CS 524: Introduction to Optimization Fall 2020

Author: A.N. Other (Joint with Yet A.N. Other)

Gerrymandering with Adjacency Constraints - Project Spec

1. What is the issue being addressed?

Gerrymandering refers to a strategy where politicians try to maximize the votes they get by proper redistricting. In the problem 1 of homework 6, we were asked to deal with a Gerrymandering problem by maximizing the votes from congressional representatives to a given party. The biggest difference between the homework problem and the real world is that we have to make sure all states in an electoral district are connected.

2. Where does the data come from and how will it be obtained?

I will use two datasets. The first one is a synthetic dataset created by myself so that I can demonstrate my model on a smaller and easier dataset. I will design a changeable dataset just like those demonstration in MIRO. The second one is USA 2016 Presidential Election[1] and County Adjacency File[2]. Depending on the computational cost, I might choose several swing states as examples to show the effectiveness of my model.

3. What is the optimization problem underlying this project?

I have two ideas for this problem. The optimization problem in this project can either be just finding a feasible solution that makes a specified party win or maximizing the total votes the party gets. Other than the original constraints of simple Gerrymandering in the homework, I have to make sure that the connectivity within an electoral district is also maintained. I am considering to use network flow to model the connectivity constraint.

4. What are the deliverables?

First, I will try to formulate a model that guarantees the connectivity when assigning nodes to each district. Second, I will combine the Gerrymandering rules into that model and make it able to find the solution under the connectivity constraint. Third, I will start to design my simple datasets (probably an mn rectangle where each pixel represent a state) and provide visualization to it. Lastly, I will use my model on US election result and use Geopandas to visualize the redistricting.

5. Other points for me to consider when evaluating. For now, I am most concerned with the computational cost of this problem.

[1] https://public.opendatasoft.com/explore/dataset/usa-2016-presidential-election-by-county/ table/?disjunctive.state

[2] https://www.census.gov/en.html