A Visualization-Based Geometrical Approach to Redistricting

Seth Drew

seth.drew@tufts.edu

***Abstract*— This paper aims to provide a basis for geometrical analysis of districting techniques. I also include an analysis of different compactness measures in the Markov Chain Monte Carlo (MCMC) techniques used in current computational redistricting literature from Chinka et al.** [1].

***Keywords*— Redistricting, Gerrymandering, Computational Geometry, Markov Chain Monte Carlo, Districting Metrics, Stationary Distribution**

I. Introduction

Redistricting and gerrymandering have become increasingly controversial topics in recent years. Technology has allowed redistricting officials increased control and searching power over a broad spectrum of plans and associated metrics, allowing them to re-draw voting districts that suit their goals. The U.S. Supreme Court plans to take two cases in the next year [2] [3] centered on proposals for new anti-gerrymandering metrics and will consider the question of whether partisan gerrymandering is justiciable.

Courts have continually asked for a unified metric to define gerrymandering, but there is no consensus on such a metric. In addition, the number of possible district plans is so large that finding global optima under a single metric is computationally infeasible. Therefore, researchers from Pittsburgh and CMU have presented *Assessing significance in a Markov chain without mixing* [1], in which they use statistical techniques to reject the null hypothesis that a specific district plan was chosen from the stationary distribution of a Markov Chain. If a district plan is not chosen from the stationary distribution, it is likely the district map was made with some partisan intent.

II. Data Used and Definitions

The data in this paper is from Pennsylvania congressional district maps where each district is a combination of n precincts. In the 50 U.S. states, the number of precincts n is about 2 orders of magnitude larger than the number of districts. Pennsylvania has 19 districts and about 9,000 precincts. This analysis could easily extend to census tracts, census block groups, or other levels of granularity.

A **valid** districting plan for a state is one that is simply connected, retains low population differences between districts (less than 2% in this case) and adheres to some test of geometrical compactness.

III. Geometry of Redistricting

A foundational part of the current research in computational redistricting is called the *1-flip* operation. If districts A and B share a border, the *1-flip* operation flips a precinct in district A to district B. Given an initial districting, this operation is used to explore valid alternative plans stemming from the initial seed.

The geometry of these precinct graphs is important when using techniques like Markov Chain Monte Carlo (MCMC). MCMC aims to explore the "configuration space" of valid district plans. The configuration space is a graph where each node corresponds to a district plan, and each edge between plan A and plan B means that using a *1-flip* operation on A gives B.

It follows that **connectivity** of this space is an important question. If the configuration space is not connected, it would be possible to choose a seed district plan from a small subspace of the full configuration space. Appendix 1 contains an outline of a proof that the configuration space is connected under the *1-flip* operation. That is, any valid district plan is achievable by using k *1-flip* operations from any seed district plan.

IV. MCMC for Redistricting

MCMC is a sampling technique for exploring a large state space in limited time. In general terms, it operates in steps on a graph with transition probabilities between nodes. After $$n$$ steps, there is probability $p(s|n)$ to be at state $$s$$. If the configuration space is connected and the chain is ergodic, as the number of steps goes to infinity, the Markov Chain converges to the stationary distribution.

In the districting problem, as defined by [1], each state in the chain is a district plan for a precinct map, and a step in the chain is an application of the *1-flip* operation. Due to the large configuration space, convergence of the Markov Chain to the stationary distribution is not guaranteed in a finite amount of time, and current applied MCMC implementations have not shown convergence. To address this issue, [1] proposes to use MCMC techniques to sample from the space to produce an ensemble of valid district plans, without requiring convergence or full state space exploration.

Sampling from many valid plans provides a basis to critique individual plans using statistical analysis of traditional anti-gerrymandering metrics.

V. Compactness & Visualizations

Using the package from [2], I have aggregated visualizations of the Markov Chain operating under different compactness constraints over time. These visualizations can help researchers gain a better understanding of how different constraints affect the result of the randomized *1-flip* operation in the Markov Chain.

In addition, I extended the package to include a new discrete constraint proposed by Moon Duchin [4] based on total population of the district and the population of its boundary district(s). This measure aims to help standardize measures of compactness across different states.

It resembles the traditional and widely cited Polsby-Popper Metric, which takes values between 0 and 1:

$$\mathrm{Cpct}(D_i) = \frac{4\pi A}{p^2}$$

Where A is the total area of the district and p is the perimeter of the district. Under this metric, 0 represents thread-like objects, and 1 represents circles. Because this measure is normalized, it seems intuitive to compare Polsby-Popper measures for different states with each other, but factors like state geography and population differences make these measures uncomparable.

The discrete metric instead uses population on the perimeter (dp) of the state versus the total population of the state (dA).

$$\mathrm{Cpct}(D_i) = \frac{dA}{dp^2}$$

As with other compactness metrics, the goal is to maximize compactness. Therefore, under the discrete metric, the goal is to maximize the interior population of a district and minimize the boundary population. Unlike Polsby-Popper, this metric is not constrained between 0 and 1. Therefore, it carries better comparison utility across state lines.

Accompanying this paper are visualizations designed to help to compare these compactness measures with each other using data from the 2017 PA congressional district map. [7] Specifically, visualizations are provided for the evolution of the district map under MCMC with four measures: no compactness constraint, the $L1$ norm, L^∞ norm of Polsby-Popper, and under the discrete metric.

These metrics each encourage good individual districts, but do not discourage bad districts enough. Therefore, we use the inverse of each of these metrics to ensure that each district in the plan is compact by penalizing the score heavily for any single districts.

VI. Future Work

Geometry, Graphical Analysis

I have claimed that the configuration space of possible districtings for precinct map redistricting is connected under the *1-flip* operation for abstract connected graphs, but ignored requirements that each intermediate step has equal population or is compact. In fact, the current proof of connectedness relies on wildly unbalanced districts.

In order to prove connectedness of the space under the method proposed in [1], we must also enforce <2% population differences, and some compactness constraint. One idea of a place to start here is by assuming the precinct graph is a triangulation, and that each precinct has equal population.

Is every state's precinct connectivity graph a triangulation (or easily made a triangulation by adding dummy nodes)? Triangulations are planar graphs that have many useful properties for analysis. This would be a good first step in any attempt to prove connectedness of the configuration space under compactness and population requirements.

What is the subgraph structure of the state graphs? Are there some kinds of subgraph isomorphisms we can find amongst the states that behave similarly under *1-flip*? In the precinct graph, are there high degree precincts in cities? In the country?

Visualization

As described in Geometry, two points in the configuration space are connected under the definition of the *1-flip* operation. It would be useful to apply visualization techniques to the configuration space. Visualization of this space could help to get an idea of possible clusterings, transition states between clusters, and to visually analyze the connectedness of the space. A good starting point is from [6]. This paper shows visualization techniques applied to exploring graphs that change over time.

Metrics

The new discrete metric has comparability and geographic independence advantages over the traditional Polsby-Popper Metric. What are the side effects of the discrete metric? For example, because the metric penalizes a district for large population on the boundary, the discrete metric aligns with the traditional redistricting principle to not cut through urban centers.

MCMC Sampling

Because of the large configuration space, it is infeasible to aim for convergence to the stationary distribution in MCMC sampling. How small does our dataset have to be to guarantee complete enumeration in a reasonable amount of time? What is the mixing time in these small examples? How does the mixing time relate to the number of precincts and districts?

What is the curvature of the configuration space for redistricting plans? Is it positively curved or negatively curved?

Legal

Some important questions to keep in mind: Do my techniques help me implement case law? Can I make the case that I have performed a best-effort implementation of current case law? Why is the current plan or proposal lacking?

References

[1] Maria Chikina, Alan Frieze, and Wesley Pegden, *Assessing Significance in a Markov Chain without Mixing*, Proceedings of the National Academy of Sciences, March 14, 2017, vol. 114 no. 11, 2860–2864

[2] Chikina et al, **markovchain** package: http://www.math.cmu.edu/~wes/files/markovchain.tgz

[3] Gill v. Whitford. (n.d.). Oyez. Retrieved May 10, 2018, from https://www.oyez.org/cases/2017/16-1161

[4] Benisek v. Lamone. (n.d.). Oyez. Retrieved May 10, 2018, from https://www.oyez.org/cases/2017/17-333

[5] Duchin, Moon. February 2018. *Outlier analysis for Pennsylvania congressional redistricting.*

[6] Stef van den Elzen, Danny Holten, Jorik Blaas, Jarke J. van Wijk, *Reducing Snapshots to Points: A Visual Analytics Approach to*

Dynamic Network Exploration, 1 IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No. 1, January 2016

[7] Seth Drew, http://sethdrew.pythonanywhere.com/

Appendix

Section 1

This section contains an outline of the connectedness under *1-flip* operation.

Conjecture: Graphs without a forbidden configuration (cut-vertex with 3 or more split-components) have connected configuration space under the 1-vertex flip operation.

This forbidden configuration is shown below. Consider a 3-precinct, 2-district state. In this picture, where each node is a precinct, each edge between these nodes implies connectivity, and the lines across these edges are district split lines. Because a district must be simply, connected, the configuration space of this graph is not connected, as flipping the cut-vertex over the district line would lead to a disconnected district on the starting side.

In order to extend this to real state examples, each of the leaf nodes is allowed to contain a subgraph beyond them of arbitrary size, but it is important that these subgraphs do not connect to one another. This configuration has not been searched for in any of the state precinct graphs as of the writing of this paper.

This configuration can be solved by adding dummy nodes (fake precincts) with population zero to change the near-triangular precinct map into a full triangulation (where this forbidden configuration occurs).

Algorithm strategy: Let be initial partition and the desired partition.

Construct the district adjacency graph for . Take a non-cut-vertex and a district in with non-empty intersection. Transform the desired district in getting a new valid partition which has one identical district as . Delete the district and recurse.

Compute the color class adjacency graph of the target partition.

Compute the block-tree of the input graph and take a leaf. This must be a 2-connected subgraph.

Compute the SPQR tree of such subgraph.

If there is an R node. choose any vertex of this node.

If not, then the graph is series-parallel. Choose a leaf of the SPQR tree that is a S node and choose a vertex of the chain (guaranteed to exist, since its an S node).

One of the color classes must contain this vertex. Contract this color class along a spanning tree and delete this vertex.

This procedure does not create a forbidden cut-vertex.

Repeat times until we have only 1 color class.

Do this for both initial and target and we get a path in the configuration space by reversing the sequence of moves for the target.

Upper bound in the number of moves: .

This proof was created in collaboration with the Tufts University computational geometry study group.