Stochastic Study of Gerrymandering

Christy Vaughn

A thesis submitted to the Department of Mathematics for honors

Duke University

Durham, North Carolina

2015

Abstract

In the 2012 election for the US House of Representatives, only four of North Carolina's thirteen congressional districts elected a democrat, despite a majority democratic vote. This raises the question of whether gerrymandering, the process of drawing districts to favor a political party, was employed. This study explores election outcomes under different choices of district boundaries. We represent North Carolina as a graph of voting tabulation districts. A districting is a division of this graph into thirteen connected subgraphs. We define a probability distribution on districtings that favors more compact districts with close to an equal population in each district. To sample from this distribution, we employ the Metropolis-Hastings variant of Markov Chain Monte Carlo. After sampling, election data from the 2012 US House of Representatives election is used to determine how many representatives would have been elected for each party under the different districtings. Of our randomly drawn districts, we find an average of 6.8 democratic representatives elected. Furthermore, none of the districtings elect as few as four democratic representatives, as was the case in the 2012 election.

1 Introduction

Every two years, elections are held to elect the US House of Representatives. Each member of the House represents a congressional district, or a contiguous region of land within a state. Every 10 years, the census is used to determine how many representatives, and thus congressional districts, each state is allotted. After every census, each state must redraw the boundaries of their congressional districts. The North Carolina General Assembly [4] lists 5 rules for drawing districts:

One Person Must Equal One Vote
Consideration of Minorities
Impermissible Consideration of Race
Districts Must Be Contiguous
Division of Counties Must Be Minimized

With these rules in mind, there are several ways to draw the congressional districts. Different congressional district boundaries can lead to different election outcomes.

In the 2012 election for the US House of Representatives, republicans took 33 more seats than democrats, despite 1.4 million more democratic votes. In North Carolina, 2,218,357 democratic votes and 2,137,167 republican votes were cast. Despite the larger democratic vote, only 4 of North Carolina's 13 congressional districts elected democrats. This raises the question of whether gerrymandering was used during the redistricting process. **Gerrymandering** is the drawing of electoral district boundaries in such a way as to give one party an unfair advantage. The results for the 2012 election in North Carolina may have been the result of gerrymandering, or possibly an unintentional result of the geographic positioning of these votes. This lends to the following question: Given a set of votes, what is the likelihood of electing a certain number of representatives of a given party? If the observed number of representatives is extemely atypical, then this suggests that the results were intentional. This study aims to quantify how atypical electing only four democratic representatives was for the 2012 election, given the votes that were observed.

2 Probability Distribution for Number of Representatives

To determine the likelihood of electing a given number of representatives of one party, we first place a probability distribution on the space of possible districtings of North Carolina into 13 congressional districts. By using the real election data from the 2012 election, we can determine the number of representatives from each party that would have been elected for a given districting of North Carolina. This allows us to extend the probability distribution on districtings of North Carolina to a probability distribution for the number of representatives elected for each party. This distribution can be used to quantify how atypical an election result is, given the votes that were cast.

To place a probability distribution on the space of possible districtings, we begin by discretizing the state of North Carolina into a graph G with vertices V and edges E. Each vertex $v \in V$ is a Voting Tabulation District (VTD) and an edge $e = (v, v') \in E$ if VTDs v and v' are adjacent on the map of North Carolina. A districting is a division of G into 13 distinct connected subgraphs. Thus, we define a districting as a map $\xi : V \to \{1, 2, ..., 13\}$. We let $D_i(\xi) = \xi^{-1}(i)$ be the ith district in districting ξ. Since the districts must be contiguous regions of land, we require that ξ be such that each $D_i(\xi)$ is a connected subgraph of G.

Our probability distribution on possible districtings should assign larger probability to districting plans which are more plausible. Ideally, zero probability should be assigned to districting plans which are not legal, such as districting plans which are not in compliance with the Voting Rights Act, or do not follow one of the 5 rules listed by the North Carolina General Assembly, which are cited in the introduction. However, many of these considerations for drawing district boundaries are not quantifiable and up to interpretation.

To keep the definition of our probability distribution straightforward and simple, we only take into account two considerations: 1) how evenly distributed the population is among the 13 districts, to support the notion of *one person equals one vote* and 2) how compact the districts are, to support the general opinion that congressional districts, which are intended to represent communities, should not be allowed to meander throughout the state. Thus, we define the following energies to evaluate the plausibility of a districting ξ:

- J_{pop}, or Population Energy: evaluates how evenly divided the population is amongst the districts. A lower J_{pop} is desired, meaning that the population is spread very evenly amongst the districts. J_{pop} is given by:

$$J_{\text{pop}}(\xi) = c_{\text{pop}} \sum_{i=1}^{13} \left(\text{Pop}(D_i(\xi)) - \frac{\text{Pop}}{13} \right)^2 \tag{1}$$

  where $\text{Pop}(D_i(\xi))$ is the population of congressional district i, Pop is the population of NC, and c_{pop} is a scaling parameter.

- J_{com}, or Compactness Energy: evaluates the compactness of congressional districts. There are multiple ways to define compactness of a region. We define compactness as a relation between a district's perimeter to its land area. Since the compactness of a region should not depend on the area of the region, and since area is proportional to perimeter squared, we define J_{com} as the following:

$$J_{\text{com}}(\xi) = c_{\text{com}} \sum_{i=1}^{13} \frac{\text{Per}(D_i(\xi))^2}{\text{A}(D_i(\xi))} \tag{2}$$

  where $\text{Per}(D_i(\xi))$ is the perimeter of district i, $\text{A}(D_i(\xi))$ is the area of district i, and c_{com} is a scaling parameter.

From these two energies, we define the total energy, J, of a districting as:

$$J(\xi) = \lambda J_{\text{pop}}(\xi) + (1 - \lambda) J_{\text{com}}(\xi) \tag{3}$$

where the parameter $0 \leq \lambda \leq 1$ allows us to vary which of these two energies is cared about the most. A smaller J means that a districting is more desirable according to the chosen energies. Now that we have an energy function that evaluates districtings, we define the measure of a districting, D_i, as the following:

$$P(\xi) = \frac{e^{-\beta J(\xi)}}{\mathcal{Z}} \tag{4}$$

The parameter $\beta > 0$ is an inverse temperature which varies the steepness of the measure. A larger β more drastically penalizes districts with a larger energy. The constant $\mathcal{Z}$ is the integration constant necessary for P to be a probability distribution:

$$\mathcal{Z} = \sum_{\xi} e^{-\beta J(\xi)} \tag{5}$$

Since a smaller J means a districting is more desirable, more desirable districts will have a larger probability.

2.1 Rerunning Elections

We have defined a probability distribution on the space of districtings of North Carolina. By rerunning the 2012 election on each possible districting, we could extend this probability distribution on districtings to a probability distribution on the number of elected representatives from each party.

To guess what the result of the 2012 election would be for a different districting, we make the assumption that a voter votes for a party, rather than a candidate. Thus a voter who voted for a democratic candidate will vote for a democratic candidate, regardless of what congressional district race they are voting in. This is clearly a simplification, but it is a plausible approximation and can still be used to gain insightful results.

With this approximation, we define the number of democratic representatives elected under districting ξ as

$$\#\text{Democratic Representatives}(\xi) = \#\{i : \text{votes}_{\text{dem}}(D_i(\xi)) > \text{votes}_{\text{rep}}(D_i(\xi)\} \tag{6}$$

where $\text{votes}_{\text{dem}}(D_i(\xi))$ is the number of democratic votes in district i and $\text{votes}_{\text{rep}}(D_i(\xi)$ is the number of republican votes in district i. Election data is available from the North Carolina Board of Elections [3].

If we could enumerate all possible districtings of North Carolina, then we could calculate the measure of each districting and the number of representatives elected from each party for each districting. This would allow us to arrive at a probability distribution for the number of representatives elected from each party over plausible districtings.

3 Estimating the Probability Distribution on Number of Representatives

In North Carolina, there are about 2750 VTDs. If we ignore the contiguity requirement, the number of possible districtings at the level of VTDs is about $13^{2500} \approx 7.2 \times 10^{2784}$. This is by far larger than the number of atoms in the universe. Although there are significantly fewer districtings with the contiguity requirement, it is not feasible to enumerate all such districtings. Thus, it is computationaly infeasible to calculate the integration constant, $\mathcal{Z}$, defined in equation 5. This motivates a need for a more efficient way to understand this probability distribution.

3.1 Metropolis-Hastings Algorithm

Rejection sampling algorithms can be used to sample from some distribution when the distribution is not known. One rejection sampling algorithm is the Metropolis-Hastings algorithm, a variant of Markov Chain Monte Carlo.

Suppose you have some Markov Chain, meaning that you have states x and transition probabilities $Q(x'|x)$ such that the probability of transitioning to state x' only depends on the current state, x. Suppose that there is a unique equilibrium probability of being at state x_i, given by $P(x_i)$. The Metropolis-Hastings algorithm can be used to sample from this equilibrium distribution. The Metropolis-Hastings algorithm is the following:

1. Pick some initial state x_0.
2. Propose a new state x' with probability $Q(x'|x_t)$
3. Calculate $a = \frac{P(x')Q(x_t|x')}{P(x_t)Q(x'|x_t)}$.
4. If $a \geq 1$, then $x_{t+1} = x'$.
5. Else, $x_{t+1} = x'$ with probability a, and $x_{t+1} = x_t$ with probability $1 - a$.

After a burn-in period, the states x_t are distributed as $P(x)$.

3.2 Application of Metropolis-Hastings

For this project, we used the Metropolis-Hastings algorithm to sample our distribution P on the space of districtings. Each state of the Markov Chain is a possible districting ξ. Our initial state is the current districting that was established after the 2010 census. To propose a new districting:

1. Uniformly select a conflicted edge at random. A **conflicted edge** is an edge $e = (u, v) \in E$ such that $\xi(u) \neq \xi(v)$, or in words, u and v are adjacent, but they are not in the same congressional district.

2. For the conflicted edge $e = (u, v)$, with probability 1/2, either let the proposed districting ξ' be

$$\xi'(w) = \begin{cases} \xi(w) & w \neq u \\ \xi(v) & u \end{cases} \qquad \text{or} \qquad \xi'(w) = \begin{cases} \xi(w) & w \neq v \\ \xi(u) & v \end{cases}$$

or in words, pick either u or v at random, and place that VTD in the same congressional district as the other VTD.

Thus, there is a possibility to transition between two districtings if and only if they differ by exactly one VTD. Since the conflicted edge is chosen uniformly at random, the transition probability $Q(\xi'|\xi)$ is given by $\frac{1}{\text{con}(\xi)}$ where $\text{con}(\xi)$ is the number of conflicted edges of districting ξ. Similarly, $Q(\xi|\xi') = \frac{1}{\text{con}(\xi')}$. Thus, we let $a = \frac{P(\xi')\text{con}(\xi')}{P(\xi)\text{con}(\xi)}$ where P is the distribution from which we would like to sample. If $a \geq 1$, the proposed step is accepted. If not, the proposed step is accepted with probability a. There is one more requirement to consider. Each proposed districting must keep all 13 congressional districts contiguous. Each proposed step is immediately checked for this contiguity requirement. If a congressional district is not contiguous, a different step is proposed until a step has been proposed which satisfies this requirement. Then the algorithm is carried out with this proposed step.

In our measure on districtings, we defined P as $P(\xi) = \frac{e^{-\beta J(\xi)}}{c}$. Therefore a is given by:

$$a = e^{-\beta(J(\xi') - J(\xi))} \times \frac{\text{con}(\xi')}{\text{con}(\xi)} \tag{7}$$

$$a = e^{-\beta \Delta J} \times \frac{\text{con}(\xi')}{\text{con}(\xi)} \tag{8}$$

Notice that the normalization constant, $\mathcal{Z}$ cancels in the numerator and denominator. Thus, we have a way to sample from the distribution $P(\xi)$ without knowing the value of $\mathcal{Z}$.

Proposed steps which lower the energy are more likely to be accepted than steps which increase the energy. One might wonder why you should ever accept steps which increase the energy, since a larger energy is not desirable. To see this, consider the following example. Suppose the energy is given by $J(x) = e^{x/5} cos(\pi x) + 10$ from $x = 0$ to $x = 10$. Figure 1 is a plot of $J(x)$. Suppose at each state x, you can move to a nearby state $x' = x \pm \delta$ for some step size δ. If we never accepted steps that increased the energy, then the state would fall into one of the valleys instead of exploring the space. By accepting districtings with a larger energy with some probability, we have a way of escaping a valley. This allows the algorithm to more fully explore the state space.

3.3 Parameters and Heating/Cooling Schedule

Our measure of districtings depends on the following parameters:

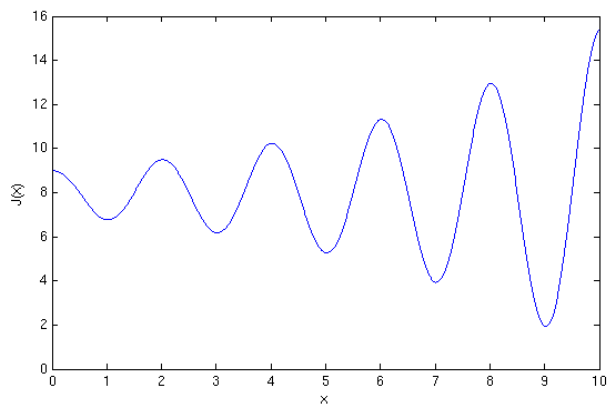

Figure 1: Example energy function. If you were not allowed to walk up the hills, you would not be able to explore different valleys.

- c_{pop}: scaling parameter for the population energy.
- c_{com}: scaling parameter for the compactness energy.
- λ: the proportion of the total energy which comes from the population energy.
- β: the steepness of the measure. A larger β more harshly penalizes districts with higher energy.

Note that this is an over parametization, since we could arbitrarily fix c_{pop} and c_{com} and then decide the values of λ and β. The parameters c_{pop} and c_{com} were used to bring the effects of the two energies on about the same scale, so that we could have greater control in finding a suitable balance between the two energies.

By fixing β, and then considering each energy factor individually (i.e. $\lambda = 1$ and $\lambda = 0$), we could determine suitable scaling factors c_{pop} and c_{com}. These factors were chosen with two things in mind:

- We do not want the samples to get stuck in a valley of the measure. We would like the districtings to keep evolving and exploring the state space.
- We do not want the energy of the districtings to trail off to infinity. Since a lower energy is more desirable, we want the algorithm to explore samples with lower energies.

This led to choosing $c_{\text{pop}} = \frac{1}{5000}$ and $c_{\text{com}} = 2000$.

After combining the two energies, we tried different values of β and λ. If β was too large, the samples would fall into a valley and stop exploring the parameter space. If β was too small, the energy would trail towards infinity.

With both the population and the compactness energies combined, there was not a suitable β to keep the districtings exploring as desired.

Our solution was to fluctuate between different choices of β to periodically heat up and cool down the energies. By heating up the energies with a small β, we allow the districtings to evolve so that they can reach different regions of the state space. Then by cooling down the energies with a large β, the districting is lowered towards a local minimum. This allows us to find desirable districtings without getting trapped in valleys of the measure. A sample districting is then taken at the end of each cooling period.

Fluctuating between $\beta = 0.001$ and $\beta = 0.01$ allowed the districts to evolve and still settle into a local minimum. We evolved the districtings with four choices of λ: 0.1, 0.2, 0.3, and 0.4, and two heating/cooling schedules: a shorter schedule where β is switched every 20,000 iterations, and a longer schedule where β is switched every 50,000 iterations. Figure 2 shows the spread of the population and compactness energies for sample districtings taken at the end of each cooling period. For comparison, the current districting of North Carolina has a population energy of $7,016$ and a compactness energy of $200,988$.

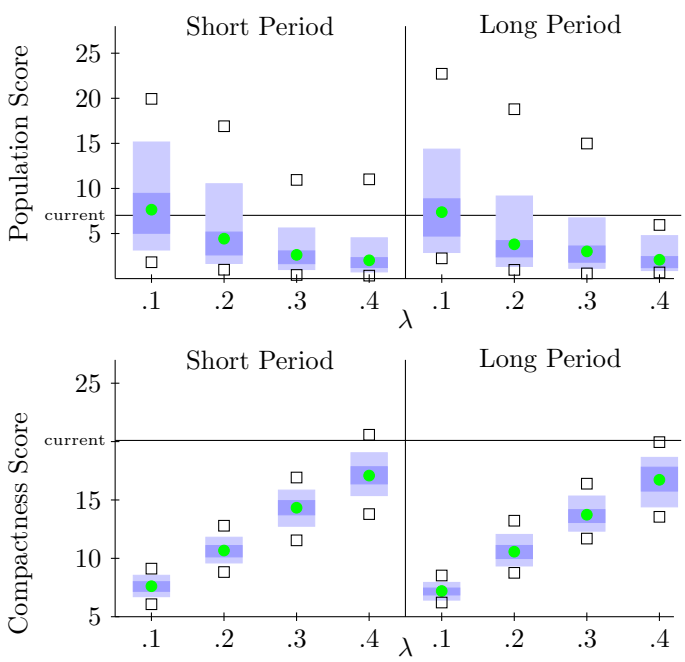

Figure 2: Population and Compactness energies for 4 values of λ and the shorter and longer heating/cooling cycles. Green dot gives the mean, darker box gives the 5th and 95th quartiles, lighter box gives the 25th and 75th quartiles, and the hollow squares give the min and max.

Notice that for $\lambda = 0.3$ or $\lambda = 0.4$, the population and compactness energies are almost always below the energies for the current districting of North Carolina. Since the current districting does an extremely well job of evenly dividing the population, but the districts do not appear very compact, we would prefer to penalize more for less compact districtings. Thus, we prefer to use $\lambda = .3$ for our analysis. Also, since the energies are not drastically different for the longer heating/cooling schedule, we prefer to use the longer schedule because it allows for the districts to evolve more between taking samples.

4 Results

After utilizing the Metropolis-Hastings algorithm to draw sample districtings from our probability distribution P, we retabulate the results of the 2012 election to obtain an empirical distribution for the number of elected representatives for each party over plausible districtings. For our preferred parameters of $\lambda = 0.3$ and with the longer heating/cooling schedule, we obtain the histogram in Figure 3 for the number of democratic representatives elected in the 2012 US House of Representatives election for North Carolina.

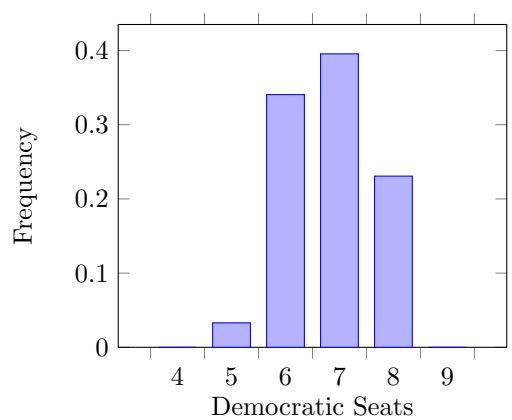

Figure 3: Histogram of number of democratic representatives elected after about 100 samples.

Notice that not once did we observe only 4 democratic representatives, as was the case in the 2012 election. To investigate any dependency of these results on our choices of parameters, Figure 4 displays summary statistics for the number of democratic representatives elected under different choices of λ and the longer or shorter heating/cooling schedule.

The results do not vary drastically between different choices of our parameters. Across all of these choices of parameters, we never observe only 4 democratic representatives, and the mean is always between 6 and 8. These results show that the event of electing 4 democratic representatives when drawing a

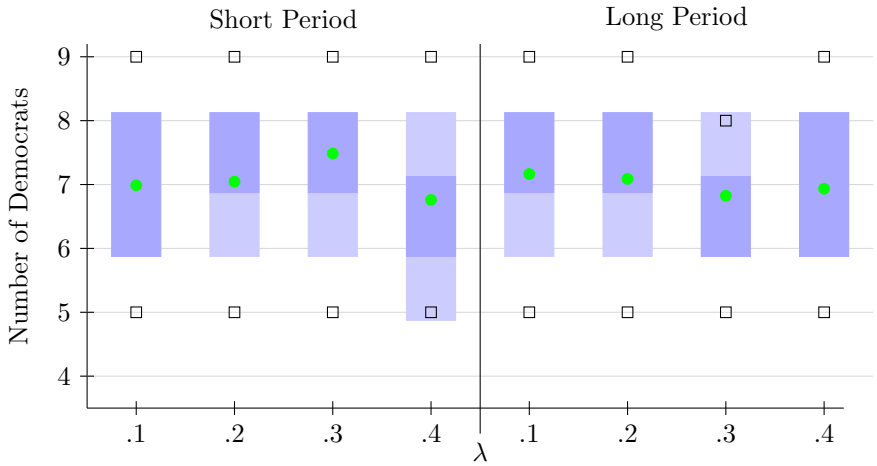

Figure 4: Summary statistics for the number of democrats elected for various choices of λ and the shorter or longer heating/cooling schedule. Green dot gives the mean, darker box gives the 5th and 95th quartiles, lighter box gives the 25th and 75th quartiles, and the hollow squares give the min and max.

districting of North Carolina according to P is unlikely.

5 Technical Notes

There are a few technical notes that we would like to address in this section. To refer to the current congressional districts, we need each vertex of our graph to belong to only one congressional district. We initially defined each vertex as a VTD. However, there are 65 VTDs which lie within two congressional districts. We split each of these VTDs to create two vertices, one for each of the two congressional districts. As before, vertices are connected when they are adjacent on the map. The congressional district boundaries are used to determine the boundary between the split VTDs, allowing us to determine which vertices are adjacent. Since population data is only available at the VTD level, we approximate the population for a split VTD as half of the population of the original VTD. Similarly, we approximate the area of a split VTD as half of the area of the original VTD. For brevity, we refer to each vertex as a VTD, even though some of the vertices are a split VTD.

Another technical note is that there are some votes that cannot be attributed to a specific VTD. For example, absentee voting allows votes to be cast outside of an individual's home VTD. The number of such votes is negligible to those votes that can be attributed to a VTD and we simply neglect them.

Not all of the necessary data to define the energy of a districting is readily available. Population data is available from the NC General Assembly [2] and areas of the VTDs can be obtained from the shapefiles of the congressional

districts [1]. Although the perimeter of a district could possibly be obtained from the shapefiles as well, there did not seem to be a convenient way to automate this task. Instead, the perimeters were estimated in the following way.

5.1 Estimating Perimeter

To determine the perimeter of a congressional district, we need to be able to estimate the shared perimeter of two neighboring VTDs as well as the shared perimeter of VTDs with the boundary of NC. To estimate the perimeter, we make the approximation that each VTD is a circle. For VTD v we can express its circumference in terms of its area by

$$\text{Circumference}(v) = 2\sqrt{\pi A(v)} \tag{9}$$

This gives an underestimate for the perimeter of a VTD, but it is a good estimate since VTDs tend to be roughly circular in shape. Next we make the approximation that the circumference of a VTD is evenly divided between it's neighbors. For an interior VTD v, let $deg(v)$ be the number of neighboring VTDs of v. For a VTD v on the boundary of NC, let $deg(v)$ be 1 more than the number of neighboring VTDs of v (since it's circumerference is also shared with the boundary of NC). Then the shared perimeter for an edge $e = (v, v')$ is approximated by:

$$\text{Shared Perimeter}(v, v') = \frac{2\sqrt{\pi A(v)}}{\text{deg(v)}} \tag{10}$$

Since this estimate could be based on either of the two VTDs, and since constants do not affect the relative sizes of the perimeters, we define the perimeter between two VTDs as

$$\text{Per}(v, v') = \frac{1}{2}\left(\frac{\sqrt{A(v)}}{deg(v)} + \frac{\sqrt{A(v')}}{deg(v')}\right) \tag{11}$$

For a VTD on the boundary of NC, we define the perimeter between the VTD and the boundary of NC (denoted o) as

$$\text{Per}(v, o) = \frac{\sqrt{A(v)}}{deg(v)} \tag{12}$$

This allows us to estimate the compactness energy, J_{com}, of a districting.

6 Future Work

This project could be extended to better quantify the extent that gerrymandering has been employed and the effects of district boundaries on election outcomes. Future work could be to incorporate other factors into the measure on districtings. For example, we could define other energies based on the following:

- Number of counties divided by congressional district boundaries
- Consideration of minorities
- Political bias (where districtings are preferred if they favor a political party)

The first two energies would make our measure more closely reflect the legislative rules for redistricting. The third energy would be particularly interesting since gerrymandering, or drawing district boundaries with the intention of favoring a political party, is the motivation for this project. Future work could also include analysis of other states with differing procedures for drawing congressional districts.

7 Conclusion

We have placed a probability distribution on the space of possible districtings of North Carolina into 13 congressional districts. This probability distribution only takes into account two factors: the compactness of the districts, and how evenly divided the population is among the districts. This probability distribution was sampled by using the Metropolis-Hastings algorithm. Then, the results of the 2012 US House election was retabulated under the randomly drawn districtings by assuming that voters would vote for the same party they voted for, even if they were voting for different candidates. By analyzing the results of the election under these randomly drawn districtings, we have an empirical distribution for the number of representatives from each party over plausible districtings. In this empirical distribution, we never see only four democrats elected, in contrast with four democrats being elected in the 2012 election. Since the results of the 2012 election are atypical under our model, this suggests that they were intentional.

8 Acknowledgements

I would like to thank my mentor, Jonathan C. Mattingly, for mentoring me throughout this project. I would also like to thank the Duke Math Department and the PRUV program for financial and material support, and Austin Graves for assisting in the construction of the graph from the VTD map.

References

[1] U.S. Census Borough. Shapefiles for congressional districts. http://www.census.gov/geo/maps-data/data/tiger-line.html, 2010.

[2] North Carolina Legislature. 2010 populations. http://www.ncleg.net/GIS/Download/Base_Data/2011/Reports/byVTD/rptBaseVTDTot.pdf, 2010.

[3] North Carolina Board of Elections. 2012 general election for US house by voting tabulation district. http://www.ncsbe.gov/ncsbe/Election-Results, 2012.

[4] North Carolina General Assenbly. Rules for redistricting. http://www.ncleg.net/

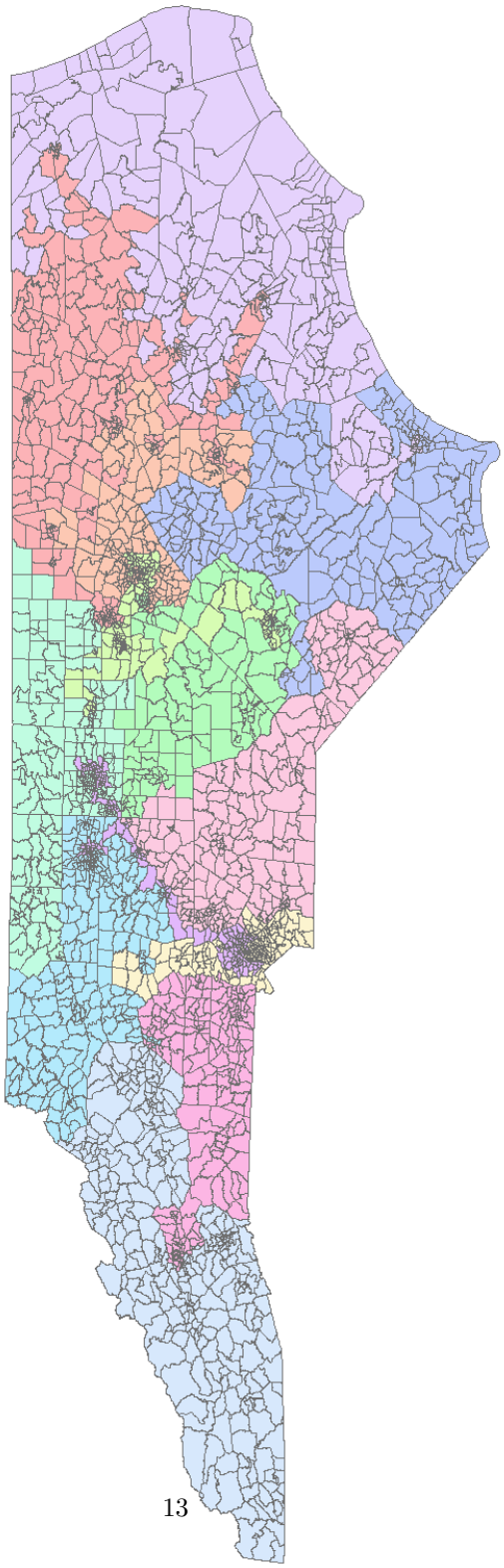

Figure 5: Current districting of NC after the 2010 census

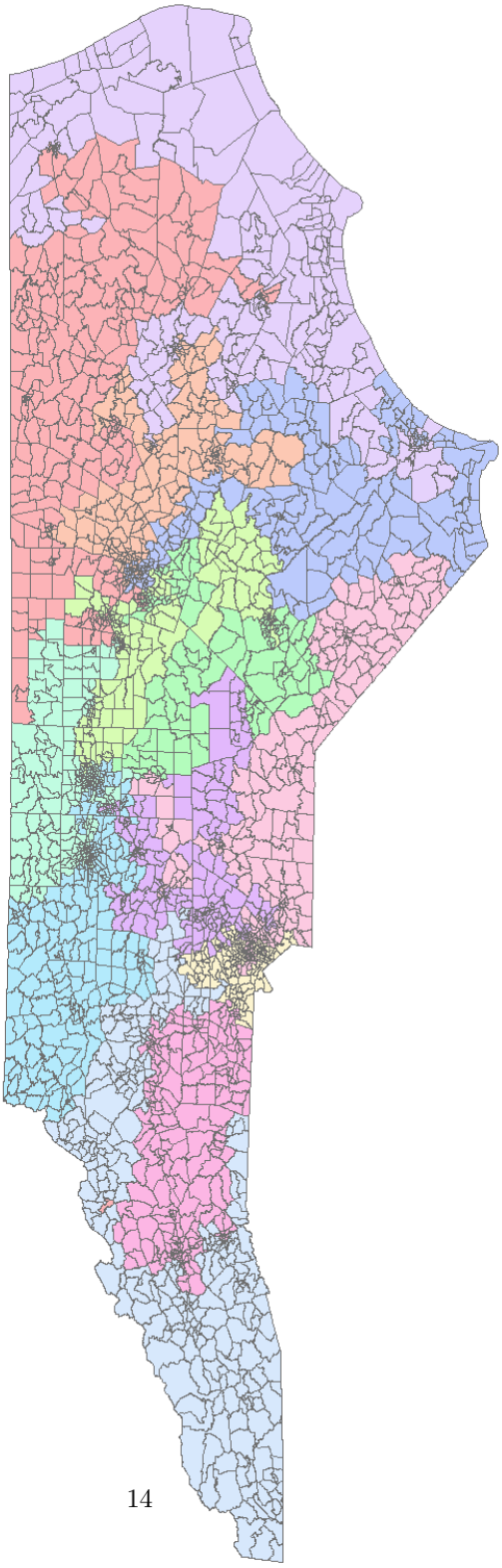

Figure 6: A sample districting from P with $\lambda = .3$ and the long heating/cooling cycle.

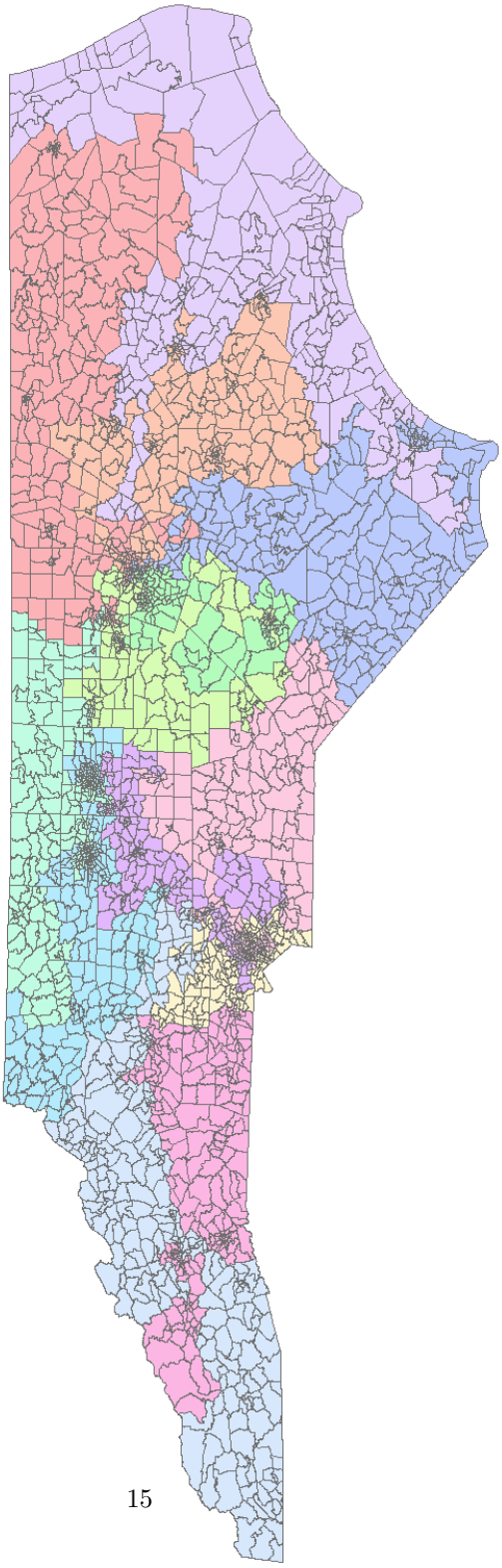

Figure 7: A sample districting from P with $\lambda = .3$ and the long heating/cooling cycle.

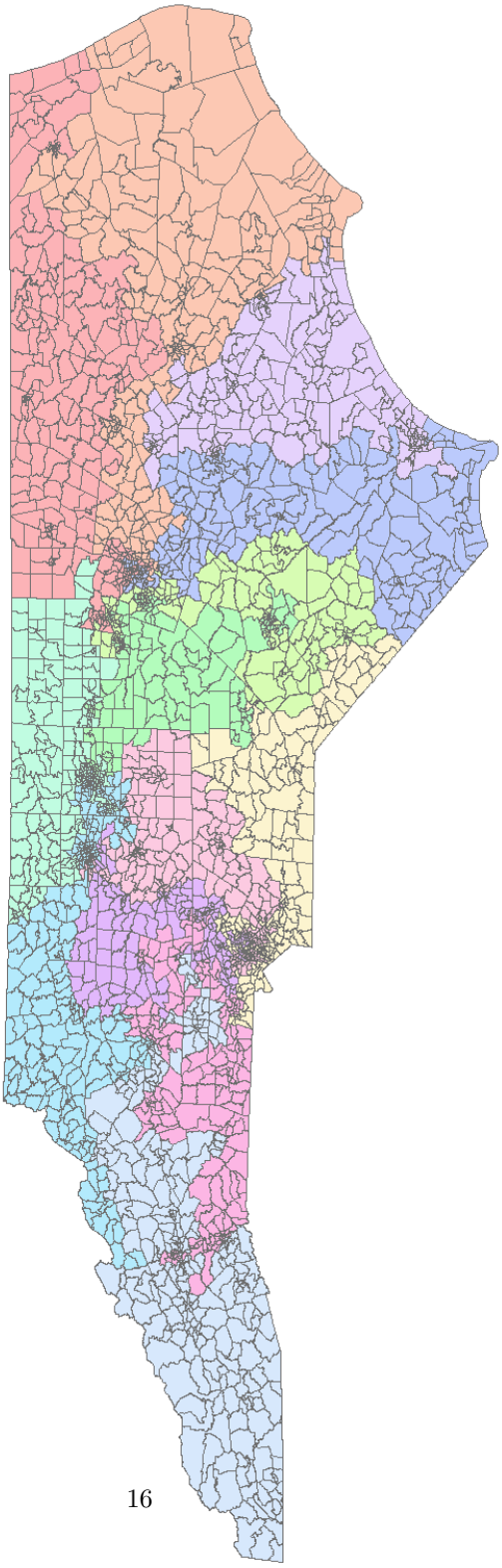

Figure 8: A sample districting from P with $\lambda = .3$ and the long heating/cooling cycle.