redistricting can be distilled to a set partition problem with additional restrictions. Let $G = (V,E)$ be a connected graph representing a given state's geography. For all vertices $v \in V$, v represents a population unit with weight $w(v)$ equal to the unit's population [8]. Population units can come in different forms and are, in some cases, specific to the redistricting problem. The smallest population unit denomination, typically census blocks, provides more granularity and accuracy when generating district plans at the cost of increased computational complexity. Vertices $v,w \in V$ share an edge if the associated population units' polygonal boundaries are adjacent, meaning they share at least a vertex.

The goal is to partition G into k connected subgraphs, each representing a district. The districts, denoted as G_i, where $Gi = (Vi, Ei)$ and $Vi \subset V$, $Ei \subset E$, should be created such that the total population of any district population(G_i) = $\sum v \in Vi \; w(v)$ is equal to the desired population L. The load capacity is the desired number of people per district under perfectly even representation [8]. The population deviation from L is the important criterion for judging the solution's efficacy to the redistricting problem and how suitable the district plan is. While there are many different ways to judge the efficacy of a redistricting plan as we will see in later sections, population distribution and the deviation of the population between districts serve as the baseline metric.

What is a redistricting algorithm:

At a high level, an algorithm is a set of instructions that a computer uses to solve a given problem. Usually, the algorithm takes in some amount of input data to produce an output solution. In the realm of redistricting, a redistricting algorithm may take as input the populations and geographies of some unit of measure, typically in the form of precincts, census tracts, census blocks, counties, etc., as well as the desired number of districts and produces an output comprised of districts, which designates which units are assigned to each created district.

While the computer executes the algorithm, humans and their potential biases write the instructions and create the algorithmic design decisions. This fact leads to the question of what makes a quality district plan, and since the computer can't identify what makes a good district plan on its own, humans must first define what makes one plan superior to another. The discussion of what qualifies an objectively good district is a broad and still raging debate. This paper focuses solely on the basic requirements: population equality between districts and compact district boundaries. In the field of algorithmic redistricting, there are three main categories of algorithms: enumeration, partitioning, and swapping.

Enumeration

Enumeration algorithms, as the name might imply, focus on generating every possible way to district a given region. The main advantage of enumeration algorithms is that they have the potential to generate every possible way to divide a region into a set number of districts. When reviewing the aspirations of people like Vickrey and Forrest, they believed these algorithms would solve the problem of human biases in redistricting. If every district plan is available for review, it is possible to identify the best based on the idealized set of criteria. While this may seem like a suitable method for generating district plans in theory, a practical look at even a small region can result in an obscenely large number of possible plans.

A natural approach to designing an algorithm for redistricting is to generate all possible valid district plans. This technique essentially boils down to utilizing a set of redistricting constraints, like population equality and contiguous districts, to identify valid district plans and having the computer generate a list of every possible plan that meets those constraints.

If we could enumerate all possible districts, the redistricting problem is a relatively straightforward optimization algorithm: score all possible plans to identify the best one; essentially, a brute force algorithm [15]. This approach achieves the idealized "perfection" quality Vickrey and Forrest hoped would solve the problem of biases in redistricting because it considers all possible plans and is guaranteed to find the best one. Given this theoretical advantage when compared to other approaches, enumeration has been proposed as a strategy to identify and evaluate plans for decades.

This begs the question: if enumeration can yield an exact result, why is it not widely used for redistricting today? There is one main reason: combinatorial explosion [15]. While in extremely small simulations, enumeration may be feasible, but the vast number of ways we can draw district lines makes the list of potentially valid plans incredibly large. As the redistricting region grows, the problem increases exponentially, quickly exceeding the practical limits of computing power and data storage. Even with the advancements in supercomputing, the approach is still unfeasible.

To solidify the intractability of the enumeration approach, consider the simple problem of partitioning an *n* x *n* grid into n equal-sized districts. As shown in figure 3, as the n value increases, the number of plans explodes relatively quickly, becoming too large for enumeration to be practical. When looking at real-life examples, the combinatorial explosion bears out. For example, the number of ways to build four congressional districts out of 99 counties of Iowa is estimated to be around 10^{24}. While the estimate for Iowa may seem large, the problem grows even larger if plans are generated from finer-grain units like precincts or census blocks, which are more commonly used for redistricting simulation. For reference, Iowa law requires that counites be used as the base unit for redistricting; if census blocks were used, this would require generating plans from 216,007 units rather than 99 units.

n	# plans
1	1
2	2
3	10
4	117
5	4,006
6	451,206
7	158,753,814
8	187,497,290,034
9	706,152,947,468,301

Figure 3 - N x N Number of District Plans [15]

Since enumeration has been demonstrated to be computationally infeasible and inappropriate for understanding real-world redistricting problems, researchers have turned to other strategies for generating and assessing district plans.

Partitioning

As the name implies, partitioning algorithms take a geographic region comprised of population units, such as a state, and divides them into a set number of contiguous districts. In most implementations, the partition algorithms fall into greedy and recursive approaches, each with different variations.

Among the greedy redistricting algorithms, many are based on Voronoi approaches. The choice of Voronoi-based approaches is driven mainly by the fact that Voronoi diagrams produce a fixed number of contiguous and compact district partitions – a noted crucial element in generating appropriate district plans. Voronoi-based redistricting algorithms work off of a set of centers to assign population units to the closest open center in the set that has reached a given population threshold. As noted by Levin, the section of centers significantly impacts the resulting district plans [8]. In Svec et al.'s implementation, Svec created centers based on the largest population units resulting in a population deviation of 0.74% between generated districts [5]. Another implementation by Ricca et al. used centers based on minimizing the longest path from a center to the most distant population unit in its district [14]. That approach resulted in significantly worse population deviation between candidate districts. Svec et al.'s approach demonstrated a superior approach based solely on population deviation, but it came at the cost of potentially partitioning a single house or other building into separate districts.

In addition to greedy-based Voronoi approaches, k-means-based algorithms are present in the field of algorithmic redistricting. K-means algorithms draw heavily on Voronoi approaches but with the ability to move centers after each iteration. As with the previous approaches, the section of centers can significantly impact the algorithm's output. In a study on the state of North Carolina, random initial centers with the k-means algorithm produced population deviations greater than 150%, far beyond what is acceptable. Bottman et al. used centers based on the current district layout and produced an output with a 2.5% population deviation [21]. The significant variance in the population deviations demonstrates the importance of initializing centers with purpose. While the selection of centers is important, another aspect to consider is the impact the size of a given population unit may have. As data granularity related to redistricting has increased, the types of population units available to researchers have also broadened. In a study produced by Cohen-Addad et al., centers were selected randomly, but census blocks were utilized instead of using census tracts. Census blocks and tracts are both geographic units used by the Census Bureau to collect and report population data, but census blocks are the smallest geographic unit used. Even with the use of random centers, which was demonstrated to be ineffective, Cohen-Addad's paper shows that even with random centers, the granularity of census blocks allowed for the production of district plans with incredibly small population deviation in six of the six states studied [10]. While Cohen-Addad's algorithm showed the utility of census blocks as a population unit, the districts generated were not as compact due to the census block's geometry being represented by a single point.

While Voronoi algorithms assign units to districts based on the distance to the closest center, Constrained Polygonal Spatial Clustering (CPSC) uses an objective function to weigh a set of criteria. CPSC works by dividing the area into census blocks and then using a Voronoi diagram to create initial districts based on the population density of each block [8]. The algorithm works iteratively to improve the district plan by adjusting the boundaries of the districts while satisfying any constraints. While a Voronoi algorithm can generate the initial district plans, existing plans can also be a starting point. This aspect presents a unique feature of the CPSC algorithm in that it can consider the existing political boundaries and attempt to respect them as much as possible. Joshi et al. implemented a CPSC algorithm with an objective function, requiring population deviation to be within 1%, that districts be contiguous and maximize district compactness [16]. The results produced by the study demonstrated that this approach could generate multiple district plans with population deviations within 1%.

Recursive algorithms are also effective for generating district plans because they break the redistricting problem into smaller subproblems. Among the recursive redistricting algorithms, the shortest split-line algorithm (SSL) is the most common [8]. The shortest-split line algorithm utilizes a divide-and-conquer approach, where the goal of every iteration is to find the shortest line that cuts the geographic region into two pieces such that these pieces contain equal portions of the population. Essentially, the algorithm works by dividing the area into a grid of cells and bisecting each cell along its shortest dimension. After which, the algorithm selects the line that minimizes the maximum distance between any two points with a district and splits the district along that line. SSL is an iterative process that continues until the desired number of districts are created. SSL does come with some tradeoffs. While the algorithm is fairly straightforward to implement and efficient, it does not consider existing political boundaries or other contextual factors that may be important in specific scenarios. In addition, as the size of the geographic region increases, the computational complexity increases significantly. Benn and German and Kalcsics et al. utilized similar SSL algorithms to generate postal districts and found that while the population deviations were within acceptable limits, the time complexity of the algorithm was $O(n^3)$ and higher for even small sample sizes[18,19].

The Diminishing Halves Algorithm (DHA) is another dive-and-conquer algorithm similar to SSL. The DHA works by recursively dividing a geographic area into two halves, with the goal being that each half has a roughly equal population [8].

Swapping

Swapping algorithms start with an existing redistricting plan based on the desired criteria. The algorithms evaluate population units on the boundary of two districts and swap boundary units between neighboring districts. Swapping algorithms can be further distinguished by how swaps are constrained. Local search swapping allows for swaps that will improve a given district, while metaheuristic swapping allows some detrimental swaps in the short term that could create a better plan later on. Distinct from partitioning algorithms, local search and metaheuristic swaps require a complete redistricting plan as input.

A local search swapping algorithm only reassigns population units between districts if the swap improves the plan. In this approach, the existing district boundaries are adjusted by swapping pairs of contiguous areas to improve specific objectives, most commonly population deviation or compactness. Each swap is evaluated by calculating the resulting change concerning the desired objective, and only swaps that improve upon the district plan are accepted. Local search swapping has been utilized in three major studies by Kaiser, Nagle, and Hayes for county-level redistricting. Kaiser produced a congressional redistricting plan for Illinois with a 26.4% population deviation, while Nagle, also working with Illinois but a smaller subset of the state, produced plans with a 4% population deviation. Hayes utilized local search swapping to generate redistricting plans in North Carolina with a similar 4.4% population deviation. A significant drawback with local search swapping is the reliance on the layout of the current districts. If the current district plan is a suboptimal starting point, local search swapping can exhaust beneficial swaps before producing a valid or optimal plan.

In order to mitigate the shortcomings of local search swapping, other works have focused on allowing detrimental swaps to avoid exhausting beneficial swaps and settling at a local optimum. Metaheuristic swapping allows more of the sample space to be explored by allowing swaps that worsen the output. Similar to local search swapping, the algorithms involve iteratively modifying an existing district plan and evaluating the results to see if it is an improvement of the previous one, with the caveat that the algorithm occasionally allows changes that make the output worse. An illustrative example of this process is simulated annealing. Simulated annealing introduces non-deterministic random swaps, which can increase population deviation to improve other objective criteria. In the previously mentioned paper by Hayes regarding redistricting in North Carolina, simulated annealing reduced the population deviation from 4.4% to under 1%.

Tabu search is another metaheuristic swapping algorithm that is present in the field of algorithmic redistricting. Unique to this approach is the ability to cut down on redundant district plans by avoiding previously explored solutions. Tabu search introduces "tabu" paths that make the result worse initially to explore different areas of the solution set. Bozkaya et al. implemented a tabu search algorithm to generate district plans for Edmonton, resulting in a population deviation of around 1%. While the results produced by Bozkaya appear promising, no existing studies extend this technique to the scale of states in the U.S[1].

Another form of metaheuristics swapping is genetic algorithms. Genetic algorithms combine features of high-scoring plans from an objective function, mimicking the evolution process of natural selection. [8]. Josh et al. utilized a genetic algorithm to tackle the redistricting problem in Nebraska and Indiana, resulting in population deviations that exceeded the legal threshold [16]. Joshi utilized a method that used a weighting function to combine constraints into an objective function. Vanneschi et al. followed with an NSGA-II technique that focused on optimizing compactness and population equality and was able to produce population deviations under the legal threshold [12]. While the results seemed promising, as we will note in the next section, the simulated states were relatively small in terms of population units and lacked extension to larger, more complex states. Baas et al. extended the work of Vanneschi to include contiguity, political competitiveness, and proportionality in the objective functions as well as utilizing current district bounds as seeds for new plans [7]. Baas showed results that met the legal population deviation requirement, but the weights used in the objective function and the number of generations required to generate legal plans were left unspecified.

Limitations of Algorithmic Redistricting

Algorithmic redistricting presents a promising field of study and an objective way to create and identify redistricting plans with minimal human bias. As developments in the field are relatively recent and the methods utilized are not entirely validated across different studies, some significant limitations and areas of concern exist.

First, there is no universal criteria base for generating redistricting plans. As was mentioned in a previous section, the criteria for generating district plans vary greatly from state to state. Certain states may only have minimal constraints on what constitutes a valid district plan, such as the minimal requirements for population equality between districts and contiguity. In contrast, others can have specifics relating to partisan fairness and minority-majority district composition. With the wildly varying constraint criteria, what constitutes a “valid” plan is still up for debate.

Second, in addition to the subjectivity of the different criteria, balancing the desired criteria in simulation is another challenge. The question of what constraint weight should be has yet to be decided and is reflected in the vastly different weights applied to the constraints among the various studies. It is challenging to generate a consensus without an agreed-upon importance of the different constraints.

Third, many experiments conducted in the field of algorithmic redistricting lack transparency. Many studies in the field do not demonstrate effectiveness in all states and usually include a closed-source code base and are thus unverifiable. Reviewing how an algorithm is constructed, how various constraints are weighted, and the ability to reconstruct the experiment to validate the finding is paramount to further development in the field.

Fourth, in between the decennial census, demographics in the United States can change dramatically. As demographics change, so would the results of a potential redistricting plan. The shifting demographics make studying redistricting difficult to validate outside of the census period. While older data can be used in theory, the ability to work with up-to-date, real-life data is essential to creating algorithms that reflect reality on the ground.

Finally, in many studies, the potential for political interference remains a paramount concern. How an algorithm is constructed to respond to various redistricting constraints can significantly impact the output plan. While many studies purport to seek a politically neutral result, the lack of transparency in some of the papers leads to concerns about the results.

Evaluating Redistricting Plans

When redistricting plans are created through algorithms or hand-drawn, they must be evaluated to ensure they meet all necessary state and federal constraints. Constraints on redistricting plans vary widely from state to state, with each state prioritizing certain constraints over others. Washington, for example, lays out requirements in Article 2, Section 43 of the state constitution as well as in the Revised Code of Washington (RCW) [3] that districts must be:

1.) Be contiguous
2.) Have equal population
3.) Be geographically compact
4.) Preserve county and municipality boundaries as much as possible
5.) Not be connected across geographic barriers, although ferries across water may establish contiguity
6.) “Provide fair and effective representation and ... encourage electoral competition.”

Given these unique and varying state and geographical requirements, a one size fits all approach to generating redistricting plans is not feasible. Instead, generating redistricting plans requires careful consideration of a multitude of different data points, including demographics, political affiliation, geography, candidates, and other state and federal-specific factors.

While the legal requirements in each state vary, there are some common standards used in the field of redistricting:

1. *Population* Equality: refers to the fundamental principle in redistricting that districts must be drawn to have roughly equal populations. This standard stems from the "one person, one vote" principle that by having equal populations, each person's vote carries equal weight.
2. *Contiguity* is another fundamental principle in modern redistricting that all parts of a district must be physically connected. When districts are formed, it should be done in a way that avoids districts that are separated by an unreasonable distance or are fully surrounded by other districts.
3. *Compactness:* is an element closely linked to contiguity and focuses on the geographic shape of the district. In this constraint, districts must be geographically compact. As popularized by the political cartoon of the 1800s, where gerrymandering derived its namesake, the district should not meander through a state but should be geographically succinct. Some studies have purported that compact districts facilitate better representation as they are not as difficult for a representative to navigate.
4. *Communities of Interest*: deals with groups of people who share common social, cultural, or geographic characteristics. This criteria reasons that people who share similar characteristics, particularly across racial or cultural boundaries, will have common concerns or political interests and should be dealt with particular concern during redistricting. Splitting these communities of interest could have negative repercussions on that particular group. While this standard has received some attention in the field of redistricting, it is difficult to quantify its importance in terms of a weighted criteria. In addition, many states do not consider this specific requirement when generating district plans.
5. Partisan Fairness: also known as political fairness, considers political party affiliation when drawing district bounds. Partisan fairness aims to ensure that a redistricting plan does not unfairly disadvantage one political party over the other.

Detecting Gerrymandering

As mentioned in the background section, gerrymandering has a long history of malicious use in the United States. While many previous methods for detecting gerrymandering relied on a visual and numerical inspection of the districts in a proposed redistricting plan, the rise of computational methods for detecting gerrymandering has seen significant attention in recent years.

As with many redistricting aspects, no one agreed upon way to detect gerrymandering exists. Instead, research has focused on analyzing potential plans for statistical anomalies. In the field of gerrymandering detection, there are five key metrics for identifying a plan that may have elements of gerrymandering:

1. *Efficiency Gap Analysis*: assesses the degree to which potential partisan gerrymandering has occurred in a candidate plan. Efficiency gap analysis quantifies one political party's advantage or disadvantage over the other. Efficiency gap analysis relies on measuring wasted votes, which looks at lost votes (votes cast for a candidate that did not win) and surplus votes (votes cast for a candidate over what is needed to win the election).
2. *Partisan Symmetry Analysis*: similar to the efficiency gap, assesses potential partisan gerrymandering by quantifying the relationship between a political party's vote share and the number of seats won in an election. Primarily, this metric seeks to measure the imbalance between a party's vote share and the number of seats won in an election to determine if one party has a significant advantage over the other.
3. *Geographic analysis:* different from the previous metrics, geographic analysis uses spatial characteristics of districts to detect signs of intentional manipulation. To score a potential district, geographic analysis uses metrics like compactness, contiguity, and previous district boundaries.
4. *Demographic analysis:* focuses on assessing a plan's impact on specific communities and quantifies the potential vote dilution or concentration. Using demographic data on attributes like racial populations, socioeconomic factors, and other demographic data points, demographic analysis looks for intentionally manipulated population distributions to gain an advantage for a particular party.
5. *Simulation and Outlier Detection*: utilizes a computer simulation to generate multiple scenarios based on a potential district plan and uses the results to determine if the plan generates results significantly different from other plans. Outlier detection typically requires a wealth of plans generated based on neutral criteria and determines if a proposed plan would create an outsized victory or loss for one party over another. Outlier detection represents one of the more prominent and cutting-edge approaches featured in legal cases involving gerrymandering, but some experts disagree on its effectiveness.

Limitations/Future Work

When looking at limitations in assessing the feasibility of implementing redistricting simulation and redistricting plan evaluation in an agent-based environment like MASS, I've come across a few limitations. First, there exists no body of work utilizing agents to increase algorithm execution speed or increase accuracy. While this limitation does not mean it's not possible, it may result in, at worst, an inefficient or ineffective solution or will only maintain current accuracy performance while taking on more significant performance overhead.

Second, as mentioned in the algorithmic redistricting limitations, there isn't a well-defined consensus on what redistricting criteria are the most effective or proper for generating redistricting plans. For a baseline implementation, population equality and contiguous district creation would be the likely constraints, resulting in a solution that does not cover all states.

Third, previous work in MASS has demonstrated its effectiveness for algorithms like K-means and Voronoi approaches. While those have been utilized for redistricting simulation, they represent some of the more basic and early algorithms. Those algorithms would generate valid solutions with the previously mentioned baseline constraints but may not be as academically worthy as more advanced approaches.

Moving forward, I am focused on two areas for finishing my preliminary research on the topic. First, I want to design and test a simplified experiment in MASS to gauge if further work is worthwhile. This experiment would most likely take the form of a small graph with synthetic population data to test potential algorithms and validate both the accuracy of the solution and the speed of execution. Second, I want to evaluate the suitability of MASS to be used to evaluate a potential district. Using the evaluation metrics previously listed, I want to see if evaluation is suitable for an agent-based approach. While I conducted a significant amount of research in algorithmic redistricting, I still need to spend more time assessing the feasibility of algorithmic redistricting and evaluation as a potential capstone research topic.

References

[1] B. Bozkaya, E. Erkut, and G. Laporte, “A tabu search heuristic and adaptive memory procedure for political districting,” *European Journal of Operational Research*, vol. 144, no. 1, pp. 12–26, Jan. 2003, doi: 10.1016/S0377-2217(01)00380-0.

[2] A. A. P. N. published yet DOI, “ALARM Project: The Algorithm-Assisted Redistricting Methodology (ALARM) Project,” *ALARM Project*. https://alarm-redist.org/about.html (accessed Mar. 16, 2023).

[3] A. 50-S. S. P. T. P. July 25 and 2022 Doi 10.7910/Dvn/Slcd3e, “ALARM Project: Washington Congressional Districts,” *ALARM Project*. https://alarm-redist.github.io/fifty-states/WA_cd_2020/ (accessed Mar. 16, 2023).

[4] “An Objective Method for Establishing Legislative Districts on JSTOR.” https://www.jstor.org/stable/2109148 (accessed Mar. 12, 2023).

[5] L. Svec, S. Burden, and A. Dilley, “Applying Voronoi diagrams to the redistricting problem.,” *UMAP*, vol. 28, no. 3, pp. 315–331, 2007.

[6] “Apportionment by Computer - Edward Forrest, 1964.” https://journals.sagepub.com/doi/10.1177/000276426400800407

[7] K. Baas, “Auto-Redistrcit,” 2016. http://autoredistrict.org/contact.php

[8] H. A. Levin and S. A. Friedler, “Automated Congressional Redistricting,” *ACM J. Exp. Algorithmics*, vol. 24, pp. 1–24, Dec. 2019, doi: 10.1145/3316513.

[9] B. Fifield, Higgins Michael, K. Imai, and A. Tarr, “Automated Redistricting Simulation Using Markov Chain Monte Carlo,” *Journal of Computational and Graphical Statistics*, vol. 29, no. 4, pp. 715–728, Oct. 2020, doi: 10.1080/10618600.2020.1739532.

[10] V. Cohen-Addad, P. N. Klein, and N. E. Young, “Balanced power diagrams for redistricting.” arXiv, Jan. 07, 2018. doi: 10.48550/arXiv.1710.03358.

[11] B. Hayes, “Computing Science: Machine Politics,” 1966. http://bit-player.org/wp-content/extras/bph-publications/AmSci-1996-11-Hayes-redistrict/compsci96-11.html

[12] L. Vanneschi, R. Henriques, and M. Castelli, “Multi-objective genetic algorithm with variable neighbourhood search for the electoral redistricting problem,” *Swarm and Evolutionary Computation*, vol. 36, pp. 37–51, Oct. 2017, doi: 10.1016/j.swevo.2017.04.003.

[13] W. Vickrey, “On the Prevention of Gerrymandering,” *Political Science Quarterly*, vol. 76, no. 1, pp. 105–110, 1961, doi: 10.2307/2145973.

[14] F. Ricca, A. Scozzari, and B. Simeone, “Political Districting: from classical models to recent approaches,” *Ann Oper Res*, vol. 204, no. 1, pp. 271–299, Apr. 2013, doi: 10.1007/s10479-012-1267-2.

[15] M. Duchin and O. Walch, Eds., *Political Geometry: Rethinking Redistricting in the US with Math, Law, and Everything In Between*. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-319-69161-9.

[16] D. Joshi, L.-K. Soh, and A. Samal, “Redistricting Using Constrained Polygonal Clustering,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2065–2079, Nov. 2012, doi: 10.1109/TKDE.2011.140.

[17] “Simplified Bipartisan Computer Redistricting on JSTOR.” https://www.jstor.org/stable/1226994 (accessed May. 12, 2023).

[18] J. Kalcsics, S. Nickel, and M. Schröder, “Towards a unified territorial design approach — Applications, algorithms and GIS integration,” *Top*, vol. 13, no. 1, pp. 1–56, Jun. 2005, doi: 10.1007/BF02578982.

[19] A. Benn, “Unbiased Congressional Districts,” 2008. Accessed: May. 16, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Unbiased-Congressional-Districts-Benn/427f9359362ef36144498b2d0213abb7b73302e8

[20] K. Imai, “Using Simulation Algorithms to Detect Gerrymandering: Evaluation of the 2022 Congressional Maps”.

[21] N. Bottman, W. Essig, and S. Whittle, “Why Weight? A Cluster-Theoretic Approach to Political Districting,” *Range Voting*, no. 1036, Feb. 2007, [Online]. Available: https://rangevoting.org/MCM_2007Smaller.pdf