Transport Distances between Districting Plans - Real-world Experiments

Si Wu, Melisande Teng, and Kristen Akey

August 2019

1 Introduction

At the Voting Rights Data Institute, we are able to generate lots of districting plans. For this project, we experiment with the notion of distance between two districting plans. That is, given two plans, we will generate a value that is zero when they are the same, small when they are similar, and large when they are different. In order to compute distances between pairs of districting plans, we compute a pairwise distance matrix D for an ensemble, where d_{ij} denotes the Wasserstein distance between plan i and j. Multidimensional Scaling (MDS), a dimensionality reduction approach, is then used to embed the pairwise distance matrix onto a 2-dimensional space.

In general, MDS works as follows: given a distance matrix D, MDS attempts to find a set of data points in a d-dimensional space that form a coordinate matrix whose configuration minimizes a loss function. In metric multidimensional scaling, the loss function is normally a stress function which is a residual sum of squares of the Euclidean distances for all pairs of points in the original high dimensional space and the Euclidean distances for all pairs of points in the embedded space.

2 Methodology and Data Analysis

The analysis is done on Iowa election shapefile. The shapefile was obtained from the U.S. Census Bureau and processed by members of the Metric Geometry and Gerrymandering Group (MGGG). The data can be found at here.

2.1 Ensemble Generation

Ensembles of districting plans are generated by the GerryChain software.

2.2 Comparison of Flip and Recom

Single edge flip ("Flip") is a design of Markov chains used to build ensembles. For Flip, at each step in the Markov chain, we (uniformly) randomly select a pair of adjacent census blocks that are assigned to different districts, then (uniformly) randomly change the assignment of one of the blocks so that they match.

For recombination ("Recom"), at each step, we (uniformly) randomly select a pair of adjacent districts and merge all of their blocks into a single unit. Then, we generate a spanning tree for the blocks of the merged unit with the Kruskal/Karger algorithm. Finally, we cut an edge of the tree at random, checking that this separates the region into two new districts that are population balanced to within 2 percent of ideal district size.

The explanations of "Flip" and "Recom" can be found here.

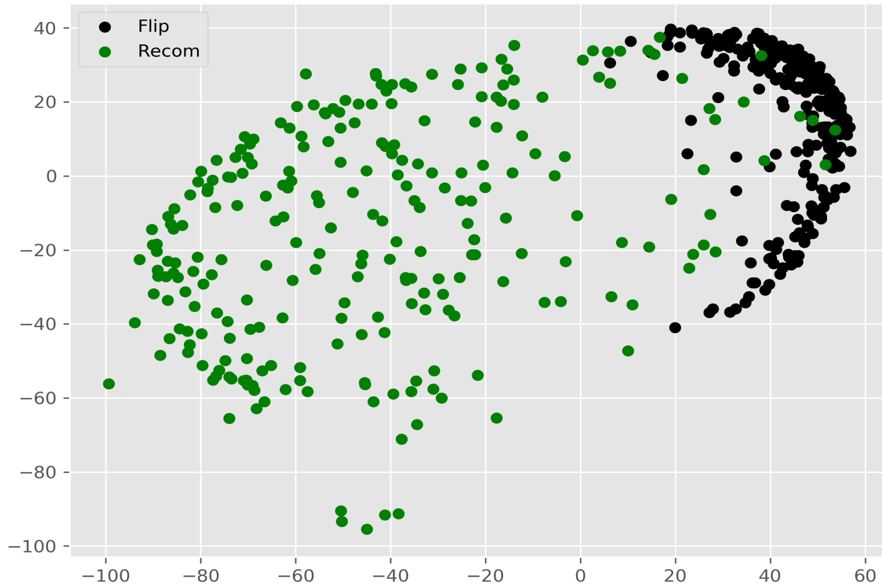

Figure 1: Comparison of Flip and Recom (MDS embedding)

To create the figure above, a Flip run of 25000 steps was done on Iowa, where we took every 100th step to get 250 total plans. The same procedure was repeated for Recom. A 500 by 500 pairwise distance matrix (250 for Flip, 250 for Recom) was computed and used for the MDS embedding.

From figure 1, we can see that compared to Flip, Recom is able to explore a larger space of possible districting plans.

2.3 MDS Visualization of Recom, Colored by Partisan Statistics

To create figures below, a Recom run of 25000 steps was done on Iowa, where we took every 100th step to get 250 total plans. The points are colored by partisan statistics.

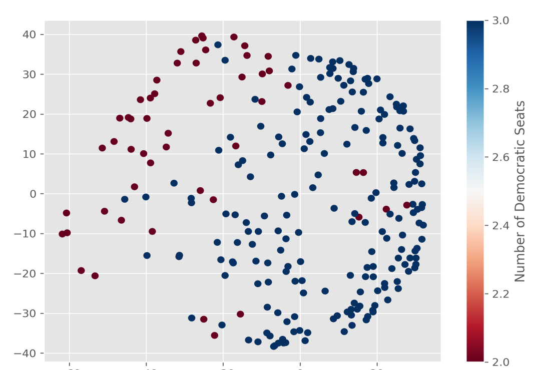

Figure 2: Recom, colored by Democratic seats (election 2000)

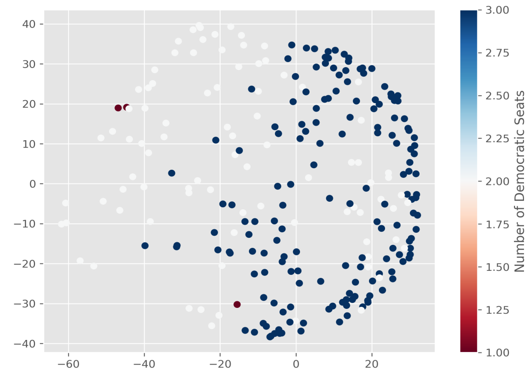

Figure 3: Same Recom run, colored by Democratic seats (election 2004)

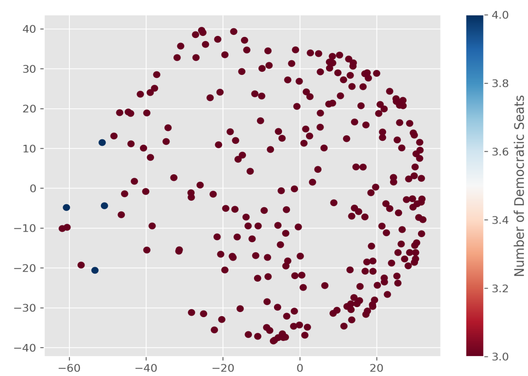

Figure 4: Same Recom run, colored by Democratic seats (election 2008)

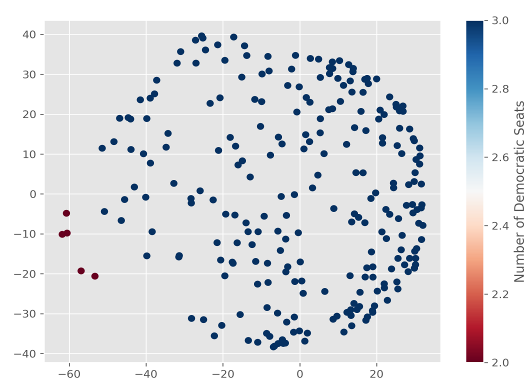

Figure 5: Same Recom run, colored by Democratic seats (election 2012)

Something interesting to note: the points that have higher Democratic seats in figure 4 seem to cluster; and the points that have lower Democratic seats in figure 5 seem to cluster.

3 Future work

Showing independence of seed for ReCom. We need to run ReCom from a few different starting partitions, and plot them all colored by seed and see how mixed the points are. While there has been lots of work showing independence of seed in metrics, this would be the first attempt to show independence at the level of plan geography.

Find out how many clusters do the most extreme (by seats) plans form, and what they look like geographically.

Conduct the same experiments on Pennsylvania, using landmark MDS.