Daniel de las Heras Garcia

1838259

Dr. Venkatesh

Introduction to Simulations and Stochastic Processes

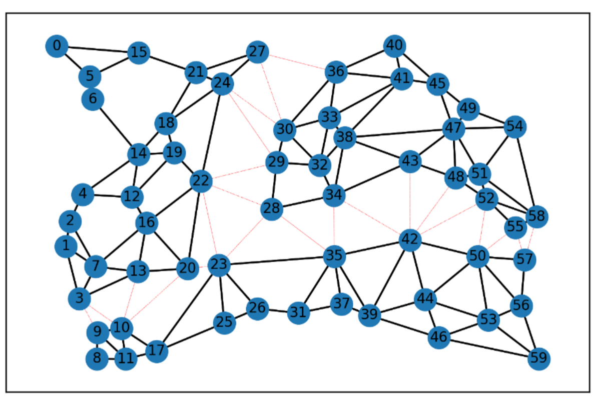
Final Project

Partisan gerrymandering is the deliberate manipulation of the political boundaries of election districts to favor a specific political party unfairly. By purposefully dividing voters among various districts, partisan gerrymandering aims to increase the number of seats a party may win in an election.

The act of gerrymandering itself is not a recent one; it has been used for ages. The capacity to design extremely accurate and successful gerrymandered districts has, however, substantially increased as a result of technological developments and access to comprehensive demographic data. The political process is distorted by partisan gerrymandering, which is frequently condemned for compromising the principles of fair representation.

In this paper we propose a MCMC (Monte Carlo Markov Chain) algorithm to sample a number of possible plans uniformly from an initial plan. By doing this we can check the probability of obtaining a particular map and then check for gerrymandering.

We developed an algorithm which samples different plans uniformly by carrying out transitions from an initial plan. This initial plan proposes the following map:



From this plan we got the following summary:

```
Number of Districts: 3
   District 0  District 1  District 2
0          20          20          20
                  district_0  district_1  district_2
Republicans             9889       10035       10027
Democrats              10111        9965        9973
Total Population       20000       20000       20000

Electoral Competitiveness 0.9955148148148149
```

From this we see that the plan is divided into three districts. The total number of nodes or precincts is 60 and this initial plan has 20 precincts in each district. The overall population is 60,000 with each district being having 20,000 inhabitants. The number of republicans and democrats are pretty even within and among all of the districts all districts averaging over 10,000 democrats and the same number of republicans. In general terms, this initial plan is very balanced overall: population, republicans-democrats ratio, number of precincts per district. And this is confirmed to a greatest extent when we see that the electoral competitiveness is also .9955.

The Algorithm

It is important to understand the transition algorithm in order to sample later. To perform one transition, we need to identify all the boundary nodes and select one randomly. From this we will then identify all the boundary districts and randomly a uniformly select one. This would leave us with a proposal plan which we would then accept using our Metropolis Criterion. The formal algorithm is as follows:

Step 1: Identifying all the nodes that are at the boundary of a district. Let this set be B={$i \in N : \exists(i,j) \in E$ and $(i,j) \notin E(p)$} and randomly select one with equal probability .

Step 2: Identifying all the districts bordering the selected edges and select one randomly. Hence from D_i = {$\forall P_k \in p : \exists(i,j) \in E, i \in P_l$, $j \in P_k$} choose one $d \in D_i$ with probability $\frac{1}{|D_{i}|}$.

Step 3: Propose the swap to the district selected. In other words, if we select $d \in Di$ then $i \in P *_{k,t}$ and $j \in P *_{l,t}$, then we propose the swap to $i \in P *_{k,t}$and $i \notin P *_{k,t}$. The probability of acceptance of this proposal is given by: $\alpha(pt{-}1 \to pt$, $i)$ = $min(1$, $|B*|$ $|B|$ $)\alpha\left(p_t \to p_{t,i}\right) = \min(1, \frac{|B| g_\beta(B^*)}{|B^*| g_\beta(\beta)})$.

The term $g_\beta(B^*)$ tells the Gibbs distribution for

$$g_\beta(p) = \frac{1}{z(\beta)} \exp\left(-\beta \sum_{V_{k \in p}} \psi(V_k)\right)$$

Where p is a redistricting plan, $\beta \geq 0$, is the inverse temperature, z(β) is the normalizing constant, V_k is a particular district, and $\psi(V_k)$ is a function which reflects population constraints to ensure certain plans can not be sampled.

Essentially, this function peaks around plans that meet the constraints therefore ensuring higher probability of sampling these plans.

It is important to introduce the Population Equality Constraint which is mathematically defined:

$$P_v = \max_{1<l<n} \left| \sum_{i \in V_k} \frac{p_i}{\hat{p}} - 1 \right| \leq \omega$$

Where ω tells us to what extent do we want the districts population to deviate from the mean of 20. Thus, $\psi(V_k) = P_v$.

By enforcing this distribution instead of the uniform distribution, we manage to sample only from plans that meet our population constraints. But we still want to sample paths uniformly in order to do that we resample weighing each plan with the inverse of their Gibbs probability

$$\frac{1}{g_{\beta}(p)} \text{ (1)}$$

After understanding the transition algorithm, we introduce what algorithm we will use to sample plans. To do this we would do 1000 transitions from the initial plan. We will get the final plan and include it in the sample. We would do this 1000 leaving us with a 1000 plan sample. From this sample we will resample weighting each plan with (1) leaving us with the desired sample. After running this code several times, I realized how power consuming was this algorithm taking up to 3873 seconds.

Tuning $\boldsymbol{\beta}$

The next challenge I faced was to find an ideal value of β to perform the transitions. After some coding I found out that the higher value of β the lower the population equality constraint which implied that lower variation of the plans since more plans where discarded.

In my code I suggested 9 values of β: 0.1, 0.5, 1.0, 4.0, 5.0, 10.0, 20.0, 80.0, 100.0. I used the sampling algorithm with less sampled plans using the different β proposed. I averaged the Electoral competitiveness, P_v, and the minimum nodes in a district for the 100 plans generated using each of the β values and got the following results:

	0.001	0.500	1.000	4.000	5.000
Pv	0.350000	0.450000	0.450000	0.300000	0.200000
EC	0.914239	0.933313	0.957341	0.938494	0.909728
Minimum District	13.000000	11.000000	12.000000	16.000000	17.000000

	10.000	20.000	80.000	100.000
Pv	0.15000	0.000000	0.000000	0.000000
EC	0.91541	0.901093	0.995515	0.995515
Minimum District	17.00000	20.000000	20.000000	20.000000

We can observe that β values below 5.0 yield P_v>.25. We want to find an equilibrium between discarding of new Plans and still get variation within plans. The stricter the constraint the more plans are not accepted, but the lower the population variation. We want to find a β which implies that the population of all districts must be within 25% of the population parity. To do this we want to find the lowest beta with $\beta < .25$. From above we can observe that $\beta = 5.0$ is the lowest value that satisfies this.

Electoral Competitiveness

In the code we implemented a method in Plan class to calculate the electoral competitiveness of a set of districts using the Tam Cho and Lin formula.

A mathematical calculation called the Tam Cho and Lin formula is used to gauge how competitive an electoral system is. To determine a district's competitiveness, the vote share and seat share are combined for each district.

The arithmetic behind the formula is broken down as follows:

1. The absolute difference between the votes received by the first candidate (pop[0]) and half of the district's total population (pop[2]) is used by the code to determine the vote share contribution for each district.

2. To calculate the average vote share (vote_share), the vote share contributions for each district are added up and divided by the total number of districts.

3. The code determines the contribution to the seat share by determining if the first candidate earned more votes than the second candidate (pop[0] > pop[1]). If so, seat_share_sum is increased by 1. In the event of a tie, seat_share_sum is increased by 0.5.

4. To calculate the average seat share (seat_share), the seat share contributions from all districts are added up and divided by the total number of districts.

5. The following formula is used to measure electoral competitiveness (ec):

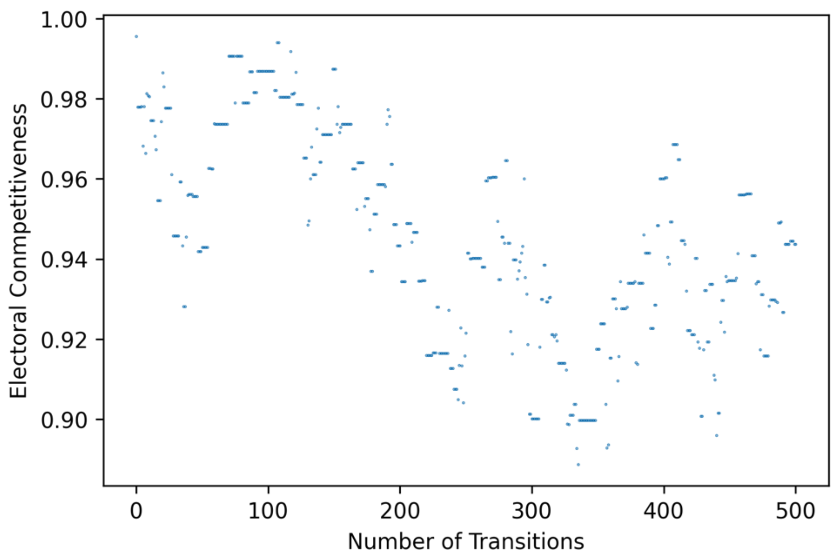
$$ec = 1 - (1 + \alpha * seat\_share) * \beta) * vote\_share$$

The average vote share for all districts is represented by the variable vote_share.

The average seat share for all districts is represented by the variable seat_share.

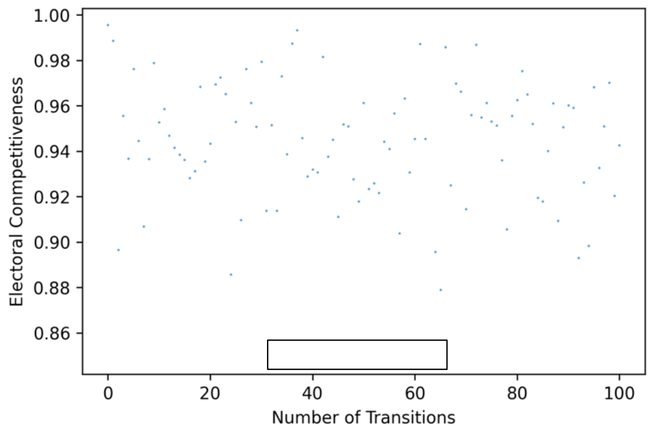
The parameters alpha and beta can be changed to reflect the significance of the seat.

To check for the Electoral Competitiveness, I performed 1000 transitions on the original plan and plot the electoral competitiveness against the transition number.



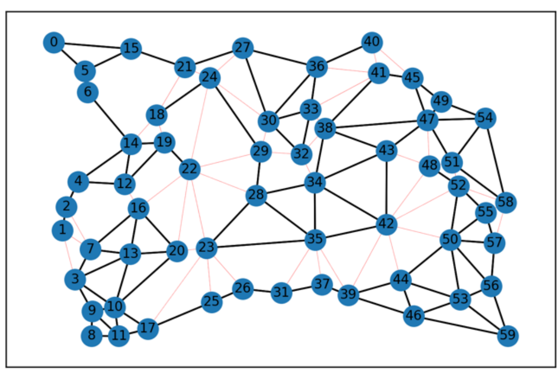
Overall, the higher the number of transitions a decrease in the Electoral Competitiveness. Overall, the variation in electoral competitiveness does not vary largely (not more than .1 up or down.). This could suggest that gerrymandering might not be an issue with the initial plan.

After I produce a scatter plot with a sample of 100 random plans:

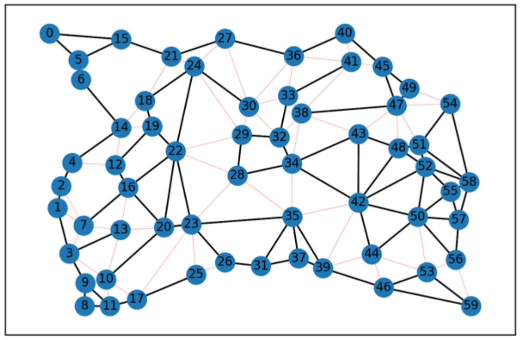


Again, showing very little variation in electoral competitiveness.

Overall, there are several clues that inclined me to reject the hypothesis that the initial plan exhibited gerrymandering. The electoral competitiveness, the node distribution within others appeared to show a very balanced and reliable plan. However, if we observe the graph, we can see that around transition 350 we get the least electoral competitiveness:



Compactness in this plan is a clear issue. We can see that node 18 lies very far away from most of the nodes in the same district. Overall, all the districts are elongated rather than in clumps. Electoral competitiveness is relatively small, close to 0. 85. All of this show some signs of possible gerrymandering.



This plan again has a similar electoral competitiveness. The district configuration is very abnormal, for example, the left-most district takes barely any inner nodes, taking all the north and east and most of the south nodes.

In conclusion, the initial plan provided appears to be possibly the best configuration of districts and precincts. The population distribution, along with the democrat and republican count along the districts is even, and the electoral competitiveness appears to be the highest that you can get. For this reason, I think we should reject the hypothesis that the initial plan exhibits partisan gerrymandering.