# I BALANCE POWER DIAGRAM FOR REDISTRICTING

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[reference].

A balanced centroidal power diagram is a particular kind of solution to an optimization problem called 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. In balanced k-means clustering, one seeks a solution of minimum cost.

A balanced centroidal power diagram arises from a solution to balanced k-means clustering that is not necessarily of minimum cost. Instead, 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.

# II 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 reperesent 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.

# III 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 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.

## IV 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 1.

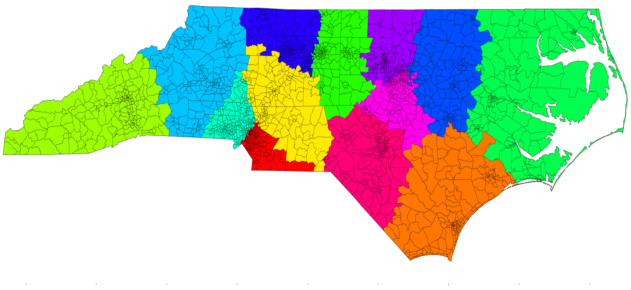

Figure 1: Districting for 2018 election

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 2. 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. Hence this map is fairer for conducting elections.

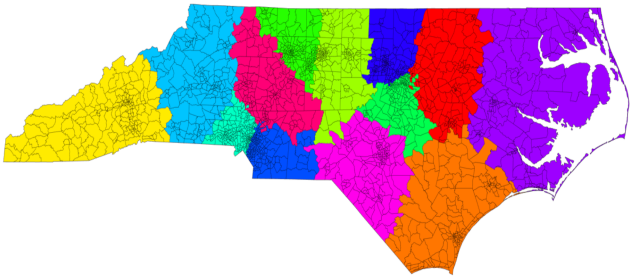

Figure 2: Districting for 2020 election using 2010 census

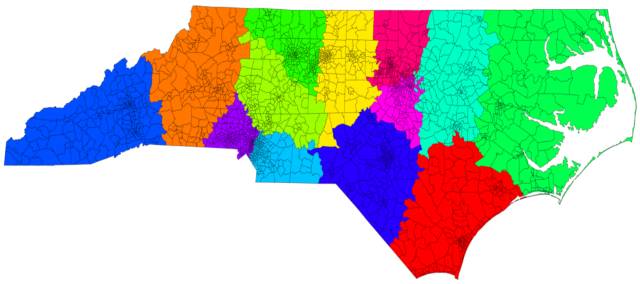

Figure 3: Districting for 2020 election using extrapolated population data

## V CONCLUSION

Current maps are gerrymandered and do not correctly represent voters decision. New maps generated by balanced power diagram should be used for election for fairer election. Instead of tract population data, more granular data like block-level population data can be used for the districting process. This will district boundaries more straight and will result in convex polygon shaped districts. Also using the latest population data is very important; this can significantly impact election results. Using correctly extrapolated population data instead of old population data will be beneficial for the districting process.