An Address-Based Precinct Definition

Max Fan

September 28, 2020

Contents

1 Abstract

Over the past few decades, several processes for quantifying gerrymandering have been proposed. When applied to real-world situations, these processes frequently rely on unreliable data from states, counties, and a variety of other jurisdictions. In some cases, precinct lines are traced by hand over scanned copied of paper maps. In such cases, identical precincts are difficult to distinguish algorithmically, making it hard to calculate district stability. Unreliable and unstable data sources pose a significant obstacle in producing reliable and accurate gerrymandering research, requiring extremely labor-intensive human data vetting. This paper proposes a simpler and more accurate method of representing districts: defining precincts as the set of addresses contained within. Additionally, the information-theoretic equivalence of such a representation of such a will be shown.

2 Introduction and Motivation

The current standard of representing voting precincts is a geo-spatial one. Precincts are represented as polygons and manipulated through various geometrical techniques. These representations are saved in a geo-spatial "shapefile", allowing for easy analysis and sharing. This format is the most visually intuitive representation of voter precincts: representing them as shapes.

One prominent metric used in gerrymandering research is the compactness metric. Intuitively, a district with a greater perimeter to area ratio is more likely to be gerrymandered than a district with a lower ratio. This is the essence of the various compactness measurement techniques proposed in the literature.

For example, a commonly proposed compactness method is exactly the method described previously: the ratio of the perimeter to the area. Another commonly used, alternative metric is the cut edges metric. The cut edges metric is generated from the dual graph of the polygon-based representation of the precincts in a district. The dual graph is generated by converting every precinct to a node, with every adjacent face of two precincts as the edges. Then, the cut edges metric is the number of edges in the dual graph that are **cut** by a districting plan.

Some compactness metrics, particularly the perimeter-area metric, can vary from year to year or even within the shapefile itself. This is particularly the case with states and counties that have inconsistent resolutions across precincts and years. If the shorelines in one precinct are extremely fine, whereas the shorelines in another are extremely coarse, the perimeter-area metric can be disproportionately inaccurate in reliably detecting gerrymandering. However, the cut edges metric does not suffer from this particular flaw for shorelines.

Several other data quality issues exist for cut edges. For example, sometimes inter-precinct boundaries can sometimes be extremely craggily, for example, if a river separates two different precincts. In such a case, many edges would be created for the same adjacent face. Without human, geographical context, it is difficult to distinguish from the legitimately many-edged precincts, such as a gerrymandered precinct, from the naturally many-edged precincts, such as those delineated by rivers or lakes. In fact, making such a determination, if possible, would be so labor intensive that these data quality issues are typically ignored due to the immense difficulty in determining which polygon edges are worthy of inclusion in the dual graph. In short, any process to clean up the quality issues in a dual graph based upon real-world precinct measurements is fraught with many subjective, arbitrary decisions. As the goal of creating objective metrics to quantify gerrymandering is to remove subjectivity in the analysis, a different approach is required.

3 New address-based representation

3.1 Advantages

3.1.1 Data Quality

3.1.2 Equivalence

3.1.3 Monads

3.2 Voronoi diagram

3.2.1 Conversion between polygon and voronoi

3.3 Drop-in replacement

4 Experimental design (TBD)

5 Results (TBD)

6 Conclusion

7 Future work