# Redistricting for Fairer Elections

Prabodh Shreedhar Katti, Mujammil Hamid Patel, DESE, IISc

## Abstract

***We solve the problem of a partitioning of a political subdivision, in this case a state, into districts by optimization of an energy function using linear programs and iterative cluster update. Two approaches to this assignment problem are discussed: volume preserving curvature flow and balanced power diagram. We set up an energy functional and solve a linear program to optimize the district assignment of population subunits subject to volume preserving constraints. Based on the results obtained, we simulate an election and compare it with the actual elections that took place.***

## I Introduction

Gerrymandering is the phenomenon of drawing boundaries of subdivisions of a state or other political entities into electoral districts to give unfair advantages to a single party, usually the party in power and in charge of the district assignment. This problem is rampant in the United States elections since its founding in the $18^{th}$ century and has been credited as a major contributing factor to a deepening political divide and partisan bickering plaguing the United States. A general illustration of gerrymandering phenomenon is given in figure 2. Most of the works on identification and rectification of gerrymandering are focussed on evolving compact, convex shaped districts; districts that do not 'look' gerrymandered.

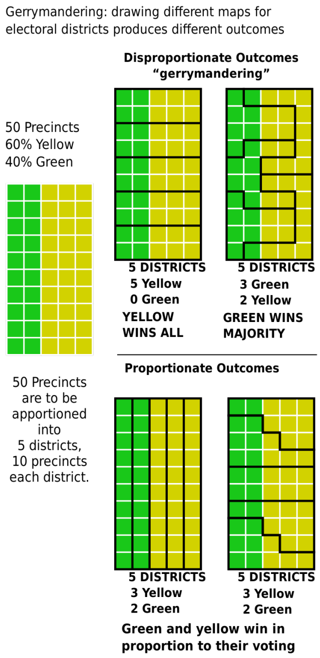

Gerrymandering: drawing different maps for electoral districts produces different outcomes
Disproportionate Outcomes "gerrymandering"
50 Precincts 60% Yellow 40% Green
5 DISTRICTS 5 Yellow 0 Green YELLOW WINS ALL
5 DISTRICTS 3 Green 2 Yellow GREEN WINS MAJORITY
Proportionate Outcomes
50 Precincts are to be apportioned into 5 districts, 10 precincts each district.
5 DISTRICTS 3 Yellow 2 Green
5 DISTRICTS 3 Yellow 2 Green
Green and yellow win in proportion to their voting

Figure 1: An illustration of gerrymandering (image credits: Mboli, Wikipedia. Released under CC)

The question of redistricting can be posed as an assignment problem over a graph that optimises an energy measure. The energy measure is constituted by placing terms that characterize the degree of gerrymander, or more importantly captures the important features that make for an ungerrymandered partitioning. A popular way is to imitate a natural process to evolve a partition, as discussed in subsequent sections. Grain boundary formation or surface tension flows are some examples. The main motivation behind these methods is a fully non-partisan partition, i.e we do not use party affiliation, key demographics like race or class or likely party-loyal voter distributions at all while drawing the mapping. This makes sure that the district maps so evolved are fully free of any partisan bias.

## II Case Study and Data

For our study, we explore the US Congress House of Representatives election for the state of North Carolina. North Carolina was picked because of its status as a 'swing state', a state that tends to swing either way -Democratic Party or Republican Party- in national elections, i.e it is not a safe Democratic or Republican state. Such states are focus of intense interference and scrutiny as these races determine the winner of presidential election or who gets a majority in Congress. A swing state also implies that the populace in the state is lot less partisan and the number of seats won by each party must be close enough. In reality, as we discuss in subsequent sections, the partisan gerrymander renders a wider gap which goes against the will of the people.

The data was obtained from various sources, chief among them being United States Census Bureau (USCB) for population and decennial census details and North Carolina State Board of Elections (NCSBE) for detailed countywise elections. The smallest unit of population was taken to be census designated subunit called tract, where population can range from anything between 0 to 14000. Figure 2 illustrates the distribution of tract centroids and the thirteen districts used for US Congress Elections of 2018.

North Carolina has 2195 such tracts with aroung 9 million people as per 2010 census, and has been allotted 13 districts in the US House of Representatives. In 2020, following a court order, the map was changed and the new one has significantly less gerrymandering and resulted in narrowing the gap between Democrats and Republicans. We will discuss this further in latter sections.

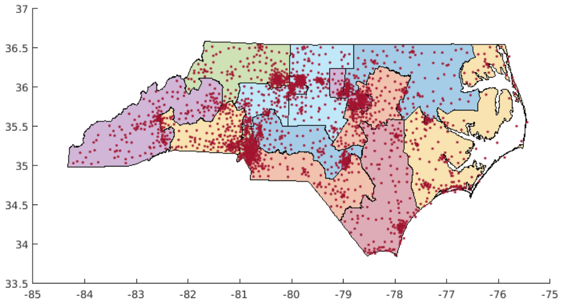


Figure 2: Tract centroid distributions along with the district partition used in 2018 midterms

## III Volume Preserving Curvature Flow

### *Formulation*

Motion by mean curvature evolves the boundaries of the partition in a graph by iteratively minimizing an energy fucntion. The energy formulation takes ints inspiration from PDEs that govern the surface tension flows of liquids or bubbles, or evolution of grain boundaries in crystal structure [1]. For a surface tension flow, the energy will be proportional to the surface area of the hypersurface, which in our 2D case would be the perimeter.

We shall formulate the surface tesntion flow problem as a partitioning problem where each district evolves to optimize the energy while preserving the volume, i.e a minimum population constraint, which is necessary to ensure nearly equal number of people in each district [2]. Let us consider a weighted graph $(\mathscr{V}, W)$. Let $x \in \mathscr{V}$ be the census tracts. To keep the weight matrix sparse, we will only keep $k$ weights for a tract and null out the rest. Here we have taken $k = 200$. The weights are then non linearly transformed into a modified version of Gaussian RBF called Zelnick-Manor and Perona weights

$$W(x,y) = exp\left(\frac{-d^2(x,y)}{\sigma(x)\sigma(y)}\right). \tag{1}$$

.

Here $\sigma(x)$ is the distance to the $\frac{k}{2}^{th}$ nearest neighbor of the given node/tract. $W$ is further normalised per row. Let $u_i(x)$ be a one-hot assignment vector for a district $i$ such that

$$u_i(x) = \begin{cases} 1 & x \in i \\ 0 & otherwise \end{cases}.$$

We can now formulate the energy as

$$E(\underline{u},\underline{c}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{x,y\in\mathscr{V}} u_i(x)A(x,y)(1-u_i(y)) + \alpha\sum_{i=1}^{N}\sum_{x,y\in\mathscr{V}} ||c_i - c(x)||^2 \tag{2}$$

Here $A = W^\top W$ has been done to make the matrix positive definite and thus guarantees the optimization of $E$ . $c_i$ are the nodes of the district which is updated after every iteration, and $c(x)$ is the tract centroid. The first term in the equation represents the graph cut with respect to $A$. This term approximates the measure of perimeter. The second term is a regularization condition that forces the nodes to form disricts that reduce dispersion. $\alpha$, the regularizing factor, determines the relative importance of the two terms. Jacobs et al. [1] have extended the MBO algorithm that basically states that for optimal mean curvature flow it is enough to minimize over the linearization the energy function, i.e.

$$u^{n+1} = \underset{u}{\text{argmin}} < \nabla E, u - u^n > \tag{3}$$

The constraints for this equation comes from the fact that elements of all assignment vectors must add up to 1 and that all districts must have roughly equal population. In other words

$$\sum_{i=1}^{N} u_i(x) = 1 \tag{4}$$

and

$$\sum_{x\in\mathscr{V}} u_i(x)p(x) \geq P \tag{5}$$

Here $p(x)$ is the population of the district, and $P = 0.999 * \frac{P_S}{N}$. There is a small relaxation given in terms of population constraints for feasibility. Here $P_S$ is the total population of the state and $N$ is the number of districts, 13 in the case of North Carolina.

### *Difference from the reference paper*

We have taken [2] as the chief reference for the volume preserving flow framework. While the formulation theory is similar, we differ at two points.

- Through experimention and dimension analysis, we found that the regularization term must be normalized to get better results. So in place of $||c_i - c(x)||^2$, we have incorporated $\frac{||c_i - c(x)||^2}{max(||c_i - c(x)||^2)}$ .
- The gradient of the energy function is given by

  $$\psi = \alpha||c - c(x)||^2 - \mathbf{A}\underline{u} \tag{6}$$

  and not. The authors have used $W$ is place of $A$, which we believe is not correct.
- The authors have used auction dynamics [1, 3] for optimization problem, where as we have used more standard solvers available in MATLAB.

The algorithm for iterative solving is thus given in 1.

We also experimented by adding an annealing term $\sim \mathscr{N}(0,T)$ to $\psi^{n+1}$ at every iteration. Experimentally, we kept the $T$ at 0.01, and reduced the intensity of the term by various annealing parameter values. The annealing parameter reduce the variance after each iteration by a fixed factor, simulating the cooling process. The results are discussed in the next section.

**Algorithm 1** Districting algorithm

```
For every iteration:
Calculate $\psi^{\mathbf{n+1}} = \alpha||c^n - c(x)||^2 - \mathbf{A}\underline{u}^n$
Solve the Linear program $u^{n+1} = \underset{u}{\text{argmin}}\,\underline{u}^\top\psi$ subject to constraints $\sum_{i=1}^N u_i(x) = 1$ and $\sum_{x\in\mathcal{V}} u_i(x)p(x) \geq P$
Update the district centroids: $c_i^{n+1} = \frac{\sum_{x\in\mathcal{V}} u_i^n(x)c(x)}{\sum_{x\in\mathcal{V}} u_i^n(x)}$
```

### *Results and Validation*

Figure 3 shows a map after running 100 iterations. We see that the map immediately began to evolve to form 'better' district partitions. As seen in figure 4a, the energy term immediately decreases to a low value, then jumps up and stabilizes. It stabilizes to a value that has net energy lower than the initial energy. Time taken for stabilization is proportional to the annealing factor. The intention of this term was to allow the solver to explore different directions and jump to a better local minima iand find a lower energy value, but eventually the optimizer is reaching almost same final energy values. We can also see the local approximation of energy, the linearization 4b, is reducing with time.

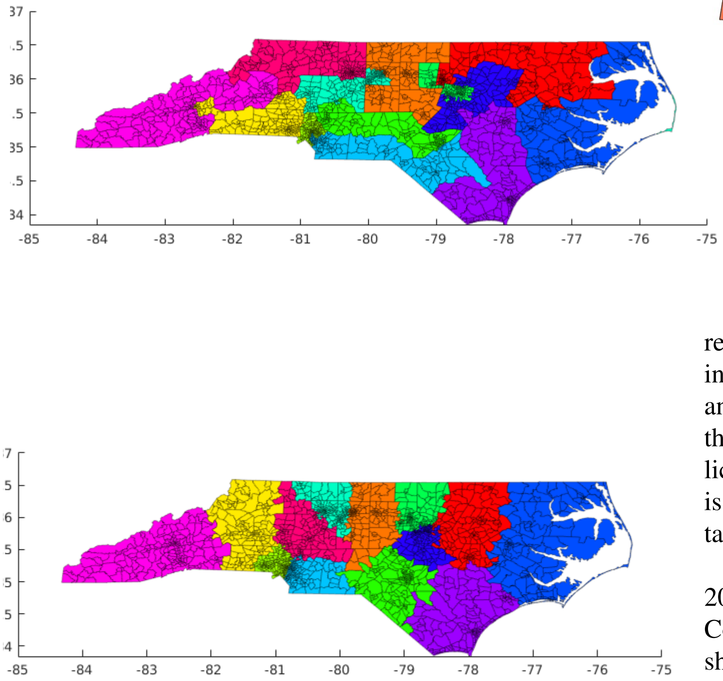
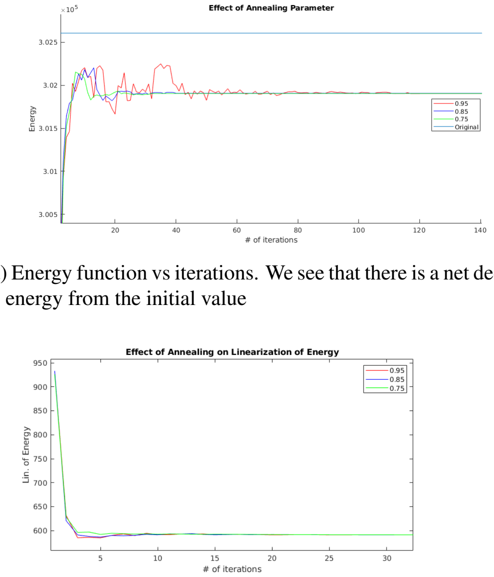


Figure 3: District before (T) and After (B). We see a better, ungerrymandered map evolved out of the original map.

We then proceed to validation by simulating a congressional election result. We used the US House of Representatives election results from 2018 and Presidential election from 2020 (by analyzing the data, we determined that the presidential election results and HOR results are highly correlated and therefore are indicative of one another. At least in case of 2020. It is easier to analyze the presidential election data, we took that). We took the results of two major parties, Democratic Party and Republican Party. Third parties have been ignored as their proportion is very insignificant, less than 2%. The results are given in the table 1.



(a) Energy function vs iterations. We see that there is a net decrease in energy from the initial value

(b) Linearization of Energy vs iterations

Figure 4: Optimization results

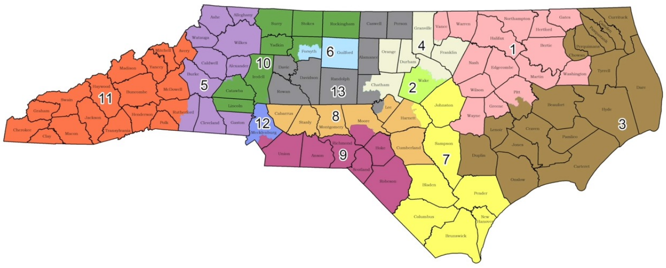


Figure 5: New districts used in 2020 elections

The districting partition we used to initialize were used for 2018 elections, and were struck down before 2020 election by Courts. The new districts redrawn by the NC state authorities shown in figure are significantly less gerrymandered. Our results too indicate that; they are closer to 2020 election results than 2018. In fact, there is significant deviation for 2018 districting, so much so that the party getting majority seats have been flipped! As indicated by percentage of votes recieved by each party in the election, our results are more proportional the actual election results.

| ELECTION | ACTUAL RESULTS | RESULTS WITH OUR MAP |
|---|---|---|
| 2018 Midterm (VOTES: 49% - 51%) | DEM-3 REP-10 | DEM-7 REP-6 |
| 2020 Elections (VOTES: 49.3% - 50.7%) | DEM-5 REP-8 | DEM-6 REP-7 |

Table 1: The results with volume preserving curvature flow method. We can see that the results with new maps are more proportionally representative.

## IV Balanced Power Diagrams

We explore a particular approach to redistricting: balanced centroidal power diagrams. Given the locations of a state's m residents and the desired number k of districts, a balanced centroidal power diagram partitions the state into k districts with the districts' populations differ by at most one [4].

A balanced centroidal power diagram is a particular kind of solution to balanced k-means clustering: given a set $P$ of $m$ points (the residents) and the desired number $k$ of clusters, a solution consists of a sequence $C$ of $k$ points (the centres) and an assignment $f$ of residents to centres that is balanced: it assigns $\lfloor m/k \rfloor$ residents to the first $i$ centres, and $\lceil m/k \rceil$ residents to the remaining $k-i$ centres (for the $i$ such that $i\lfloor m/k \rfloor + (k-i)\lceil m/k \rceil = m$). The cost of a solution $(C, f)$ is the sum, over the residents, of the square of the Euclidean distance between the resident's location and assigned centre. We seeks a solution of minimum cost.

The solution $(C, f)$ only needs to be a local minimum, meaning that it is not possible to lower the cost by just varying $f$ (leaving$C$ fixed), or just varying $C$ (leaving $f$ fixed). Hence we have used a variant of Lloyd's algorithm: start with a random set $C$ of centres, then repeat the following steps until an equilibrium is reached: (1) given the current set $C$ of centres, compute a balanced assignment $f$ that minimizes the cost; (2) given that assignment $f$, change the locations of the centres in $C$ to minimize the cost.

### Formulation and Algorithm

The problem in step 1 can be converted minimum cost flow problem. Firstly we consider the formation of power diagram $P(C,w)$ which is defined as follows. For any centre $x \in C$, the weighted squared distance from any point $y$ to $x$ is $d^2(y,x) - W_x$. The power region $C_x$ associated with $x$ consists of all points whose weighted squared distance to $x$ is no more than the weighted squared distance to any other centre. We calculate the total cost as follows:

$$\sum_{y \in P, x \in c} d^2(y,x) - W_x$$

We convert the above problem to minimization problems as follows,

Minimize

$$\sum_{y \in P, x \in c} d^2(y,x) a_{yx}$$

Subjected to

$$\sum_{y \in P} a_{yx} = \mu_x \qquad (x \in C)$$

$$\sum_{x \in c} a_{yx} = 1 \qquad (y \in P)$$

$$a_{yx} \geq 0 \qquad (y \in P, x \in C)$$

Here, $u_x \in \lfloor m/k \rfloor, \lceil m/k \rceil$ and $a_{yx}$ represent the assignment of $y$ to centre $x$. As the data of individual residents was not available, instead a tract wise data was available, we modified the above problem in the following manner,

Minimize

$$\sum_{y \in P, x \in c} d^2(y,x) a_{yx}$$

Subjected to

$$\sum_{y \in P} P_y a_{yx} = \mu_x(1+T) \qquad (x \in C)$$

$$\sum_{x \in c} a_{yx} = 1 \qquad (y \in P)$$

$$a_{yx} \geq 0 \qquad (y \in P, x \in C)$$

Here $P$ represent set of tracts and $C$ represent set of centres of districts. $u_x \in \lfloor m/k \rfloor, \lceil m/k \rceil$ and $a_{yx}$ represent the assignment of a tract $y$ to district centre $x$. $P_y$ is the population of tract $y$. $T$ introduced reperesents the tolerence limit set by us.

Once the cost flow problem is solved, we move to the second stage in which we balance the centre of districts by the following update

$$x = \frac{\sum_{y \xrightarrow{f} x} P_y X_y}{\sum_{y \xrightarrow{f} x} P_y}$$

Where $X_y$ is the centroid of tract $y$.

We repeat step 1 and step 2 in one iteration, and we keep repeating this process until the new assignment becomes equal to the last iteration assignment.

### *Experiment*

We run our experiment into two parts. In the first part, we perform election based on a newly districted map created by us using 2018 election data. We have used the data of tract population from 2010 census data. In election data of 2018, county wise number of votes for the Republican party and the Democratic party was available. We assumed that the tract

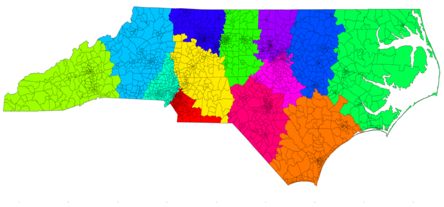


Figure 6: Districting for 2018 election

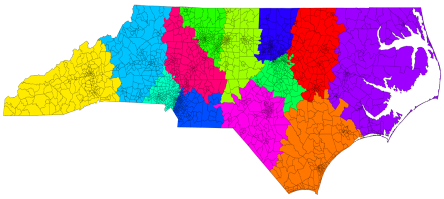


Figure 7: Districting for 2020 election using 2010 census

voter choice ratio for Republic and Democrat is equal to voter choice ratio of the county to which it belongs. Also, the turnout percentage and percentage of eligible voters of the tract was assumed to be same as of that county to which the tract belongs.

In the second part, we perform election based on newly districted maps created by us using 2020 presidential election data. We made two maps. In one map we have used census 2010 data to get tract population and maps was generated using this population data. For the second map, we took an extrapolated data of the population. We had a data of county wise extrapolated population available to us. We assumed the percentage change in the tract population to be equal to the percentage change in the county population to which it belongs. As the data was extrapolated, it better represents the tract population in the present compared to 2010 census data. We have used the same assumption used in the first part of the experiment with the only difference being the election data used now is 2020 presidential election data.

### *Results*

In the 2018 election, 49% voters voted for the Democratic party and 51% voted for Republic party. But due to improper districting only 3 out of 13 seats were won by Democratic party. Clearly, this result does not represent the voter's choices. We conducted an election on maps generated by us using election data we got Democratic party to lead 7 out of 13 seats. This is a better representation of voters call. The map generated by us is shown in figure 6.

In the 2020 election, 49.3% voters voted for the Democratic party and 50.7% voted for Republic party. But again due to improper districting only 5 out of 13 seats were won by Democratic party. After conducting the election on the map generated by us using 2010 census data shown in figure 7. we get the same results. But when we used extrapolated and developed a map shown in figure 3. the results changes. Conducting election on this map show Democratic party to win 6 seats out of 13.

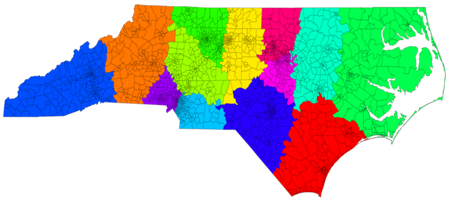


Figure 8: Districting for 2020 election using extrapolated population data

## V CONCLUSIONS AND FUTURE WORK

Fair district maps and ungerrymandered partitions are fundamental to a healthy functioning of democracy, and methods demonstrated here can detect and correct these issues, and produce a more proportionate representation. We have used tract as the fundamental unit of census data. A finer granularity, of say block level, could result in better flow and lesser energy configurations in case of volume preserving curvature flow (VPCF) method, and we can achieve convex polygonal districts, as demonstrated by Cohen et al [4] for balanced power diagrams (BPD) method. For VPCF, we let the regularization function dominate because it reduces the tract splitting. But since our district boundaries are restricted by tract shape, we can explore allowing splitting and subsequently suggesting a boundary shape corresponding to the split of our own, although finer granularity will solve this problem too. For now, we resolved splitting offline by thresholding in the assignment vector. For VPCF, a more efficient solver like auction dynamics [1, 2] can be explored.

Our analysis has also suggested that using population projection data generated by the Census Bureau instead of decennial census data may be more helpful, since census data is updated only every ten years. We have used $L_2$ distance measure for computing weight matrix. In real life, the geographical and physical features like mountains, water-bodies can split communities, so a straight-forward 'as the crow flies' $L_2$ can be replaced by driving distance to better take on ground connectivity into account.

Data and codes can be accessed at: `https://indianinstituteofscience-my.sharepoint.com/:f:/g/personal/mujammilh_iisc_ac_in/ErffKeRPua1Imzrbsa3yf84BwEht1P5epHTJAdT1Zw9UuA?e=iuMuk8`

## REFERENCES

[1] Matt Jacobs, Ekaterina Merkurjev, and Selim Esedoáž¡lu. Auction dynamics: A volume constrained mbo scheme. *Journal of Computational Physics*, 354:288–310, 2018.

[2] Matt Jacobs and Olivia Walch. A partial differential equations approach to defeating partisan gerrymandering. *arXiv preprint arXiv:1806.07725*, 2018.

[3] Dimitri P Bertsekas. Auction algorithms. *Encyclopedia of optimization*, 1:73–77, 2009.

[4] Vincent Cohen-Addad, Philip N Klein, and Neal E Young. Balanced power diagrams for redistricting. *arXiv preprint arXiv:1710.03358*, 2017.