Comparing Algorithms for Detecting Gerrymandering: An Investigation Into Bias In District Selection In Ensemble Map Generation

Jonas Eichenlaub, 2022

Gerrymandering is a pervasive issue in American politics and over the last decade mathematicians have made notable contributions towards its detection and avoidance. One focus of these efforts has been using computers to generate an ensemble of potential legislative maps and then comparing a real map against this distribution of maps. Vote counts from recent elections can then be plugged into both the real and simulated maps to understand whether the number of districts a given party wins under the real map is statistically likely under the distribution of generated maps. In other words, if one party wins far more seats in the real map than the mean number of seats they win in the simulations, it is likely the real map is the result of gerrymandering. Since gerrymandering is often hard to precisely define, it is easy for policymakers to circumvent any one metric of how gerrymandered a district is; the strength of this technique is that it does not rely on a single metric, instead wholistically comparing the real map with the ensemble of possible maps.

A prevalent way of creating such an ensemble of simulated maps is with a *merge-split* algorithm, which involves encoding a state as a graph with voting precincts as nodes. A spanning tree is traced over this graph and then edges are cut away such that the graph is broken into equally sized connected components, each of which corresponds to a legislative district. Constraints within the algorithm make sure that the edges that are cut away create districts that adhere to whatever legal criteria legislative maps must meet, such as equal population, compactness, municipality-preservation, etc. By cutting different edges of the original graph, different maps are formed, which collectively create an ensemble of possible maps. Duke Professors Jonathan Mattingly and Gregory Herschlag have proposed formalizing such an approach by placing a merge-split algorithm into a Metropolis Hastings Markov Chain, where the proposal of a new map is separated from the acceptance of that proposal into the ensemble. Instead of making decisions about how maps are added to the ensemble based the structure of the algorithm, Mattingly and Herschlag's work applies the existing mathematical theory on Metropolis Hastings chains to justify their approach; among other benefits, this allows the maps to be sampled uniformly over possible partitions of the state, rather than uniformly over possible spanning trees of the graph of the state. Each map represents a stage in the chain, and new maps are proposed with the merge-split technique described above. Maps are then added to the chain based off an acceptance probability that quantifies how well the proposed map fits within the legal criteria, which is designed to be changeable to meet the specific policy parameters of different states. By making the choices within the algorithm explicit, Mattingly and Herschlag offer an approach to detecting gerrymandering that is more assessable and defensible when used as evidence in legal challenges to gerrymandered maps.

Part of their contribution is to sample maps uniformly from possible partitions of a state rather than uniformly from possible spanning trees of the graph of the state, since each partition of a state can be formed by a different number of spanning tree configurations. My research this summer focused on determining whether sampling from partitions generates different ensembles of maps than that from the original spanning tree method; more specifically, I investigated whether algorithms that sample from the space of spanning trees are biased against generating maps that draw district lines along the border of areas with high and low population density (i.e., along the borders of cities). Working with Prof. Mattingly and Herschlag, I ran a version of their algorithm that sampled from the partition space and a version that mirrors the merge-split approach sampling from the spanning tree space on a variety of "test states" that I constructed. I represented each test state simply as a lattice graph, with a few nodes on the lattice replaced with smaller, denser lattice graphs representing cities. I varied the placement, size, and number of these "cities" to gain insight into how often each of the two versions of the algorithm cut the edges between the city and noncity nodes. To run these experiments, I needed to make sure the Metropolis Hastings chain was reaching its stationary distribution. Accordingly, a secondary part of this research revolved around understanding ways to check for the convergence of Markov Chains and then writing code to create statistics that apply those ideas to ensembles of maps.

Ultimately, I found that the algorithm that samples from spanning trees is biased against making districts that have edges along the border of cities, but only under certain situations. The two main factors that determine whether such bias appears in an ensemble are the district population size and the city population size. When the city population fits perfectly within a single or multiple districts, the spanning tree algorithm is significantly more likely to create maps that do not break up cities compared to the partition-space algorithm; hence, a hidden bias in the original merge-split approach is evident. However, if the city population does not fit evenly into one or more districts, both algorithms are forced to break up the city and there is not a notable difference between the ensembles they generate. These results reenforce why it is important to use Mattingly and Herschlag's approach, as their formalized algorithm allows for policymakers to explicitly decide to whether to keep municipalities within the same district, but does not automatically do so in a hidden and uncontrollable manner. Given that the redistricting process from the 2020 Census is currently underway, this insight could find application in the upcoming legal challenges to the inevitably gerrymandered maps that politicians propose.

Bowdoin Faculty Mentor: Jack O'Brien; External Mentors: Jonathan Mattingly, Gregory Herschlag
Funded by the Gibbons Summer Research Program