Measuring the Compactness of Political Districting Plans

Roland G. Fryer, Jr. and Richard Holden*

August 11, 2007

Abstract

The United States Supreme Court has long recognized compactness as an important principle in assessing the constitutionality of political districting plans. We propose a measure of compactness based on the distance between voters within the same district relative to the minimum distance achievable – which we coin the relative proximity index. We prove that any compactness measure which satisfies three desirable properties (anonymity of voters, efficient clustering, and invariance to scale, population density, and number of districts) ranks districting plans identically to our index. We then calculate the relative proximity index for the 106th Congress, requiring us to solve for each state's maximal compactness; an NP-hard problem. Using two properties of maximally compact districts, we prove they are power diagrams and develop an algorithm based on these insights. The correlation between our index and the commonly-used measures of dispersion and perimeter is -.22 and -.06, respectively. We conclude by estimating seat-vote curves under maximally compact districts for several large states. The fraction of additional seats a party obtains when their average vote increases is significantly greater under maximally compact districting plans, relative to the existing plans.

Keywords: Compactness, gerrymandering, power diagrams, redistricting.
JEL Codes: H70, K19

*We are grateful to Alberto Alesina, Rosalind Dixon, Edward Glaeser, Emir Kamenica, Lawrence Katz, Gary King, Glenn Loury, Barry Mazur, Franziska Michor, Peter Michor, David Mumford, Barry Nalebuff, Ariel Pakes, Andrei Shleifer, Andrew Strominger, Jeremy Stein, and seminar participants at Brown (Applied Math) and Harvard (Labor Economics) for helpful discussions and suggestions. Alexander Dubbs provided exceptional research assistance – going far beyond the call of duty. Shiyang Cao, Laura Kang and Eric Nielsen also provided excellent research assistance. Financial support was provided by the Alphonse Fletcher Sr. Fellowship. Fryer thanks the Erwin Schrödinger International Institute for Mathematical Physics in Vienna, Austria for their hospitality. Correspondence can be addressed to Fryer at Department of Economics, Harvard University, 1875 Cambridge Street, Cambridge MA 02138 or Holden at MIT Sloan School of Management, E52-410, 50 Memorial Drive, Cambridge MA 02142. E-mail: (Fryer) rfryer@fas.harvard.edu or (Holden) rholden@mit.edu.

1 Introduction

The architecture of political boundaries is at the heart of the political process in the United States.[1] When preferences over political candidates are sufficiently heterogeneous, altering the landscape of political districts can have large effects on the composition of elected officials. For instance, prior to the 2003 Texas redistricting the congressional delegation was comprised of 17 Democrats and 15 Republicans; after the 2004 elections there were 11 Democrats and 21 Republicans.[2] Politically and racially motivated districting plans are believed to be a significant reason for the lack of adequate racial representation in state and federal legislatures, and there is a debate as to whether the creation of majority-minority districts to ensure some level of minority representation have led to less minority-friendly policies (see Shotts, 2002 for an excellent overview and critique).

There are several factors which weigh on the constitutionality of districting plans: (i) equal population (the Supreme Court first established this principle for congressional districts in Wesberry v. Sanders, 376 US 1 (1964)), (ii) contiguity (which is a requirement in 49 state constitutions), and (iii) compactness. The latter consideration – distinct from the mathematical notion of a finite subcover of a topological space – refers to how "oddly shaped" a political district is. The Supreme Court has acknowledged the importance of compactness in assessing districting plans for nearly half a century.[3] Yet, despite its importance as a factor in adjudicating gerrymandering claims, the court has made it clear that no manageable standards have emerged (see the judgment of Scalia J in *Vieth v. Jubelirer*). There is no consensus on how to adequately measure compactness.[4]

In this paper, we propose a simple index of compactness based on the distance between voters within the same political district in a state relative to the minimum such distance achievable by any districting plan in that state – which we coin the relative proximity index.[5] The index satisfies three desirable properties: (i) voters are treated equally (*anonymity*), (ii) increasing the distances between voters within a political district leads to a larger value of the index (*clustering*), and

[1] Article I, §4 of the United States Constitution provides that "The Times, Places and Manner of holding Elections for Senators and Representatives shall be prescribed in each State by the Legislature thereof; but the Congress may at any time by Law make or alter such Regulations, except as to the Places of choosing Senators."

[2] In the US, political boundaries are typically redrawn every 10 years, after the decennial census. The 2003 "mid-decade" redistricting in Texas is a notable exception. The US Supreme Court recently held that this was not unconstitutional in *Jackson, et al. v. Perry, et al.* (docket number 05-276).

[3] The Apportionment Acts of 1842, 1901 and 1911 contained a compactness requirement. In Davis v. Bandemer, 476 US 173 (1986)) Justices Powell and Stephens pointed to compactness as a major determinant of partisan gerrymandering, and Justices White, Brennan, Blackmun and Marshall cited it as a useful criterion. Nineteen state constitutions still contain a compactness requirement (Barabas and Jerit, 2004).

[4] An important argument against the use of compactness as a districting principle is that it may disadvantage certain population subgroups. As Justice Scalia put it in *Vieth v. Jubelirer*: "Consider, for example, a legislature that draws district lines with no objectives in mind except compactness and respect for the lines of political subdivisions. Under that system, political groups that tend to cluster (as is the case with Democratic voters in cities) would be systematically affected by what might be called a "natural" packing effect. See Bandemer, 478 U. S., at 159 (O'Connor, J., concurring in judgment)." First, the courts use compactness as one of several criterion. Second, it is an open question whether or not more compact districting plans have a positive or negative effect on racial or political representation.

[5] For the emprical analysis and characterization of the optimally compact district plan we use Euclidean distance. But since many of our results are proven in an arbitrary metric space one can extend much of the analysis here by using driving distance or what many legal scholars refer to as "communities of interest."

(iii) the index be invariant to the scale, population density, and the number of districts in a state (*independence*). Further, we show that any compactness index that satisfies these properties ranks districting plans identically to the relative proximity index.

The relative proximity index has several advantages over existing measure of compactness. First, it is the only compactness index which permits meaningful comparisons across states. Second, the index does not assume (implicitly or otherwise) that voters are uniformly distributed across political districts. Many previously proposed measures adopt a geometric approach (perimeter length of political districts, e.g.) and fail to consider the distribution of voters within a state. Third, our measure is constructed at the state level. Some measures apply to political districts.[6] Yet, the districting problem is fundamentally about partitioning; the shape of one element of the partition affects the shapes of the other elements. Analyzing individual pieces of a larger partition in isolation can be misleading. Fourth, though our index is simple, it is based on desirable properties that compactness measures should satisfy. Existing measures have been proposed in a relatively ad hoc fashion. At a minimum, our approach is a more principled way of narrowing the field of competing measures.

We apply the index to the districting plans of the 106th congress using tract-level data from the US census. In doing so, we are required to calculate each state's maximal compactness. This number is the denominator of our index. But calculating this number by brute force, enumerating the set of all feasible partitions and maximizing compactness over this set, is impossible.[7] Similar partitioning problems arise in applied mathematics (computer vision), computer science and operations research (the k-way equipartition problem), and computational biology (gene clustering), which have given rise to several important algorithms and candidate functionals. Unfortunately, none of these techniques are directly applicable to our districting problem as they are either designed for very small samples (≈100) or do not require partitions to be of even approximately equal cardinality.

We develop an algorithm for approximating this partitioning problem which is suitable for very large samples and guarantees nearly equal populations in each partition. The algorithm is based on *power diagrams* – a generalization of classic Voronoi diagrams – which have been used extensively in algebraic and tropical geometry (Passare and Rullgard, 2004), condensed matter physics (Richter-Gebert, Sturmfels and Theobald, 2003), and toric geometry/string theory (Diaconescu, Florea, and Grassi, 2002). Power diagrams are a powerful tool to partition euclidean space into cells by minimizing the distance between points in a cell and the centroid of that cell. We prove that maximally compact districts are power diagrams and that the line separating two adjacent districts are perpendicular to the line connecting their centroids, and all such lines separating three adjacent districts meet at a single point. It follows that the resulting districts are convex polygons.

The empirical results we obtain on the compactness of districting plans are interesting and in some cases quite surprising. The five states with the most compact districting plans are Idaho,

[6] See Young (1988), however, and section 2.2 below.

[7] A back of the envelope calculation reveals that, for California alone, the cardinality of this set is larger than the number of atoms in the observable universe.

Nebraska, Arkansas, Mississippi, and Minnesota. The five least compact states are Tennessee, Texas, New York, Massachusetts, and New Jersey. The districting plan that solves the minimum partitioning problem is more than forty percent more compact than the typical districting plan. States which are more compact tend to be states with a larger share of blacks and a larger difference between the percent who vote Republican and Democrat. The latter is intuitive: states with more to gain from altering the design of political districts tend to do it more. Whether or not a state is forced to submit their districting plans to the Department of Justice (under Section 5 of the Voting Rights Act) is also highly correlated with compactness. The rank correlation between the relative proximity index and the most popular indices of compactness, dispersion and perimeter, is -.22 and -.06, respectively.

We conclude our analysis by estimating a counterfactual of the 2000 Congressional elections in Califorina, New York, Pennsylvania, and Texas using optimally compact districts derived from our algorithm. To better understand the impact a strict policy of maximal compactness might have on those elected, we estimate a seat-vote curve for the actual and hypothetical districting plans of each state. We are interested in two parameters: bias and responsiveness. Recall, bias reports, when the vote is split, twice the difference between the seat share the Democrats get and 50%. Responsiveness is the fraction of seats the Democrats get if the average vote goes up 1%.

The results of this exercise are quite illuminating. California, New York, Pennsylvania, and Texas all have substantially more responsive seat-vote curves under our new partition, but bias is unchanged. These results prove that maximally compact districts would have a statistically significant effect on voting outcomes – making election outcomes more responsive to actual votes.

The structure of the paper is as follows. Section 2 provides a brief legal history of compactness and an overview of existing measures. Section 3 presents the relative proximity index, discusses its properties, and proves our main result. Section 4 implements the index using data from the 106th congress. Section 5 provides a counterfactual estimate of the congressional elections in four large states using the partitions derived from our index. Section 6 concludes. There are two appendices. Appendix A contains proofs of all technical results. Appendix B provides a guide to programs to calculate the Relative Proximity Index.

2 Background and Previous Literature

2.1 A Brief Legal History of Compactness

Compactness has played a fundamental role in the jurisprudence of gerrymandering, both racial and political. Since *Gomillion v. Lightfoot* 364 U.S. 339 (1960), where the court struck down Alabama's plan to redraw the boundaries of the city of Tuskegee, the court has recognized compactness as a relevant factor in considering racial gerrymandering claims. In *Gomillion* the court referred to the proposed district as "an uncouth 28-sided figure." Although *Gomillion* is considered by many to be a jurisprudential high-water mark, the role of compactness in considering racial gerrymandering claims has been affirmed in other decisions.[8] As Justice O'Connor put it: "we believe that reapportionment is one area in which appearances do matter."

Compactness has also played an important role in partisan gerrymandering claims. It has been recognized by the court as a "traditional" districting principle. In *Davis v. Bandemer*, Justices Powell and Stevens described compactness as a major criterion (at 173), and Justices White, Brennan, Blackmun and Marshall described it as an important criterion (at 2815). In *Vieth*, the plurality acknowledged compactness as a traditional districting principle. Justice Kennedy, in his concurring opinion, states that compactness is an important principle in assessing partisan gerrymandering claims: "We have explained that "traditional districting principles," which include "compactness, contiguity, and respect for political subdivisions," are "important not because they are constitutionally required...but because they are objective factors that may serve to defeat a claim that a district has been gerrymandered on racial lines." ...In my view, the same standards should apply to claims of political gerrymandering, for the essence of a gerry-mander is the same regardless of whether the group is identified as political or racial."

Despite different views about what a judicially manageable standard is or might be, the court has been unanimous that it must include some notion of compactness.

2.2 Existing Measures of Compactness

There is a large literature in political science on the measurement of compactness. Niemi et al (1990) provide a comprehensive account of the various measures which have been proposed (see also Young (1988)).[9] Niemi et al (1990) classify existing measures into four categories: (i) dispersion measures, (ii) perimeter measures, (iii) population measures, and (iv) other miscellaneous measures. We draw heavily on their summary and classification.

One class of dispersion measures are based on length versus width of a rectangle which circumscribes the district (Harris, 1964; Eig and Setizinger, 1981; Young, 1988). A second uses circumscribing figures other than rectangles and considers the area of these figures.[10] At least two "moment-of-inertia" measures have been suggested. Schwartzenberg (1966) and Kaiser (1966) consider the variance of the distances from each point in the district to the districts areal center. Boyce and Clark (1964) consider the mean distance from the areal center to a point on the perimeter reached by equally spaced radial lines.

A second set of measures are those based on perimeters. The sum of perimeter lengths was suggested by Adams (1977), Eig and Setizinger (1981) and Wells (1982), but this measure is potentially intractable for reasons highlighted in the classic work of Mandelbrot (1967) on the length of the coastline of Great Britain. In fact, a fractal dimension based measure was proposed by

[8] In *Shaw v. Reno* 113 S. Ct. 2816. 92-357 (1993), the court upheld a challenge to North Carolina's redistricting plan on the basis that the ill-compactness of the districts was indicative of racial gerrymandering. See also *Thornburg v. Gingles* 478 U.S. 30 (1986) or *Growe v. Emison* 278 U.S. 109 (1993).

[9] Some of these measures were originally proposed for purposes other than to do with legislative districts - but were later applied by other authors to that issue. We cite the original authors.

[10] Reock (1961) proposes a circle, Geisler (1965) a hexagon, Horton (1932) and Gibbs (1961) a circle with diameter equal to the districts longest axis, still others use the smallest convex figure (see Young (1988)).

Knight (2004). Various authors have proposed measures which compare the perimeter to the area of the district. Cox (1927) considers the ratio of the district area to that of a circle with the same perimeter.[11]

There are three population based measures. Hofeller and Groffman (1990) propose two: the ratio of the district population to the convex hull of the district, and the ratio of the district population to the smallest circumscribing circle. Weaver and Hess (1963) suggest the population moment of inertia, normalized to lie in the unit interval.

Niemi et al's (1990) final miscellaneous category includes three measures: (1) the absolute deviation of district area from average area in the state (Theobald 1970); a measure based on the number of reflexive and non-reflexive interior angles (Taylor 1973); and the sum of all pairwise distances between the centers of subunits of the district, weighted by subunit population (Papayanopolous 1973). Finally, Mehrotra, Johnson and Nemhauser (1998) use a branch-and-price algorithm to compute a districting plan for South Carolina. Their objective function is how far people are from a graph theoretic measure of the center of the district.

All of these measures either fail to account for the population distribution or are not invariant to geographical size. As such, meaningful comparisons across states or time cannot be made.

3 The Relative Proximity Index

3.1 Preliminaries

Let $\mathbf{S}$ denote a collection of states with typical element $S \in \mathbf{S}$. A finite set S, whose elements we call individuals or voters, is a metric space with associated distance function $d_{ij} \geq 0$, which measures the distance between any two elements $i, j \in S$. Let $V_S = \left\{v_1^S, ..., v_n^S\right\}$ denote a finite partition of S into elements $v_i \in V_S$ which we shall refer to as "voting districts", or "districts". We will routinely refer to the partition V_S as a "districting plan" for state S. We restrict voting districts to be equal in size, up to integer rounding.[12] In symbols: $\left|v_i^S\right| \in \{\lfloor |S| / |V_S| \rfloor, \lceil |S| / |V_S| \rceil\}$ for all $v_i^S \in V_S$, where $\lceil x \rceil = \inf \{n \in \mathbb{Z} | x \leq n\}$ and $\lfloor x \rfloor = \sup \{n \in \mathbb{Z} | n \leq x\}$. Let $\mathcal{V}_S$ denote the set of all partitions of S which satisfy this restriction. We say a districting plan V_S is *feasible* if and only if $V_S \in \mathcal{V}_S$.

Definition 1 *A compactness index for a state S is a map $c : V_S \mapsto \mathbb{R}_+$.*

3.2 The Relative Proximity Index

Consider voter i in element $v \in V_S$ and define:

[11] For variants of Cox (1927) see Attneave and Arnoult (1956), Horton (1932), Schwartzberg (1966), or Pounds (1972).

[12] This was first held as a requirement by the Court in *Baker*, and is becoming a very strict constraint. For instance, a 2002 Pennsylvania redistricting plan was struck down because one district had 19 more people (not even voters) than another. The 2004 Texas redistricting had each district with the same number of people up to integer rounding. Yet, the population may grow at drastically different rates across political districts between redistrictings. For instance, in the 2000 census, a typical state had a 23% difference in the population of its smallest and largest district.

$$\pi\left(V_S\right) = \sum_{v \in V} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 \tag{1}$$

Similarly, let $V_S^* = \underset{V_S \in \mathcal{V}_S}{\arg\min} \left\{\pi\left(V_S\right)\right\}$. The Relative Proximity Index (RPI), for a partition of state S, V_S, is given by

$$RPI = \frac{\pi(V_S)}{\pi(V_S^*)}.$$

The RPI is well defined so long as $\pi(V_S^*) \neq 0$ which holds so long as all voters are not located at the same point. In the non-degenerate case, the RPI ranges from 1 to infinity; higher numbers indicate less compactness. The index has an intuitive interpretation: a value of 3 implies that the current districting plan is roughly 3 times less compact than a state's maximal compactness.

3.3 An Example

[insert figure 1]

Consider the state depicted in Figure 1. The nodes represent voters. There are two voting districts separated by the bold dashed line. Voters are spread evenly across the state; each adjacent voter is 1 kilometer apart. Voter 1 is 1 kilometer away from voters 2 and 4, $\sqrt{2}$ kilometers away from voter 5, $\sqrt{5}$ kilometers away from voter 6, and so on.

There are two steps involved in calculating the Relative Proximity index. First, we calculate the numerator. For voter 1 the sum of squared distances is 5, since she is 1 kilometer away from voter 2 and 2 kilometers away from voter 3–and they are the only other voters in her district. For voter 2 the total is $1^2 + 1^2 = 2$ and for voter 3 it is $1^2 + 2^2 = 5$. Voters 4,5 and 6 are symmetric to voters 1,2 and 3 respectively. Thus the numerator of our index is $2(5 + 2 + 5) = 24$.

The second step in calculating RPI is to account for state specific topography. This will represent the denominator of our index. There are nine other feasible partitions in addition to $\{\{1, 2, 3\}, \{4, 5, 6\}\}$.[13] We perform the same calculation as above for each of those partitions and then take the min of these ten values. The minimizing partition is $\{\{1, 4, 5\}, \{2, 3, 6\}\}$–although $\{\{1, 2, 4\}, \{3, 5, 6\}\}$ achieves the same value. That value turns out to be $2\left(1^2 + 2 + 1^2 + 2 + 1^2 + 1^2\right) = 16$. The index is thus $24/16 = 3/2$.

The example provides a snap-shot of the Relative Proximity Index and previews some of its properties. For instance, because the index is calculated relative to a state specific baseline, neither the size of states nor their population density can solely alter the index. If we increased the distance between any two nodes in figure 1 to 2 kilometers, the index would not change. Similarly, if we imputed 10 more individuals to each node – thinking of them in terms of neighborhoods rather than households – the index would be unaltered.

[13] They are: $\{\{1, 2, 4\}, \{3, 5, 6\}\}$, $\{\{1, 2, 5\}, \{3, 4, 6\}\}$, $\{\{1, 2, 6\}, \{3, 4, 5\}\}$, $\{\{1, 3, 4\}, \{2, 5, 6\}\}$, $\{\{1, 3, 5\}, \{2, 4, 6\}\}$, $\{\{1, 3, 6\}, \{2, 4, 5\}\}$, $\{\{1, 4, 5\}, \{2, 3, 6\}\}$, $\{\{1, 4, 6\}, \{2, 3, 5\}\}$, $\{\{1, 5, 6\}, \{2, 3, 4\}\}$.

3.4 Three Desirable Properties of a Compactness Index

We now describe three properties which any compactness index should satisfy and discuss each in turn.

Axiom I, an anonymity condition in the same spirit as that typically used in social choice theory (Arrow, 1970), requires that all individuals be treated equally. That is, any compactness index should not depend on the particular identities (race, political affiliation, wealth, etc.) of voters.

Axiom I (Anonymity) Consider a state S with associated partition V and compactness index $c(V,S)$. For any bijection $h : S \to S$ and compactness index $c_h(V,S)$, $c_h(V,S) = c(V,S)$.

Compactness is fundamentally a mathematical partitioning problem; deciding who to group with whom in a political district. Clustering is the quintessential objective (Bartal, Charikar, and Raz, 2001).[14] Our second axiom requires that if two states with the same number of voters, voting districts, and the same value for the minimum partitioning problem have different weighted intra-district distances, then the state with the larger value is less compact.

Axiom II (Clustering) Let $\gamma_k = \sum_{i,j \in v} \alpha_{ij} \left(d_{ij}\right)^{\delta}$, for $k = \{1, ..., n\}$ and let $g\left(\gamma_1, ..., \gamma_n\right) : \mathbb{R}^n \to \mathbb{R}$ be a monotonic, increasing function. Consider two states, S_1 and S_2 and partitions V and V' respectively such that S_1 and S_2 have: the same number of voters, the same number of districts and

$$\min_{V \in \mathcal{V}_{S_1}} g_{S_1}\left(\gamma_1, ..., \gamma_n\right) = \min_{V \in \mathcal{V}_{S_2}} g_{S_2}\left(\gamma_1, ..., \gamma_n\right).$$

Then

$$g_{S_1}\left(\gamma_1, ..., \gamma_n\right) > g_{S_2}\left(\gamma_1, ..., \gamma_n\right) \Longrightarrow c\left(V, S_1\right) > c\left(V', S_2\right).$$

Our final axiom requires that any measure of compactness of a state be insensitive to its physical size, population density, and number of districts. This is vital for making cross-state comparisons of districting plans.

Density independence means that if we replicate a state by multiplying the number of people in each household by λ, the index of compactness is unaltered. For instance, when comparing two voting districts (Cambridge, MA and New York, NY, e.g.) who differ in their population density, the index provides the same cardinal measure of compactness.

Scale independence provides a similar virtue, permitting comparisons across states that differ in the distances between individuals (Massachusetts and Texas, say), allowing one to increase the distances between all individuals in a state by a constant with no resulting change in the index. Independence with respect to the number of districts is also vital in making cross-state comparisons.

Before stating the property formally, we need some further notation. We say that a state $\widehat{S}$ is an n-Replica of S if and only if $\forall i \in S, \exists j_1, ..., j_n \in \widehat{S}$ such that $d_{ij} = 0, \forall i$ and $d_{j_i j_k} = 0, \forall i, k$.

[14] Other common objectives are distance from the geographic centroid of each partition or distance from a representative (typically the center of a cluster and not necessarily the center of the partition).

It is also useful to have a shorthand for the realized value of the minimum partitioning problem. Consider two partitions of state S, V and V' with ρ and ρ' elements respectively. Let $V_S^{\min_\rho}$ and $V_S^{\min_{\rho'}}$ be the respective minimizing partitions.

Axiom III (Independence) *Consider $S, \widehat{S} \in S$ with cardinality $|S|$ and $\left|\widehat{S}\right|$ respectively.*

1. *(Scale) If $d_{ij} = \lambda d_{ij}$, for all $i, j \in S, \widehat{S}$. Then $c(V, S) = c(V, \widehat{S})$,* for all V.
2. *(Density) If $\left|\widehat{S}\right| = \lambda |S|$ and $\widehat{S}$ is a λ-replica of S then $c(V, S) = c(V, \widehat{S})$,* for all V.
3. *(Number of Districts)*

$$\text{If } \frac{\sum_{v \in V_S^\rho} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2}{V_S^{\min_\rho}} = \frac{\theta \sum_{v \in V_S^{\rho'}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2}{V_S^{\min_{\rho'}}} \Longrightarrow c\left(V, S\right) = \theta c\left(V', S\right).$$

3.5 The Main Result

Let $O_c = (\mathbb{R}_+, \succeq)$ denote the ordered set generated by the relative proximity index c, and let $O_{\widehat{c}}$ denote the ordered set over elements $V_S \in \mathcal{V}_S$ generated by any other compactness index. We say that two indices, c and $\widehat{c}$, are ordinally isomorphic if $O_c = O_{\widehat{c}}$. We are now equipped to state our main result. The proof of this, as with all others, can be found in the appendix.

Theorem 1 *(1) The RPI satisfies Anonymity, Clustering, and Independence; (2) Suppose $\delta = 2$ and $g_{S_i}(\cdot)$ is symmetric for all i, then any compactness index which satisfies Anonymity, Clustering and Independence is ordinally isomorphic to the Relative Proximity Index.*

Proof. See Appendix A ■

The structure of the proof is as follows. The first part simply verifies that the RPI satisfies the three axioms. Because we weight individuals identically, anonymity is satisfied. The numerator of the index is almost the definition of a special case of clustering. The normalization provided by the denominator of the index ensures that independence is satisfied.

The difficulty is in showing that any other compactness index which satisfies these axioms ranks districting plans identically to the RPI. We accomplish this by transforming states so that they have the same number of voters and districts, and have equal minimum values. Independence and anonymity allows for such a transformation. The clustering axioms when $\delta = 2$ and $g_{S_i}(\cdot)$ is symmetric for all i, then implies that districting plans can then be ranked by their total intra-cluster pairwise distances. Any compactness index that ranks plans differently than RPI, after our transformation, must violate clustering.

4 Implementing the Relative Proximity Index

In this section, we apply the relative proximity index to the districting plans of the 106th congress.

4.1 The Minimum Partitioning Problem

Calculating the denominator of the relative proximity index is a complicated combinatorial problem. When partitioning n voters into d districts the number of feasible partitions is $\left(\frac{(n-1)!}{(n/d-1)!(n-n/d)!}\right)^{d-1}$. So, for California alone, using data at the tract level, involves $n = 6,800$ and $d = 53$. The cardinality of the set of feasible partitions is $78.4 \times 10^{59,351}$. Technically speaking, the problem is *NP-hard*.

Similar problems arise in fields such as applied mathematics (computer vision), computer science and operations research (k-way equipartitioning problem), and computational biology (gene clustering). The celebrated Mumford-Shah functional is a candidate functional designed to segment images (Mumford and Shah, 1989). The structure of the functional contains two penalty functions: one to ensure that the continuous approximation is close to the discrete problem, and another to penalize perimeter length. While the Mumford-Shah functional is a powerful tool for myriad problems, it cannot guarantee even nearly equal population size across districts.

If our objective function was simply distance, rather than distance squared, the problem is precisely the k-way equipartition problem which has received considerable attention in computer science and related to a literature in computational biology employing minimum spanning trees to partition similar genes into clusters.[15] Good algorithms for the k-way equipartition problem when sample sizes are small (≈ 100) can be found in Ji and Mitchell (2005) and Mitchell (2003). This restriction makes these algorithms impractical for our purposes.

Below, we develop an algorithm to approximate the minimum partitioning problem for large samples, based on *power diagrams* (a concept we make precise below), that guarantees nearly equal populations in each partition and runs in $O(m \log(n))$ time, where n is the number of voters and m is the number of districts in a state.

4.2 Optimally Compact Districting Plans and Power Diagrams

In this section, we show that optimally compact districting plans are *power diagrams*, a generalization of Voronoi diagrams due to Aurenhammer (1987). Consider a set of *generator points* $m_1, \ldots, m_n$ in a finite dimensional Euclidean space. The *power* of a point/voter $x \in S$ with respect to a generator point m_i is given by the function $pow_\lambda(x, m_i) = \|x - m_i\|^2 - \lambda_i$, where $\|\cdot\|$ is the Euclidean norm. The total number of voters assigned to generator point m_i is called its *capacity*, denoted K_{m_i}. A *power diagram* is an assignment of voters to generator points such that point x is assigned to generator point m_i if and only if $pow_\lambda(x, m_i) < pow_\lambda(x, m_j)$ for all $j \neq i$. Let the points assigned to generator point m_i be denoted D_i, which is referred to as a *cell*. Note that no two D_is can intersect, and furthermore, every $x \in S$ is in some D_i, and hence $\{D_1, \ldots, D_n\}$ is a partition of S. Note also that the dividing line between cells D_i and D_j in a power diagram satisfies $\|x - m_i\|^2 - \|x - m_j\|^2 = \lambda_i - \lambda_j$.

[15] Without the constraint that each district have an equal number of voters the problem is the *min-sum k-clustering problem* which was shown by Sahni and Gonzales (1976) to be NP-complete. An approximation for it in a general metric space which runs in $n^{O(1/e)}$ time has been found by Bartal, Charikar and Raz (2001). It is also closely related to the classic graph partitioning problem, which is also known to be NP-hard.

When $\lambda_i = \lambda$ for all i then the power diagram is a Voronoi diagram. Power diagrams are thus a generalization of Voronoi diagrams.

Definition 2 *An optimally compact districting plan for state S is a feasible districting plan, V_S, with an associated total distance $\sum_{v \in V_S} \sum_{i,j \in v} (d_{ij})^2$ such that there does not exist another feasible districting plan, V_S' with an associated total distance $\sum_{v \in V_S'} \sum_{i,j \in v} (d_{ij})^2$ such that $\sum_{v \in V_S'} \sum_{i,j \in v} (d_{ij})^2 < \sum_{v \in V_S} \sum_{i,j \in v} (d_{ij})^2$.*

We can now state our second key result.

Theorem 2 *Optimally compact districting plans are power diagrams.*

Proof. See Appendix A. ■

This theorem follows from three lemmas which partially characterize an optimal districting plan and establish that these characteristics imply a power diagram. The first lemma shows that our objective function is equivalent to a variant of the k-means objective function. This is important because it allows one to focus attention on district centroids.

The second lemma shows that any pair of districts are separated by a line perpendicular to a line connecting their centroids. This separating line is the locus of points at which the power of the two centroids are equal. It represents all points in which one is indifferent between placing voters in one district and the other. Finally, we establish that all such lines separating any three adjacent districts meet at a single point; they are concurrent.

To see that these properties imply a power diagram, recall that a power diagram is a set of lines dividing a euclidean space into a finite number of cells. The line separating two adjacent cells are such that the power of the points along this locus is equal to their respective centroids. And the power of a point is measured as a function of the difference between a point and the centroid of its district – which we have already established is equivalent to our objective function. It is important to note that if the line separating two adjacent districts was not perpendicular to the line connecting their centroids then one could not be indifferent between points being in one district and the other everywhere along the line. This holds for all such pairs of districts, which implies concurrent lines. Taken together, these imply that optimally compact districtings are power diagrams[16]. Notice, since all subsets of a convex set formed by drawing straight lines are convex, it follows that the resulting districts must be convex polygons.

Theorem 2 provides an important insight for building an algorithm, allowing us to use all we know about a partial characterization of optimally compact districts. There are three important

[16] Aurenhammer et al. (1998) prove a closely related theorem, taking squared distance from the centroid as the objective function. Their proof proceeds by showing that if an algorithm can be designed to find a power diagram then it is an optimal partition. By contrast, we provide a constructive proof based on the parallel and concurrent line lemmas. We could, of course, state our lemma on the equivalence of the objective functions and then appeal to their result, but our current proof provides more information about optimal districtings.

caveats. First, we have not yet proven that there is a unique power diagram for every set of starting values. Second, we are only able to map optimal districting plans into power diagrams when distance is quadratic, because this guarantees that optimal districting involves straight lines. Mathematically, this is an obvious limitation. Practically, however, it boils down to assuming that courts punish outliers in a district more. Given this assumption, we are hard pressed to find a principled reason for courts to prefer higher order exponents.

Third, power diagrams do not guarantee a global optimum to the minimum partitioning problem because their structure depends on exogenously given starting values.

[insert figure 2]

Panel A of figure 2 depicts the optimally compact districting plan for a hypothetical state. There are nine voters, arranged so the state is a lattice. The stars represent centroids of the resulting districts. Note that the line separating districts 1 and 2 is perpendicular to a line connecting their centroids (the same is true for districts 1 and 3, and also 2 and 3). This is an illustration of the Perpendicular Line Lemma alluded to above. The Concurrent Line Lemma is also illustrated by the intersection of the lines separating districts 1,2 and 3 at a single point. The partition depicted is indeed the globally optimal partition. Once one knows that, the centroids of the districts are easy to compute.

In our problem, however, we do not know the optimal districts in advance, and so we must choose generator points which will not in general be the centroids of the optimal districting plan. An important part of the approximation problem is selecting and improving upon the generator points. To illustrate this point, consider panel B of figure 2 which chooses alternative generator points than those used to partition the panel A. The generator point used for district 1 differs from that used above resulting in four voters being placed in district 1 and only 2 in districting 2, thereby violating the equal size constraint.

4.3 An Algorithm Based on Power Diagrams

The algorithm we propose is a modification of the second algorithm presented in Aurenhammer et. al (1998). Since we know by Theorem 2 that local optima of the RPI are power diagrams, we search within the set of power diagrams for one that is a feasible districting. However, as power diagrams are generated around sites, which we call $z_1, \ldots, z_n$, it is necessary to update the locations of the sites as well as the design of the districts.

First we explain the (Aurenhammer et al, 1998) algorithm for finding a power diagram which minimizes $\Psi_{z_1,\ldots,z_d}(D_1, \ldots, D_d)$ with $|D_i| \approx n$ for all i. Since a power diagram is defined by its sites and their weights, $\lambda_1, \ldots, \lambda_d$, assuming fixed sites each district D_i is a function of $\lambda_1, \ldots, \lambda_d$, or $D_i = D_i(\lambda_1, \ldots, \lambda_d)$. We suppress this dependence for simplicity. Let

$$\xi(\lambda_1, \ldots, \lambda_d) = \sum_{i=1}^{d}(n - |D_i|) \cdot \lambda_i + \Psi_{z_1,\ldots,z_d}(D_1, \ldots, D_d).$$

Aurenhammer et al, (1998) simplifies the problem by continuing as if each D_i does not change locally with respect to each λ_i everywhere, as this is true almost everywhere (at all but finitely many points). Therefore, $|D_i|$ and $\Psi_{z_1,\ldots,z_d}(D_1,\ldots,D_d)$ are locally constant with respect to λ_i, so,

$$\frac{\partial \xi}{\partial \lambda_i} = n - |D_i|.$$

Let $\Lambda = (\lambda_1, \ldots, \lambda_d)$. Using some choice of Λ_0, we can update it by gradient descent,

$$\Lambda_{t+1} = \Lambda_t + \epsilon_t \cdot \nabla \xi(\Lambda_t).$$

In our implementation we set Λ_0 to be the zero vector. It remains to pick the step sizes $\{\epsilon_t\}_{t\geq 0}$. To do this, one first determines an overestimate of the minimum value of ξ, call it $\overline{\xi}$. This can be done by setting $\overline{\xi} = \Psi_{z_1,\ldots,z_d}(D_1,\ldots,D_d)$ for any feasible districting $(D_1,\ldots,D_d)$. We use the notation $D_i(\Lambda_t)$ to mean one of the districts induced by the power diagram weights contained in the vector Λ_t, and let:

$$\epsilon_t = \frac{\overline{\xi} - \xi(\Lambda_t)}{\sum_{i=1}^{d} |D_i(\Lambda_t)|^2}$$

This step size is iterated until the minimum is either reached or missed, which happens when $\sum_{i=1}^{d} |D_i(\Lambda_t)| \cdot |D_i(\Lambda_{t+1})| > 0$. Then, $\overline{\xi}$ is updated by solving the equation:

$$\frac{\overline{\xi} - \xi(\Lambda_t)}{\sum_{i=1}^{d} |D_i(\Lambda_t)|^2} = \frac{\overline{\xi} - \xi(\Lambda_{t+1})}{\sum_{i=1}^{d} |D_i(\Lambda_{t+1})|^2}$$

ϵ_{t+1} is chosen accordingly. This algorithm is repeated until the $|D_i|$'s are within some predetermined error bound around n.

Once optimal districts $D_1,\ldots,D_d$ for sites $z_1,\ldots,z_d$ are chosen, by Lemma 7 (see Appendix A) the function $\Psi_{z_1,\ldots,z_d}(D_1,\ldots,D_d)$ is improved by moving the z_i's to the centroids of the D_i's and keeping the $\lambda_1,\ldots,\lambda_d$ constant. Yet, all of the D_i's are not necessarily of size n, so they need to be adjusted by the above procedure. This process is repeated until moving the $z_1,\ldots,z_d$ still leaves the sizes of the D_i's within the prescribed error bound.

Note: The algorithm described in Aurenhammer et al. (1998) tends to fail when one of the districts is randomly set to size 0. Our solution to this issue was to move z_i to a random new location if $|D_i|$ became zero during any point in the process. Random new locations were chosen using a uniform distribution function ranging from the minimum to the maximum of the longitude and the latitude of the state in question.

4.4 The Compactness of Political Districting Plans of the 106th Congress

The ideal data to estimate the relative proximity index would contain the geographical coordinates of every household in the US, their political district, some measure of distance between any two households within a state, and a precise definition of communities of interest. This information is not available.

In lieu of this, we use tract-level data from the 2000 US Census from the Geolytics database which contains the latitude and longitude of the geographic centroid of each tract, the political district each centroid is in, and its total population.[17] Census tracts are small, relatively permanent statistical subdivisions of a county. The spatial size of census tracts varies widely depending on the density of settlement, but they do not cross county boundaries. Census tracts usually have between 2,500 and 8,000 persons and, when first delineated, are designed to be homogeneous with respect to population characteristics, economic status, and living conditions. The latter consideration is our main interest in using this level of aggregation (relative to blocks or block-groups), as census tracts are more likely to contain some notion of communities of interest.

An important consideration in the application of RPI is how to handle tracts with different density. The equal representation constraint – districting plans must have the same number of individuals in each district up to integer rounding – is predicated on individuals, not tracts. Our algorithm, described below, addresses this issue by allowing one to divide tracts into arbitrarily small units. There is an important trade-off between computational burden and the variance in population across districts, a burden that lessens with technological progress.

For ease of implementation, we have chosen not to split any tracts. As a robustness check, we split tracts of small states into 4 smaller parts and assigned them to the same longitude and altered their latitude by 0.001 degrees. In all cases, accuracy (and computing time) were substantially increased with little effect on the RPI.

To calculate the RPI for each state, we begin with the numerator of the index: $\sum_{v \in V} \sum_{i,j \in v} (d_{ij})^2$, where i and j are population centroids of tracts and v are voting districts. We weight the total distances by the population density of each tract. An identical calculation is performed for the denominator, but V is constructed by our power diagram algorithm.

The empirical results we obtain on the compactness of districting plans are displayed in Table 1. The first column list each state, the second provides the relative proximity index, the third and fourth give the maximum deviation from equal partitions in the actual data and that resulting from our algorithm – an indication of the degree to which the equal size constraint holds. The final columns report the results from a bootstrapping technique which we describe below. It is important to realize that for every state, the elements of our partitions are more balanced than what appears in the actual districting plans. Further, the largest deviation from equal partitions in the actual data (Florida 0.46) is substantially larger than our largest deviation (California 0.22).

Table 1 illustrates that the five states with the most compact districting plans are Idaho, Washington, Arkansas, Mississippi, and New Hampshire. The five least compact states are Idaho, Nebraska, Arkansas, Mississippi, and Minnesota. The five least compact states are Tennessee, Texas, New York, Massachusetts, and New Jersey. The districting plan that solves the minimum

[17] For roughly 5,000 census tracts, information on congressional district was not provided. In these cases, we mapped the coordinates of the centroid of the tract and manually keypunched the congressional district to which it belonged.

partitioning problem is more than forty percent more compact than the typical districting plan. The rank correlation between the relative proximity index and the most popular indices of compactness, dispersion and perimeter, is -.22 and -.06, respectively.

Axiom III ensures that the RPI can be compared across states, but it does not guarantee that the distribution of RPI values across states are the same. It is entirely plausible that Texas finds it "easier" (a lower percentile of the distribution of RPI values from feasible partitions) to obtain a given value of RPI than say, Florida. Thus, gleaning an understanding of how "sensitive" RPI values are for a given state is difficult.

To try and address this issue, we calculated 200 RPI values for each state by randomly generating starting values for the algorithm. Columns 5 and 6 in Table 1 report the means and associated standard deviations from this process. The final column reports what percentile in the distribution our original RPI value lies, if the distribution of RPI values is assumed to be normal. In all but one case, our original estimates are higher than the mean of the simulated distribution and in most cases, under the normality assumption, we are at the far extreme of the right tail of the distribution. There are four notable exceptions: Oklahoma, Oregon, Rhode Island, and Wisconsin. In these states, our estimate of RPI is at the median or below in the simulated distribution. This is likely due to the fact that the current partitions of these states generate starting values that are highly non-optimal. To obtain maximal compactness in these states, a significant restructuring is likely needed.

To understand what state demographics are correlated with compactness, we estimate a state-level OLS regression where the dependent variable is the RPI and the independent variables are percent black, percent Asian, percent Hispanic, population density, difference in presidential vote shares between Democrats and Republicans, and whether or not the state is required to submit their districting plans to the Department of Justice under the preclearance provision of section 5 of the Voting Rights Act. States which are more compact tend to be states with a larger share of blacks and a larger difference between the percent who vote Republican and Democrat. The latter is intuitive: states with more to gain from altering the design of political districts tend to do it more. Whether or not a state is forced to submit their districting plans is also highly correlated with compactness. Consistent with Axiom II, RPI is uncorrelated with population density.

Beyond the technical considerations, perhaps the best evidence in favor of our approach can be illustrated visually. Figures 3-11 present side-by-side comparisons of congressional district maps for actual districting plans and those obtained from our algorithm.[18] Figures 3 and 4 illustrate this comparison for the least and most compact states, Tennessee and Idaho, respectively. Tennessee, under the current districting plan, resembles the salamander shaped districts drawn by Eldridge Gerry they gave rise to name "gerrymandering." Under the algorithm, however, Tennessee is transformed into a neat set of convex polygons. Idaho is at the other extreme. Because it need only cut the state into two equal parts, the existing cut and our preferred cut are very similar to one another. Further, our partition provides a more equitable distribution of voters across the districts,

[18] A complete set of maps are available at http://www.economics.harvard.edu/faculty/fryer/fryer.html

which explains why the calculate RPI is slightly less than one.

These figures illustrate 3 key points. First, the geometric properties discussed above (the perpendicular and concurrent line lemmas and the convexity of political districts) are immediately apparent. Second, those states which rank relatively high (resp. low) in terms of the RPI appear to quite different (resp. similar) to the partition resulting from our algorithm. Third, figures 5 and 8 (Hawaii and Nevada), suggest that communities of interest are an important consideration. In the actual plans, Honolulu and Las Vegas are their own districts while the rest of the state is contained in the other. The issues faced by residents of the outer islands might well be more similar than those of residents in Honolulu. This serves to highlight why compactness is only one factor which weighs on the redistricting question. RPI in its current implementation ignores this consideration. An RPI with a more general notion of distance or carefully selected starting values for the power diagram can address this issue.

5 Election Counterfactuals

Thus far, we have derived an index of compactness, shown how one implements the index, and provided some basic facts about the most and least compact districting plans and what correlates with these plans. We conclude our analysis with some suggestive evidence on the impact of maximally compact districting plans on election outcomes in four large states.

In winner-take-all election contests, such as elections for representatives for the U.S. Congress and for electoral votes for the U.S. Presidency, the winner of a contest is determined by which candidate receives the plurality of the votes. In most of these cases, only the top two parties need to be considered, yielding an easy condition for an election win in a district.

Assuming there are n districts, labeled $i \in [1, ..., n]$, let ϕ_i denote the proportion of the two-party vote received by the candidate from the first party (in examples to follow, the Democratic Party). The candidate's victory can then be expressed as $s_i = w_i \mathbb{I}(\phi_i > \frac{1}{2})$, where w_i denotes how many seats are determined by the vote; 1 for single-member districts, or 3 or more for the Electoral College, for example. Two important summary statistics are the average district vote, $\Phi = \frac{1}{n}\sum_{i=1}^{n} \phi_i$, and the seat share, $S = \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} w_i}$.

Many other statistics can be generated using the vote and seat outcomes directly, but we are particularly interested in partisan bias and responsiveness. Namely: $\mathit{Bias} = 2E(S|\Phi = 0.5) - 1$ estimates the deviation from the median share of seats if each side receives an identical average district vote; $\mathit{Responsiveness} = \frac{dS}{d\Phi}|\Phi$ estimates how a small shift in the average district vote would translate into a shift in the share of seats. This estimate is taken either at the observed average district vote or the median vote.

5.1 Data and Statistical Framework

We use voter tabulation district (VTD) level election return data from US elections of the 105th and 106th congress for four large states; California, New York, Pennsylvania, and Texas. These states were chosen because of their large number of congressional districts (roughly 30 or greater) and the availability of vote shares by VTD. There are approximately 300 VTDs in a typical congressional district, though there is substantial variation. In our data, for instance, California has 7,000 VTDs for 50 districts; Texas has 8,000 for 30. Pennsylvania has 9,000 for 20, and New York it's 13,000 for 30 districts.

The intuition behind our approach is straightforward. Consider figure 9, which depicts the existing districting plan of New York and the plan derived from our algorithm. To fix ideas, concentrate on the western portions of the state. There are roughly 433 VTDs in each congressional district in New York. Suppose an election takes place. Currently, a congressional representative is chosen by aggregating the votes from the VTDs within each district. In figure 9, this amount to adding votes from roughly 433 voting centers in districts 27 through 31. Now, suppose we want to estimate how these representatives will change if the districting plan were drawn to maximize compactness. To do this, we simply take note of which VTDs are in the new partitions and aggregate within each new district. In short, we disaggregate down to the VTD level, take note of the new districting lines, and then aggregate up taking these boundaries into account. As before, the winner of the new districts (in Figure 9 this now amounts to district 4, 6, 8, and 17) is determined by aggregating the votes from VTDs.

There are a few complications. First, we need to assign candidates to the new districts in a reasonable manner. Second, we need to take into account the results of previous elections and whether or not the candidate is an incumbent – as both of these factors weigh heavily on the prediction of future elections. Third, we need to think about how to get standard errors on our estimates.

To formalize the intuition above, we employ techniques from elementary Bayesian statistics developed in Gelman and King (1994). We provide a short synopsis of their approach below.[19] The crux of the Gelman-King method is a linear model with two distinct error components of the form:

$$\phi_i = X\beta + \gamma_i + \varepsilon_i. \tag{2}$$

The vector X consists of an intercept term, results from the previous election, and an incumbent dummy.

To derive precise predictions in this framework, more structure has to be placed on the error terms. Let $\gamma_i \sim N(0, \sigma_\gamma^2)$ represent the systematic error component; an expression of the unobserved variables that took place before the election campaign began and would be identical if the election were to be re-run again. This might include the result in the previous election, the race of the candidates, or a relevant change in election law. The unpredictability of the behavior of voters is also a source of systematic error.

The second source of error is a random component which can be explained by random events during the election, such as the weather on election day or the reaction of the public to an unintentional gaffe. Let $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.

[19] For more details, see Gelman and King (1994).

The key assumption in the Gelman-King Method is to express these errors in terms of two parameters: σ^2, the sum of the individual variances σ_γ^2 and σ_ε^2, and λ, the proportion of the total variance attributed to the systematic component; $\lambda = \sigma_\gamma^2/(\sigma_\gamma^2 + \sigma_\varepsilon^2)$.

Estimating λ and σ^2

In practice, a districting map is constant over a series of elections. Thus, λ and σ^2 are found by taking the mean of individual estimators from each year. In each year, σ^2 is the variance of the random error term in equation (2) and λ, the fraction of the error attributed to systematic error, is estimated by including the results of the previous election as an explanatory variable in the current one. By calculating this for each election that did not follow a redistricting (i.e. where the electoral map is identical), and taking the mean, we have an estimator for λ.[20]

Generating Hypothetical Future Elections

To predict the properties of a subsequent election using the same districting plan, a series of hypothetical elections are simulated using the estimates for β and σ^2. A new set of explanatory variables X is used to demonstrate the conditions at the election. Since no information can be derived about the nature of the systematic error component beforehand, one error term is used, $\omega = \gamma + \varepsilon$, with variance σ^2. Thus, a single hypothetical election is then generated by drawing from

$$\boldsymbol{\phi}_{hyp} = \mathbf{X}_{hyp}\beta + \delta_{hyp} + \omega \tag{3}$$

where β is the posterior distribution, with mean $\widehat{\beta} = (X'X)^{-1}X'\phi$ and (with a normality assumption) variance $\Sigma_\beta = \sigma^2(X'X)^{-1}$. The δ term is used to produce hypothetical elections whose average district vote is desired to be different from the original. Integrating out the conditional parameters β and γ one obtains the marginal distribution:

$$\boldsymbol{\phi}_{hyp}|\boldsymbol{\phi} \sim N(\lambda\mathbf{v} + (\mathbf{X}_{hyp} - \lambda\mathbf{X})\widehat{\beta} + \delta, (\mathbf{X}_{hyp} - \lambda\mathbf{X})\Sigma_\beta(\mathbf{X}_{hyp} - \lambda\mathbf{X})'^2)\sigma^2 I).$$

To evaluate the election system, let $\mathbf{X}_{hyp} = \mathbf{X}$; to evaluate under counterfactual conditions, set $\mathbf{X}_{hyp}$ to the desired explanatory variables.

Comparing Districting Plans

With the above statistical model in hand, we can predict elections under different partitions of a state into voting districts. The procedure is as follows. First, we estimate the model in equation (2). Second, having generated a new map through our algorithm, we determine the values for the explanatory variables for each district, either by aggregating and averaging the previous values in each precinct or by making sensible predictions for their value (e.g. incumbency). In terms of vote shares, we simply aggregate the VTDs in the new partitions. For incumbency, we assign each incumbent to the latitude and longitude of the centroid of their district. Under the new districting

[20] Ideally, one would have historical votes for many years to tease out the systematic error component. We have only two years of such data.

plan, if there is one such incumbent per district, s/he becomes the incumbent. In the rare cases where there was more than one incumbent assigned to a district under a new districting plan, we break the tie by choosing the incumbent closest to the resulting centroid and replacing another district with the other incumbent to keep the numbers constant. Finally, with our new map we simulate the model 1000 times; deriving the relevant parameters is straightforward.

5.2 Analyzing Seat-Vote Curves

Using the methodology described above, figures 13-16 provide seat-vote curves for California, New York, Pennsylvania, and Texas under each state's actual districting plan and the plan that maximizes its compactness. The vertical axis depicts the proportion of seats won by democrats. The horizontal axis depicts the share of votes that the democrats earned in the election. Each figure reports two interesting quantities: Vote is the average district vote the Democrats received in the election; and Seats report the fraction of seats the Democrats received in the election (not the hypothetical seat share). The dark line represents our estimate of the seat vote curve, the two parallel lines around it are 95% confidence intervals. Visually, one can see that there is a marked difference between the seat-vote curves estimate, from the actual data and those estimated from the partition developed by our algorithm, in California and New York. The slope of the curve is significantly steeper in both these states. Texas and Pennsylvania are also slightly steeper, but the difference is much less dramatic.

To get a better sense of the magnitudes involved, table 2 presents our estimates of Bias and Responsiveness for the actual partition of our four states and those gleaned from the algorithm. We also report the t-statistic on the difference between them. Under maximally compact districting, measures of bias are slightly smaller in all states except Pennsylvania, though none of the differences are statistically significant. In terms of responsiveness, however, there are large and statistically significant differences between the existing partitions and those that are maximally compact. New York, in particular, has a five fold increase; from .482 to 2.51. In other words, under the current partition, a 1% increase in vote share for Democrats results in a .482% increase in seats under the current system. When maximally compact, however, a 1% increase results in a 2.51% increase. The next largest change is California - increasing from 1.086 to 1.731. Pennsylvania and Texas show smaller increases, which are statistically significant at the 10% level.

6 Concluding Remarks

There will be continued debate about the design of districting plans. We have developed a simple but principled measure of compactness. Our measure can be used to compare districting plans across state and time, a feature not found in existing measure, and our algorithm provides a way of approximating the most compact plan. Further, the impact a maximally compact districting plan can have on the responsive of votes is encouraging. These are first steps toward a more scientific understanding of districting plans and their effects.

References

[1] Adams, Bruce. (1977). "A Model State Reapportionment Process: The Continuing Quest for 'Fair and Effective Representation'." *Harvard Journal on Legislation* 14. 825-904.

[2] Attneave, Fred and Malcolm D. Arnout. (1956). "The Quantitative Study of Shape and Pattern Perception." *Psychological Bulliten* 53. 472-451.

[3] Arrow, Kenneth. (1970). *Social Choice and Individual Values.* Cowles Foundation Monograph Series. New Haven, CT. Yale University Press.

[4] Aurenhammer, Franz. (1987). "Power Diagrams: Properties, Algorithms and Applications," *SIAM Journal on Computing* 16. 78-96.

[5] Aurenhammer, Franz., Hoffmann, F., and Aronov, B. (1998). "Minkowski-Type Theorems and Least-Squares Clustering," *Algorithmica* 20. 61-76.

[6] Barabas, Jason and Jennifer Jerit. (2004). "Redistricting Principles and Racial Representation," *State Politics and Policy Quarterly* 4. 415-435.

[7] Bartal, Yair, Moses Charikar and Danny Raz. (2001). "Approximating min-sum k-clustering in Metric Spaces,"

[8] Boyce, Ronald R. and W.A.V. Clark. (1964). "The Concept of Shape in Geography," *Geographical Review* 54. 561-572.

[9] Cox, E.P. (1927). "A Method of Assigning Numerical and Percentage Values to the Degree of Roundness of Sand Grains." *Journal of Paleontology* 1. 179-183.

[10] Diaconescu, D.E., B. Florea, and A. Grassi. (2002). "Geometric transitions, del Pezzo surfaces and open string instantons," Advanced Theorertical and Mathematical Physics. 6 . 643–702.

[11] Eig, Larry M. and Michael V. Seitzinger. (1981). "State Constitutional and Statutory Provisions Concerning Congressional and State Legislative Redistricting." Congressional Research Service Report No. 81-143A.

[12] Geisler, R. Gene. 1985. "Measuring Gerrymandering of Legislative Districts." Unpublished Manuscript.

[13] Gelman, Andrew and Gary King. (1990). "Estimating the Electoral Consequences of Legislative Redistricting," *Journal of the American Statistical Association* 85. 274-282.

[14] Gelman, Andrew and Gary King. (1994). "A Unified Method of Evaluating Electoral Systems and Redistricting Plans," *American Journal of Political Science* 38(2). 514-554.

[15] Gibbs, J.P. (1961). "A Method for Comparing the Spatial Shape of Urban Units." in *Urban Research Methods* (J.P.Gibbs, ed.) New York, NY: Van Nostrand.

[16] Harris, Curtis C., Jr. (1964). "A Scientific Method of Districting." *Behavioral Science* 9. 219-225.

[17] Hofeller, Thomas and Bernard Groffman. (1990). "Comparing the Compactness of California Congressional Districts under Three Different Plans: 1980, 1982, and 1984." in *Toward Fair and Effective Representation* (Bernard Groffman, ed.) New York, NY: Agathon.

[18] Horton, R.E. (1932). "Drainage Basin Characteristics." *Transactions of the American Geophysical Union* 13. 350-361.

[19] Issacharoff, Samuel, Pamela S. Karlan and Richard H. Pildes (2002). *The Law of Democracy: Legal Structure of the Political Process* (2nd ed.) New York, NY: Foundation Press.

[20] Kaiser, Henry F. (1966). "An Objective Method for Establishing Legislative Districts." *Midwest Journal of Political Science* 10. 200-213.

[21] Knight, Jeffrey L. (2004). "GIS BASED COMPACTNESS MEASUREMENT USING FRACTAL ANALYSIS." http://www.spatial.maine.edu/ucgis/testproc/knight/knight.html.

[22] MacQueen, J.B. (1967). "Some Methods For Classification and Analysis of Multivariate Observations," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley, CA. University of California Press. 281-297.

[23] Mandelbrot, Benoit B. (1967). "How long is the cost of Great Britain? Statistical self-similarity and fractal dimension." *Science* 156. 636-638.

[24] Mitchell, John E. (2001). "Branch-and-cut for the k-way equipartition problem," Technical report, Mathematical Sciences, Rensselaer Polytechnic Institute.

[25] Mitchell, John E. (2003). "Realignment in the National Football League: Did They Do It Right?" *Naval Research Logistics* 50. 683-701.

[26] Mumford, David and Jayant Shah. (1989). *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems.* Harvard University Press. Cambridge, MA.

[27] Niemi, Richard G., Bernard Groffman, Carl Calucci and Thomas Hofeller. (1990). "Measuring Compactness and the Role of Compactness Standard in a Test for Partisan and Racial Gerrymandering". *The Journal of Politics* 52. 1155-1181.

[28] Papayanopolous, L. (1973). "Quantitative Principles Underlying Legislative Apportionment." *Annals of the New York Academy of Sciences* 19. 181-191.

[29] Passare, Mikael and Hans Rullgard. (2004). "Amoebas, Monge-Ampere measures, and triangulations of the Newton polytope," Duke Mathematics Journal 121. 481–507.

[30] Pounds, Norman J.G. (1972). *Political Geography* (2nd ed.) New York, NY: McGraw-Hill, Inc.

[31] Pildes, Richard H. and Richard G. Niemi. (1993). "Expressive Harms, 'Bizarre Districts,' and Voting Rights: Evaluating Election-District Appearances After Shaw v. Reno". 92 Michigan Law Review. 483-531.

[32] Richter-Gebert, Jurgen, Bernd Sturmfels, and Thorsten Theobald. (2003) "First steps in tropical geometry", arXiV.math. 1–29.

[33] Rudin, Walter. (1964). *Principles of Mathematical Analysis.* New York, NY: McGraw Hill, Inc.

[34] Sahni, S. and T. Gonzales. (1976). "P-Complete Approximation Problems." *Journal of the ACM* 23. 555-566.

[35] Schwartberg, Joseph E. (1966). "Reapportionment, Gerrymanders, and the Notion of Compactness." *Minnesota Law Review* 50. 443-452.

[36] Taylor, Peter J. (1973). "A New Shape Measure for Evaluating Electoral District Patterns." *American Political Science Review* 67. 947-950.

[37] Weaver, James B. and Sidney W. Hess. (1963). "A Procedure for Nonpartisan Districting: Development of Computer Techniques." *Yale Law Journal* 72. 288-308.

[38] Wells, David I. (1982). "How to Inhibit Gerrymandering: The Purpose of Districting Criteria." *National Civic Review* 71. 183-187.

[39] Young, H.P. (1988). "Measuring the Compactness of Legislative Districts." *Legislative Studies Quarterly* 13. 105-115.

7 Appendix A: Technical Proofs

7.1 Proof of Theorem 1

Proof of Theorem 1, Part 1:

That the RPI satisfies the three axioms follows from five simple lemmas which we now state and prove.

Lemma 1 *The Relative Proximity Index satisfies Anonymity.*

Proof. Consider a partition V of state S and an associated compactness index $c(V,S)$. Now consider a bijection $h : S \to S$.

$$\sum_{v \in V_S} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2$$

is unchanged since h is a bijection and hence there are the same number of point in each element of V and they are at the same points. For identical reasons the denominator of the RPI does not change, and hence $c(V,S) = c_h(V,S)$ for any bijection h. ■

Lemma 2 *The Relative Proximity Index satisfies Clustering.*

Proof. Let there be two partitions, V_S^1 and $V_{S'}^2$ such that

$$\sum_{v \in V_S^1} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 > \sum_{v \in V_{S'}^2} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 \tag{4}$$

Clustering requires:

$$c(V_S^1, S) > c(V_S^2, S)$$

Suppose, by way of contradiction, that (4) holds, and

$$c(V_1, S) < c(V_2, S). \tag{5}$$

That is

$$\frac{\sum_{v \in V_S^1} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2}{\min_{V \in \mathcal{V}_S} \sum_{v \in V} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2} < \frac{\sum_{v \in V_S^2} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2}{\min_{V \in \mathcal{V}_S} \sum_{v \in V} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2} \tag{6}$$

The denominators are identical and hence the supposition requires:

$$\sum_{v \in V_S^1} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 < \sum_{v \in V_{S'}^2} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2, \tag{7}$$

a contradiction. ■

Lemma 3 *The Relative Proximity Index satisfies Density Independence.*

Proof. Consider S and $\widehat{S}$, with $|S|$ and $\left|\widehat{S}\right|$ respectively with $\widehat{S}$ a λ-replica of S. We need to show that $RPI\,(V,S) = RPI(V,\widehat{S})$ for all $V \in \mathcal{V}_S, V \in \mathcal{V}_{\hat{S}}$. That is

$$\frac{\sum_{v\in V_S}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\min_{V\in\mathcal{V}_S}\sum_{v\in V}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2} = \frac{\sum_{v\in V_{\hat{S}}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\min_{V\in\mathcal{V}_{\hat{S}}}\sum_{v\in V}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2},$$

for all $V \in \mathcal{V}_S, V \in \mathcal{V}_{\hat{S}}$. By the definition of a λ-replica, the right-hand side of the above equation is simply

$$\frac{\lambda\sum_{v\in V_S}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\lambda\min_{V\in\mathcal{V}_S}\sum_{v\in V}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2},$$

which is clearly equal to the left-hand side for any partition. ■

Lemma 4 *The Relative Proximity Index satisfies Scale independence.*

Proof. Scale Independence requires that for two states, S and $\widehat{S}$ with $d_{jk} = \lambda d_{jk}$, for all $j,k \in S,\widehat{S}$. Then $c(V,S) = c(V,\widehat{S})$, for all $V \in \mathcal{V}_S, V \in \mathcal{V}_{\hat{S}}$. That is

$$\frac{\sum_{v\in V_S}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\min_{V\in\mathcal{V}_S}\sum_{v\in V}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2} = \frac{\sum_{v\in V_{\hat{S}}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\min_{V\in\mathcal{V}_S}\sum_{v\in V}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2},$$

for all $V \in \mathcal{V}_S, V \in \mathcal{V}_{\hat{S}}$. Scale independence means that the right-hand side of the above equation is simply

$$\frac{\sum_{v\in V_S}\sum_{i\in v}\sum_{j\in v}(\lambda d_{ij})^2}{\min_{V\in\mathcal{V}_S}\sum_{v\in V}\sum_{i\in v}\sum_{j\in v}(\lambda d_{ij})^2} = \frac{\lambda^2\sum_{v\in V_S}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\lambda^2\min_{V\in\mathcal{V}_S}\sum_{v\in V}\sum_{i\in v}\sum_{j\in v}d_{ij}^2}.$$

which is clearly equal to the left-hand side for any partition. ■

Lemma 5 *The Relative Proximity Index satisfies Number of Districts independence.*

Proof. Follows immediately from the definition of independence with respect to number of districts. ■

We can now prove the second part of Theorem 1. It is proved by transforming a given state so that it can be compared to another state. Anonymity and Independence ensure that this can be done in a way which does not alter the compactness index, and Clustering then allows a comparison of two districting plans to be made based on their total intra-cluster pairwise distances.

Proof of Theorem 1, Part 2.

Proof. From part 1 we have $RPI\,(V,S_m) > RPI\left(\hat{V},S_n\right) \Rightarrow c\,(V,S_m) > c\left(\hat{V},S_n\right)$, for any m,n. Suppose part 2 is not true. This implies that

$$c\,(V,S_m) > c\left(\hat{V},S_n\right) \text{ and } RPI\,(V,S_m) < RPI\left(\hat{V},S_n\right), \tag{8}$$

or

$$c\,(V,S_m) < c\left(\hat{V},S_n\right) \text{ and } RPI\,(V,S_m) > RPI\left(\hat{V},S_n\right),$$

for some m, n.

If $S_m = S_n$ then the argument is straightforward. Begin with the first pair of inequalities. Note that Equality implies that $\mu_{ij} = \mu$ for all i, j and that symmetry of g implies combined with Equality implies that g is additively separable in its arguments. Then by Equality and Clustering we have

$$\sum_{v \in V_{S_m}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 > \sum_{v \in \hat{V}_{S_n}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 \Longrightarrow c(V, S_m) > c\left(\hat{V}, S_n\right),$$

since $RPI(V, S_m) < RPI\left(\hat{V}, S_n\right)$ and

$$S_m = S_n \Rightarrow \min_{V \in \mathcal{V}_{S_m}} \sum_{v \in V_{S_m}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 = \min_{V \in \mathcal{V}_{S_n}} \sum_{v \in \hat{V}_{S_n}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2,$$

we have

$$\sum_{v \in V_{S_m}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 < \sum_{v \in \hat{V}_{S_n}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2.$$

By Clustering this implies that $c(V, S_m) < c\left(\hat{V}, S_n\right)$–a contradiction. Identical reasoning rules out the case where

$$c(V, S_m) < c\left(\hat{V}, S_n\right) \text{ and } RPI(V, S_m) > RPI\left(\hat{V}, S_n\right).$$

Now consider the case in which $S_m \neq S_n$, and suppose that S_m contains γ_m districts and S_n contains γ_n districts. Consider the following transformation of state n. First, make a $\lambda-$replica of S_n and a $\mu-$replica of S_m so that the number of voters is the same as in state the transformed S_m. Note that $c(V, S_m)$ and $RPI(V, S_m)$ are unchanged due to Independence. In a slight abuse of notation we will continue to use V and S_m in reference to the $\mu-$replicated state. Second, expand or contract the state in the sense that the distance between any two points, d_{ij} say, in state S_n is αd_{ij} in state $S_{n'}$. Note that any partition of state n is a well defined partition of state $S_{n'}$ as it contains the same voters, scaled by α. Choose α such that

$$\alpha = \frac{|n| \min_{V \in \mathcal{V}_{S_n}^{\gamma_m}} \sum_{v \in \hat{V}_{S_n}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2}{\mu |m| \min_{V \in \mathcal{V}_{S_m}} \sum_{v \in V_{S_m}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2},$$

where $|n|$ and $|m|$ are the number of voters in states S_n and S_m respectively, and the γ_m superscript denotes a partition into γ_m elements. Note that

$$\min_{V \in \mathcal{V}_{S_m}} \sum_{v \in V_{S_m}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 = \min_{V \in \mathcal{V}_{S_{n'}}^{\gamma_m}} \sum_{v \in V_{S_{n'}}} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2. \tag{9}$$

Third, select a feasible partition of $S_{n'}$ with γ_m elements, and denote this partition $\hat{V}'$. Suppose

$$\sum_{v\in\hat{V}'_{S_{n'}}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2=\theta\sum_{v\in\hat{V}_{S_n}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2,$$

and that

$$\min_{V\in\mathcal{V}_{S_n}^{\gamma_m}}\sum_{v\in\hat{V}_{S_n}}\sum_{i\in v}\sum_{j\in v}f(d_{ij})=\beta\min_{V\in\mathcal{V}_{S_n}^{\gamma_n}}\sum_{v\in\hat{V}_{S_n}}\sum_{i\in v}\sum_{j\in v}f(d_{ij}).$$

Hence

$$\frac{\sum_{v\in\hat{V}'_{S_{n'}}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\min_{V\in\mathcal{V}_{S_n}^{\gamma_m}}\sum_{v\in\hat{V}_{S_n}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}=\frac{\theta}{\beta}\frac{\sum_{v\in\hat{V}_{S_n}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}{\min_{V\in\mathcal{V}_{S_n}^{\gamma_n}}\sum_{v\in\hat{V}_{S_n}}\sum_{i\in v}\sum_{j\in v}(d_{ij})^2}$$

By Independence

$$c\left(\hat{V}',S_{n'}\right)=\frac{\theta}{\beta}c\left(\hat{V},S_n\right)$$

and

$$RPI\left(\hat{V}',S_{n'}\right)=\frac{\theta}{\beta}RPI\left(\hat{V},S_n\right).$$

From (8)

$$c(V,S_m)>\frac{\beta}{\theta}c\left(\hat{V}',S_{n'}\right)\text{ and }RPI(V,S_m)<\frac{\beta}{\theta}RPI\left(\hat{V}',S_{n'}\right).\tag{10}$$

But since S_m and $S_{n'}$ have the same number of voters, the same number of districts, and (9) holds, it follows that (10) implies that c violates Clustering.

Identical reasoning rules out the case where

$$c(V,S_m)<c\left(\hat{V},S_n\right)\text{ and }RPI(V,S_m)>RPI\left(\hat{V},S_n\right),$$

and hence the proof is complete. ■

7.2 Proof of Theorem 2

Let districts of state S be denoted $D_1,\ldots,D_d$. A districting plan is *feasible* if $|D_i|=n$ for all $i\in\{1,\ldots,d\}$. The set of feasible districtings is $\mathcal{V}$. Let the centroid of district D_i be m_i, so $m_i=\frac{1}{n}\sum_{x\in D_i}(x)$. Define the functions:

$$\psi(D_i)=\sum_{x\in D_i}\|x-m_i\|^2,\qquad\Psi(D_1,\ldots,D_d)=\sum_{i=1}^{d}\psi(D_i)$$

We say that districting is optimally compact if it minimizes $\Psi(D_1, \ldots, D_d)$ over all $(D_1, \ldots, D_d) \in \mathcal{V}$. For $z_1, \ldots, z_d \in \mathbb{R}^2$, let:

$$\psi_{z_i}(D_i) = \sum_{x \in D_i} \|x - z_i\|^2, \qquad \Psi_{z_1,\ldots,z_d}(D_i) = \sum_{i=1}^{d} \psi_{z_i}(D_i)$$

A Power Diagram with sites $z_1, \ldots, z_d$ is a partition of $\mathbb{R}^2$ into districts $D_1, \ldots, D_d$ such that for fixed constants $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$,

$$D_i = \left\{ q \in \mathbb{R}^2 : i = \arg\min_j \left[\|q - z_j\|^2 - \lambda_j \right] \right\}$$

It is clear that a power diagram is described by its edges and the fact that if x is on the same side as D_i of any complete set of linear separators between D_i and other districts then $x \in D_i$, and otherwise not. The edges of D_i are described by the set of $q \in \mathbb{R}^2$ such that $\|q - z_i\|^2 - \lambda_i = \|q - z_i\|^2 - \lambda_j$, or $\|q - z_i\|^2 - \|q - z_i\|^2 = \lambda_i - \lambda_j$.

Lemma 6 $\Psi(D_1, \ldots, D_d)$ *is proportional to the RPI for* $(D_1, \ldots, D_d) \in \mathcal{V}$*, so minimizing one is equivalent to minimizing the other. Specifically,*

$$\sum_{i=1}^{d} \sum_{x \in D_i} \sum_{y \in D_i} \|x - y\|^2 = 2n \sum_{i=1}^{d} \sum_{x \in D_i} \|x - m_i\|^2 .$$

Proof of Lemma 6.

$$\begin{aligned}
\sum_{i=1}^{d}\sum_{x\in D_i}\sum_{y\in D_i}\|x-y\|^2 &= \sum_{i=1}^{d}\sum_{x\in D_i}\sum_{y\in D_i}\left(\|x\|^2+\|y\|^2-2x\cdot y\right)\\
&= \sum_{i=1}^{d}\sum_{x\in D_i}\left(n\|x\|^2-2nm_i\cdot x+\sum_{y\in D_i}\|y\|^2\right)\\
&= \sum_{i=1}^{d}\left(\sum_{x\in D_i}\left(n\|x\|^2-2nm_i\cdot x\right)+n\sum_{y\in D_i}\|y\|^2\right)\\
&= \sum_{i=1}^{d}\left(\sum_{x\in D_i}\left(2n\|x\|^2-2nm_i\cdot x\right)\right)\\
&= \sum_{i=1}^{d}\left(2n\sum_{x\in D_i}\left(\|x\|^2-m_i\cdot x\right)\right)\\
&= \sum_{i=1}^{d}2n\left(\sum_{x\in D_i}\left(\|x\|^2\right)-n\|m_i\|^2\right)\\
&= \sum_{i=1}^{d}\left(2n\left(\sum_{x\in D_i}\left(\|x\|^2-2m_i\cdot x+\|m_i\|^2\right)\right)\right)\\
&= \sum_{i=1}^{d}\left(2n\left(\sum_{x\in D_i}\|x-m_i\|^2\right)\right)\\
&= 2n\sum_{i=1}^{d}\sum_{x\in D_i}\|x-m_i\|^2
\end{aligned}$$

■

Lemma 7 *For all* $(D_1,\ldots,D_d)\in\mathcal{V}$,

$$(m_1,\ldots,m_d)=\arg\min_{(z_1,\ldots,z_d)}\Psi_{z_1,\ldots,z_d}(D_1,\ldots,D_d)$$

Proof of Lemma 7. It suffices to show that substituting m_i for z_i minimizes the expression on the right. Its first order condition with respect to the z_i is:

$$\forall D_i,\qquad 2\sum_{x\in D_i}(x-z_i)=0\qquad\Rightarrow\qquad z_i=\frac{1}{n}\sum_{x\in D_i}x=m_i$$

■

Lemma 8 *In an optimally compact districting, every pair of adjacent districts is separated by a line perpendicular to a line connecting their centroids.*

Proof of Lemma 8. Let $(D_1, \ldots, D_d)$ be optimally compact. Without loss of generality we can prove the lemma for districts D_1 and D_2. By isometry we can assume that $m_1 = (0,0)$ and $m_2 = (\xi, 0)$. Pick $v_1 = (x_1, y_1) \in D_1$ and $v_2 = (x_2, y_2) \in D_2$. Let $D_1' = D_1 \cup \{v_2\} - \{v_1\}$ and $D_2' = D_2 \cup \{v_1\} - \{v_2\}$. By the optimality of $(D_1, \ldots, D_d)$ and the optimality lemma,

$$\psi(D_1) + \psi(D_2) \leq \psi(D_1') + \psi(D_2') \leq \psi_{m_1}(D_1') + \psi_{m_2}(D_2')$$

$$\Rightarrow \qquad \|v_1 - m_1\|^2 + \|v_2 - m_2\|^2 \leq \|v_1 - m_2\|^2 + \|v_2 - m_1\|^2$$

$$\Rightarrow \qquad -2v_1 \cdot m_1 - 2v_2 \cdot m_2 \leq -2v_1 \cdot m_2 - 2v_2 \cdot m_1$$

$$\Rightarrow \qquad (v_2 - v_1) \cdot (m_1 - m_2) \leq 0$$

$$\Rightarrow \qquad (x_2 - x_1) \cdot (-\xi) + (y_2 - y_1) \cdot (0) \leq 0$$

$$\Rightarrow \qquad x_1 \leq x_2$$

Since v_1 and v_2 are arbitrary, we can pick them such that v_1 is the point in D_1 with greatest x_1 and v_2 is the point in D_2 with least x_2, showing that there is a line of the form $x = c$ for $c \in \mathbb{R}$ separating the two districts. Isometrics preserve perpendicularity, so applying one moving m_1 and m_2 away from $(0,0)$ and $(\xi, 0)$ leaves the separator between D_1 and D_2 perpendicular to the segment connecting m_1 and m_2. ■

Lemma 9 *Let $(D_1, \ldots, D_d)$ be optimal. For every three districts, there exist three concurrent lines each of which separates two of the three districts, with one line separating each pair of districts.*

Proof of Lemma 9. Without loss of generality we prove this for the three districts D_1, D_2, and D_3. By the Straight Line lemma, there exist linear separators between D_1 and D_2, D_2 and D_3, and D_3 and D_1 perpendicular to the lines connecting their centroids. We can characterize these lines by the equations $\|r - m_1\|^2 - \|r - m_2\|^2 = \mu_{1,2}$, $\|s - m_2\|^2 - \|s - m_3\|^2 = \mu_{2,3}$, and $\|t - m_3\|^2 - \|t - m_1\|^2 = \mu_{3,1}$, for free variables $r, s, t \in \mathbb{R}^2$. If the lines are concurrent, that means there exist $q \in \mathbb{R}^2$ satisfying all three equations. Adding them together gives $\mu_{1,2} + \mu_{2,3} + \mu_{3,1} = 0$. Therefore, if the lines are concurrent then for all r, s, and t on the lines,

$$\|r - m_1\|^2 - \|r - m_2\|^2 + \|s - m_2\|^2 - \|s - m_3\|^2 + \|t - m_3\|^2 - \|t - m_1\|^2 = 0$$

Assume there is no choice for $\mu_{1,2}$, $\mu_{2,3}$, and $\mu_{3,1}$ such that the lines are concurrent. Then, for all r, s, and t on the three edges,

$$\|r - m_1\|^2 - \|r - m_2\|^2 + \|s - m_2\|^2 - \|s - m_3\|^2 + \|t - m_3\|^2 - \|t - m_1\|^2 \neq 0$$

If any one of $\mu_{1,2}$, $\mu_{2,3}$, or $\mu_{3,1}$ induces an optimal separator at both the values ν_1 and ν_2 in $\mathbb{R}^2$, then it must also at the value $\lambda\nu_1 + (1 - \lambda)\nu_2$ for $\lambda \in [0, 1]$. So the expression above is either strictly greater or strictly less than 0 for all permissible values of r, s, and t. We assume without loss of generality that it is greater. Then, there exist $v_1 \in D_1$, $v_2 \in D_2$, and $v_3 \in D_3$ such that when substituted for r, s, and t, respectively, the above expression reaches a positive infimum. The expression cannot be at an infimum unless the extreme values of r, s, and t are specifically chosen to be in D_1, D_2, and D_3, respectively, otherwise $\|r - m_1\|^2 - \|r - m_2\|^2$, for example, could be decreased by moving r in the direction $m_1 - m_2$ while still separating D_1 and D_2. Therefore,

$$\|v_1 - m_1\|^2 - \|v_1 - m_2\|^2 + \|v_2 - m_2\|^2 - \|v_2 - m_3\|^2 + \|v_3 - m_3\|^2 - \|v_3 - m_1\|^2 > 0$$

$$\Leftrightarrow \|v_1 - m_1\|^2 + \|v_2 - m_2\|^2 + \|v_3 - m_3\|^2 > \|v_1 - m_2\|^2 + \|v_2 - m_3\|^2 + \|v_3 - m_1\|^2$$

Let $D_1' = D_1 \cup \{v_3\} - \{v_1\}$, $D_2' = D_2 \cup \{v_1\} - \{v_2\}$, and $D_3' = D_3 \cup \{v_2\} - \{v_3\}$. Then,

$$\psi(D_1) + \psi(D_2) + \psi(D_3) > \psi_{m_1}(D_1') + \psi_{m_2}(D_2') + \psi_{m_3}(D_3') > \psi(D_1') + \psi(D_2') + \psi(D_3')$$

This contradicts the optimality of $D_1, \ldots, D_d$, and the lemma follows.

■

Proof of Theorem 2. We prove that any optimal districting is a power diagram with cites equal to their centroids, $m_1, \ldots, m_d$. For any pair of districts D_i and D_j, we can pick $\mu_{i,j}$ such that $\|q - m_i\|^2 - \|q - m_j\|^2 = \mu_{i,j}$ is a linear separator between the districts, and if we add a third district D_j, we can similarly pick $\mu_{j,k}$ and $\mu_{k,i}$ such that the districting lines are concurrent, or $\mu_{i,j} + \mu_{j,k} + \mu_{k,i} = 0$. Note that $\mu_{a,b} = -\mu_{b,a}$. We prove that there exist constants $\lambda_1, \ldots, \lambda_d$ such that $\lambda_i - \lambda_j = \mu_{i,j}$ by induction. This is obviously true when $n = 2$. Assume it is true for districts $D_1, \ldots, D_k$. For $i, j < k + 1$,

$$\mu_{i,k+1} = \mu_{i,j} + \mu_{j,k+1} = \lambda_i - \lambda_j + \mu_{j,k+1}$$

$$\Rightarrow \lambda_i - \mu_{i,k+1} = \lambda_j - \mu_{j,k+1}$$

Thus, $\lambda_i - \mu_{i,k+1}$ is constant over choice of i, call the constant λ_{k+1}. That makes $\mu_{i,k+1} = \lambda_i - \lambda_{k+1}$ for any i, and the induction is complete. Clearly any $x \in D_i$ is on the m_i side of a boundary line between D_i and another district, so it follows that optimal districtings are power diagrams. ■

8 Appendix B: A Guide to Programs

All programs to compute feasible districtings minimizing the RPI are written for MATLAB. There are two main programs, Main.m and Compute_Index.m, and support programs District.m, getRandGP.m, Psi.m, Weighted_Assign.m, Weighted_FirstTryAssign.m, and Weighted_PowerDiagram.m. We briefly describe each below.

Main.m and Compute_Index.m are both shell programs which call District.m, the actual algorithm, and store its output in text files. Typing Compute_Index(filename, Iterations) reads demographic data about a state from a text file, say 'indiana.out', and creates a new districting Iterations times. The file should have the latitudes and longitudes of the census tracts of the states in columns two and three (respectively), the FIPS code of the state repeated in every entry of column four, the current districts of all census tracts in column five, and the populations of all census tracts in column six. Compute_Index.m generates two output files. The first, in this case 'indiana.out.output' contains the latitudes and longitudes of the census tracts in the first two columns, and their new district numbers in the subsequent columns. Each column after the second represents a different iteration of the algorithm. The second output file, in this case 'indiana.out.stats', contains statistics from each iteration of the algorithm on a different row. The first column has the RPI's, the second has the accuracy of the districting, and the third has the accuracy of the current districting. Accuracy is measured:

$$\max_{i\in\{1,\ldots,d\}}\left|\frac{|D_i|-n}{n}\right|$$

Compute_Index.m has the following hard-coded parameters which are passed to District.m: outside_tol_ratio, tol_ratio, outside_bail, and bail. tol_ratio and bail are the stopping criteria for the sub-routine Weighted_Assign.m which creates the best districting around randomly-initiated sites. If the accuracy falls below tol_ratio or the number of iterations of the gradient-descent procedure rises above bail, the algorithm terminates. Likewise, outside_tol_ratio and outside_bail are the stopping criteria for the larger districting algorithm. If the accuracy of the districting falls below outside_tol_ratio or the number of times the sites are moved rises above outside_bail, the algorithm terminates. The set values for outside_tol_ratio, tol_ratio, outside_bail, and bail are .9 times the real accuracy, whichever is the lesser between .9 times the real accuracy or .05, 35 times the number of districts in the state, and 35 times the number of districts in the state.

Main(filename) reads a list of states and iterations for each state to be run by Compute_Index. The file is of the form:

states, bootstraps
alabama 4
arizona 7
arkansas 3
california 1

Names of states and numbers of iterations are separates by tabs. If 'arizona' is written in this file, Compute_Index will open a file called 'arizona.out'. Main.m creates an additional file called index.txt which lists the FIPS code for every state next to the best RPI the algorithm has found for it such that the accuracy for the districting corresponding to that RPI is better than the state's current accuracy.

This procedure yields an RPI > 1 and an accuracy better than the current accuracy nearly all of the time for all states other than Connecticut, Idaho, Minnesota, and Nebraska, which already are well-districted and usually require quite a few bootstraps to improve on the current districting.

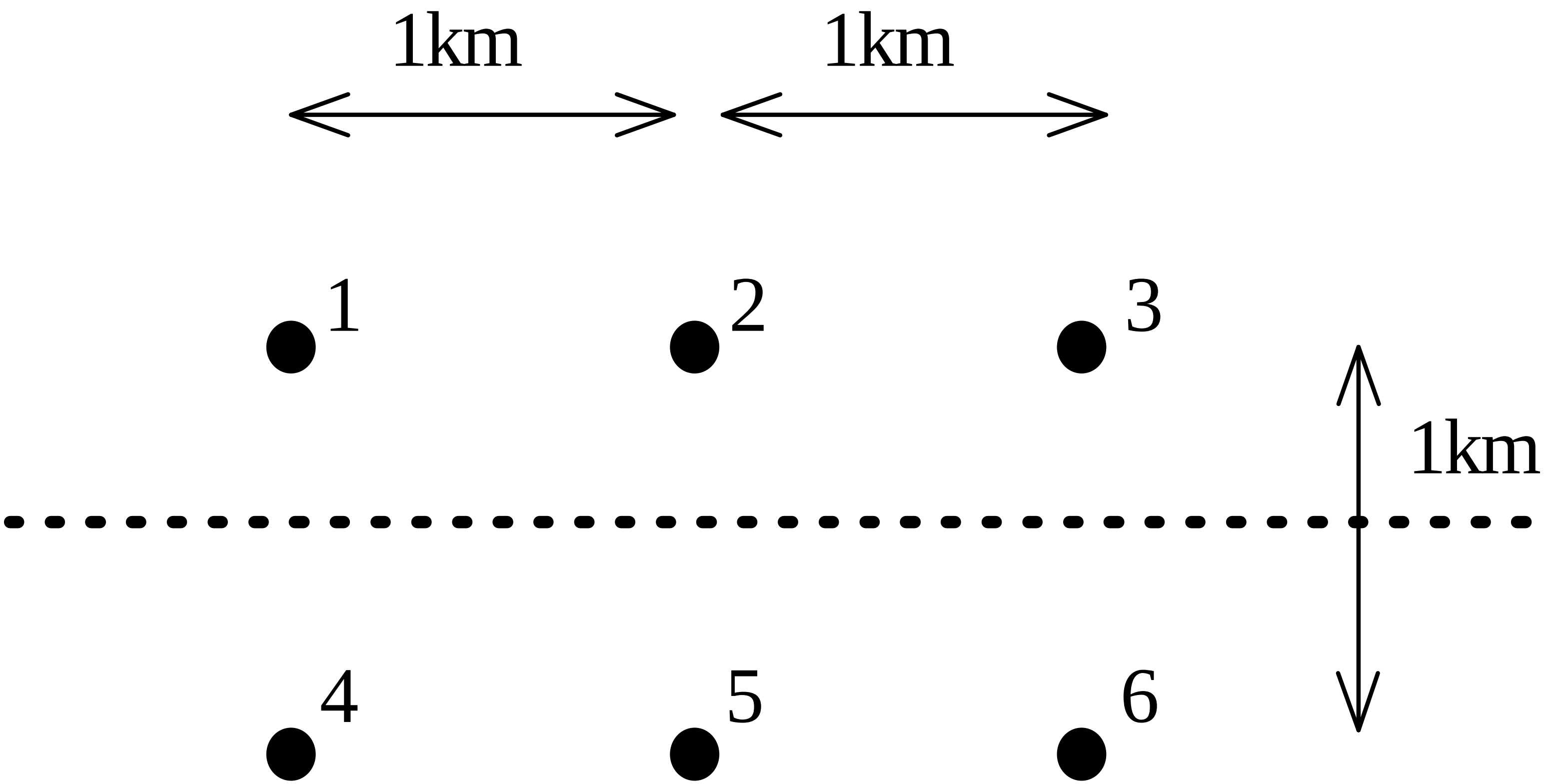

Figure 1: A Simple Example

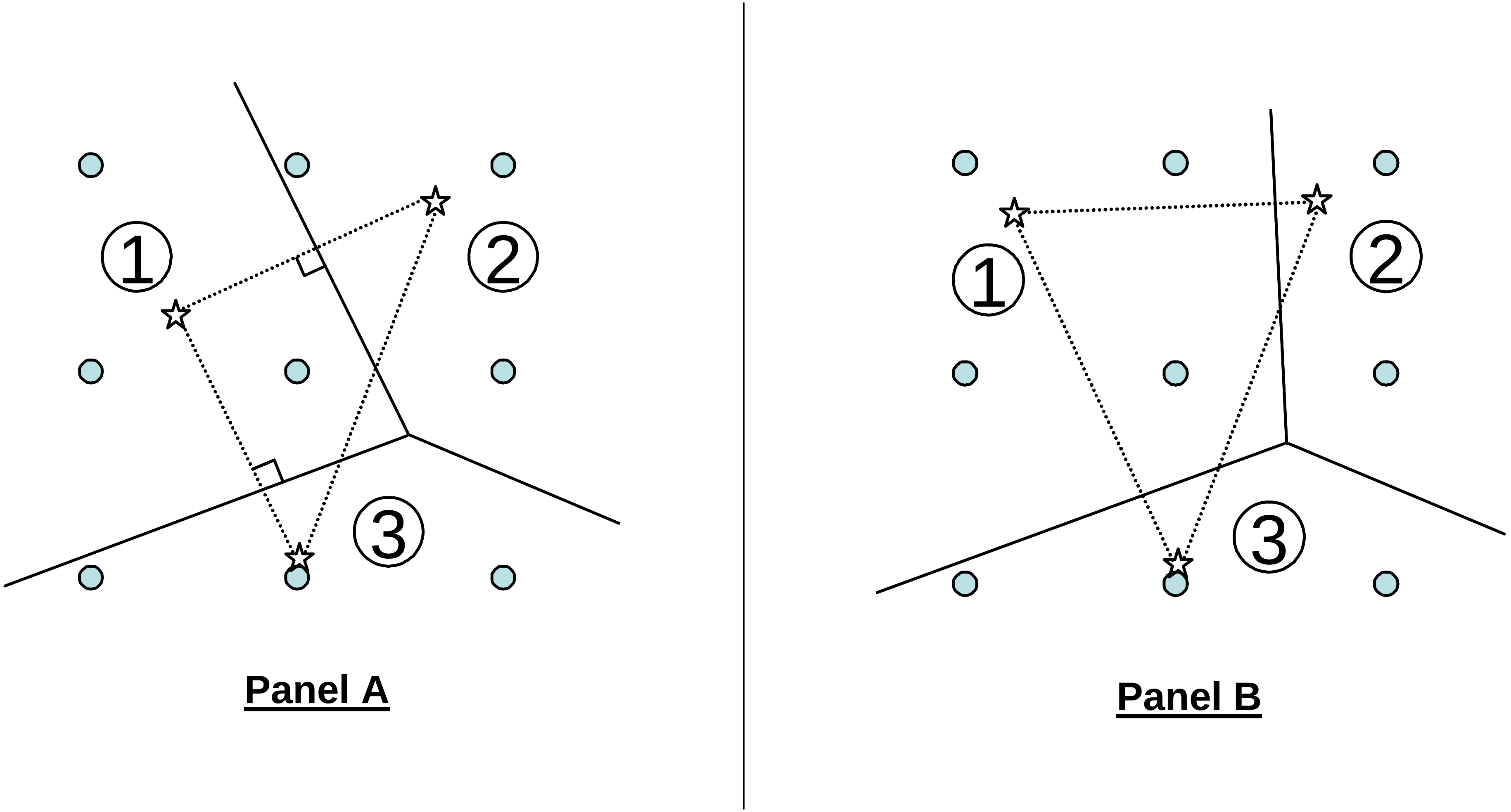

Figure 2: Good and Bad Generator Points

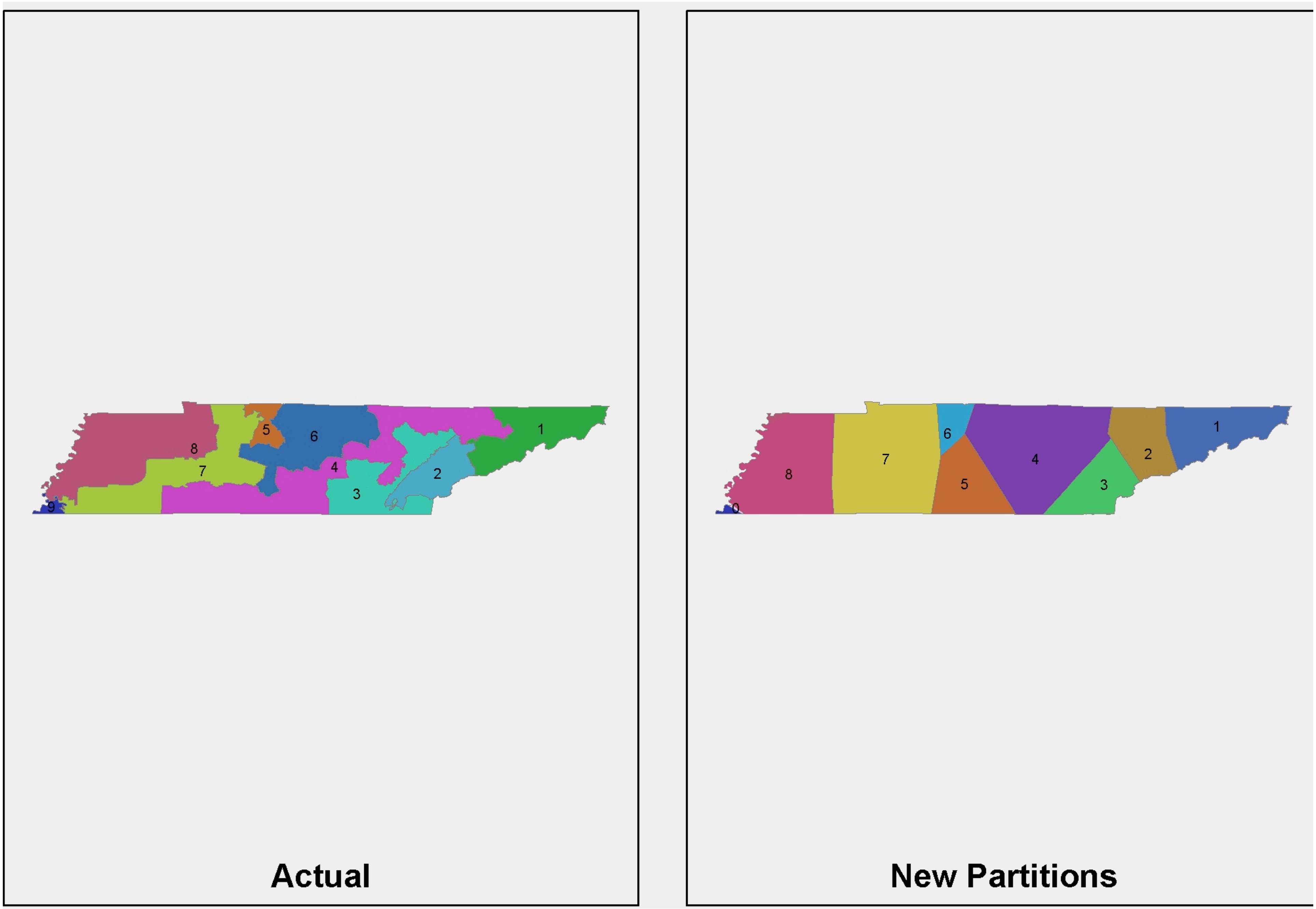

Figure 3: Tennessee 106th Congress Districting Plans, Actual v. Algorithm

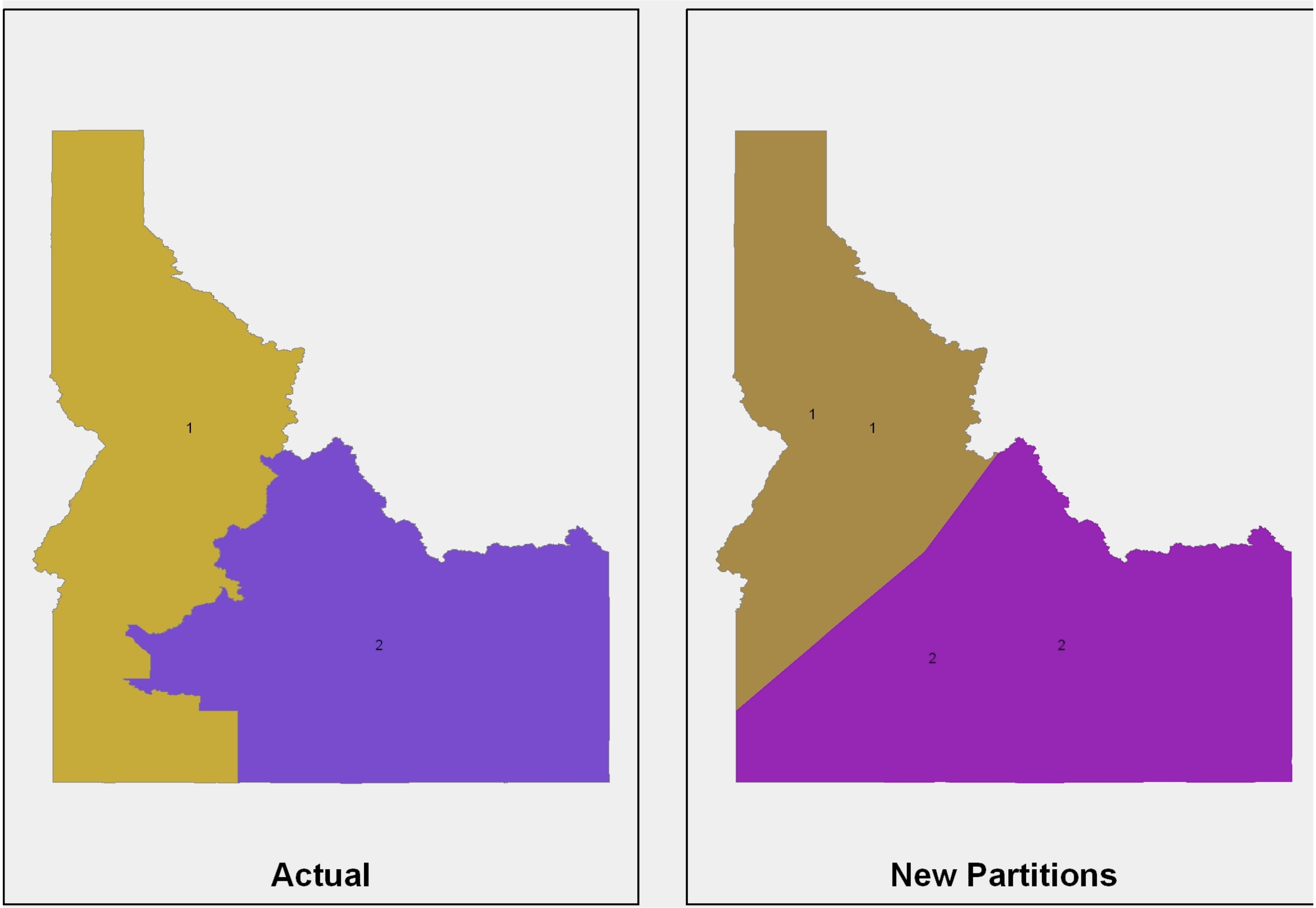

Figure 4: Idaho 106th Congress Districting Plans, Actual v. Algorithm

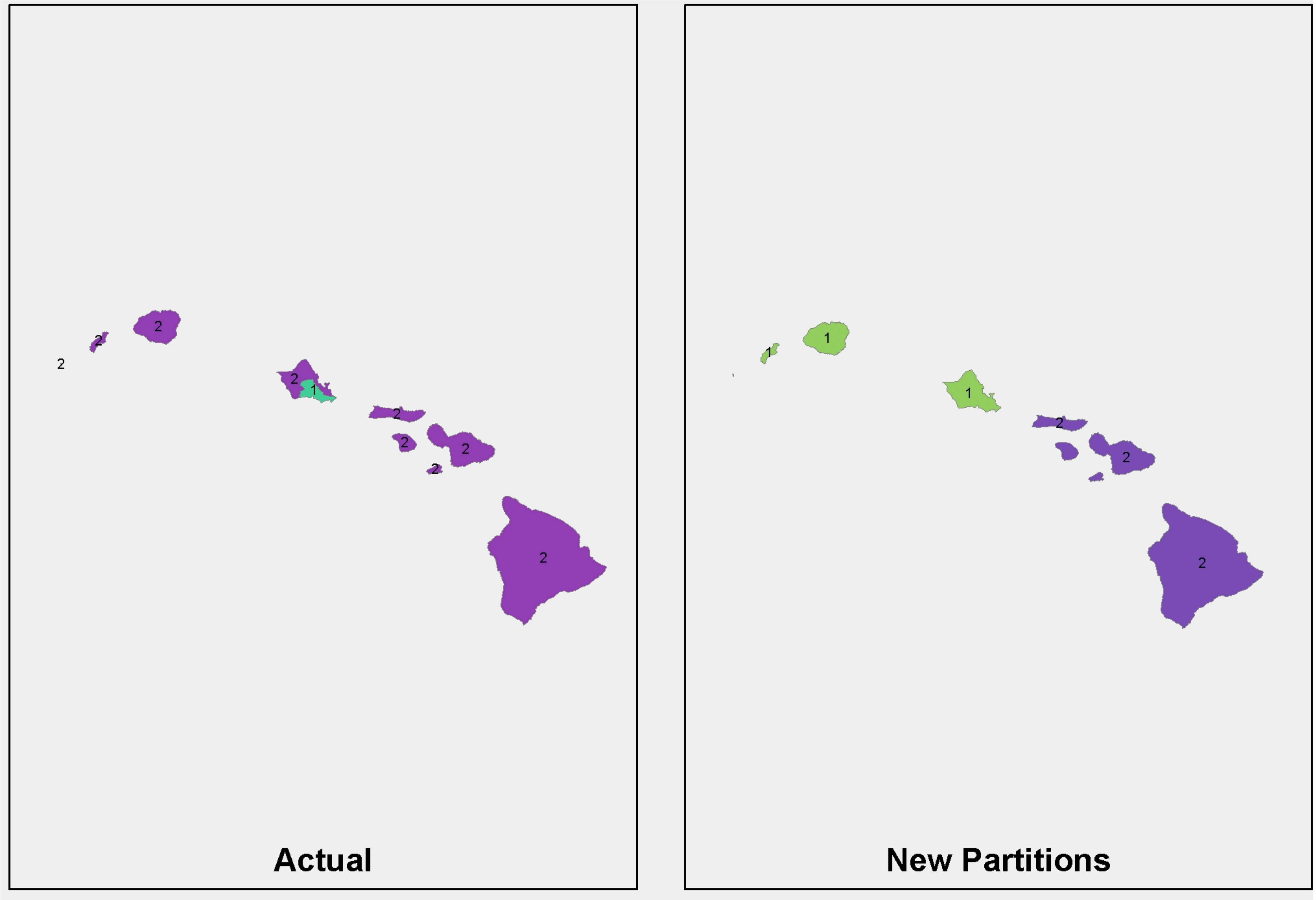

Figure 5: Hawaii 106th Congress Districting Plans, Actual v. Algorithm

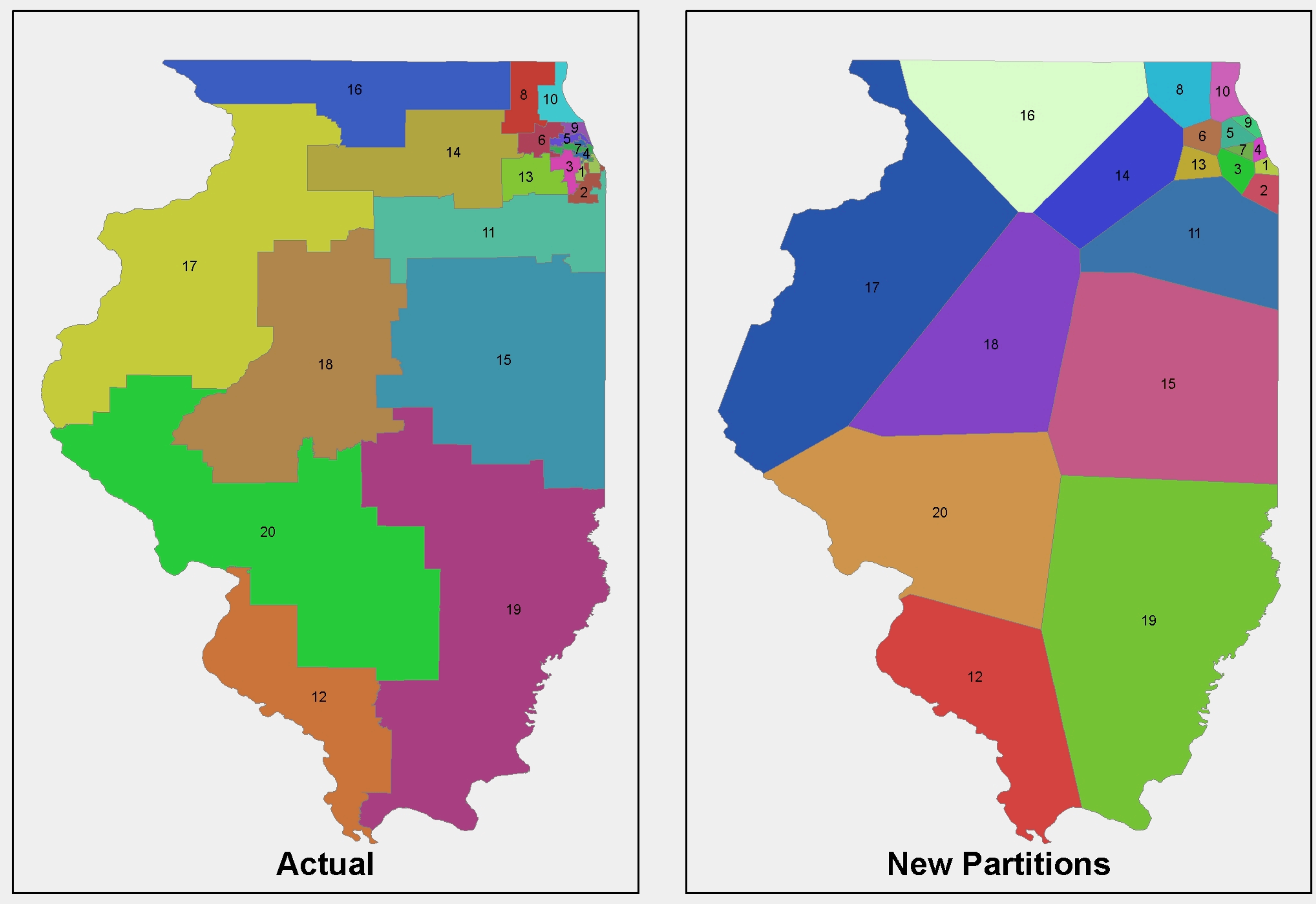

Figure 6: Illinois 106th Congress Districting Plans, Actual v. Algorithm

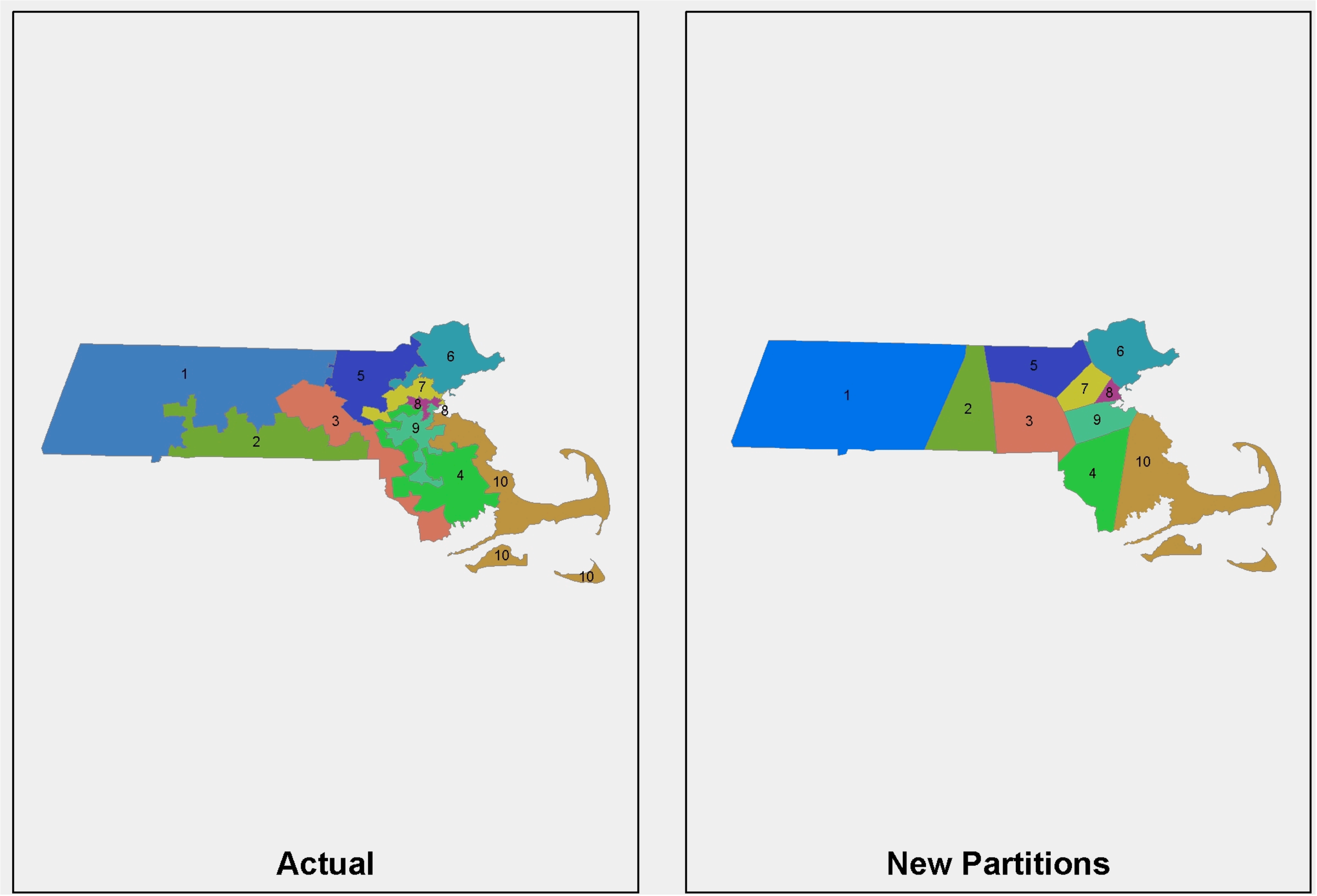

Figure 7: Massachusetts 106th Congress Districting Plans, Actual v. Algorithm

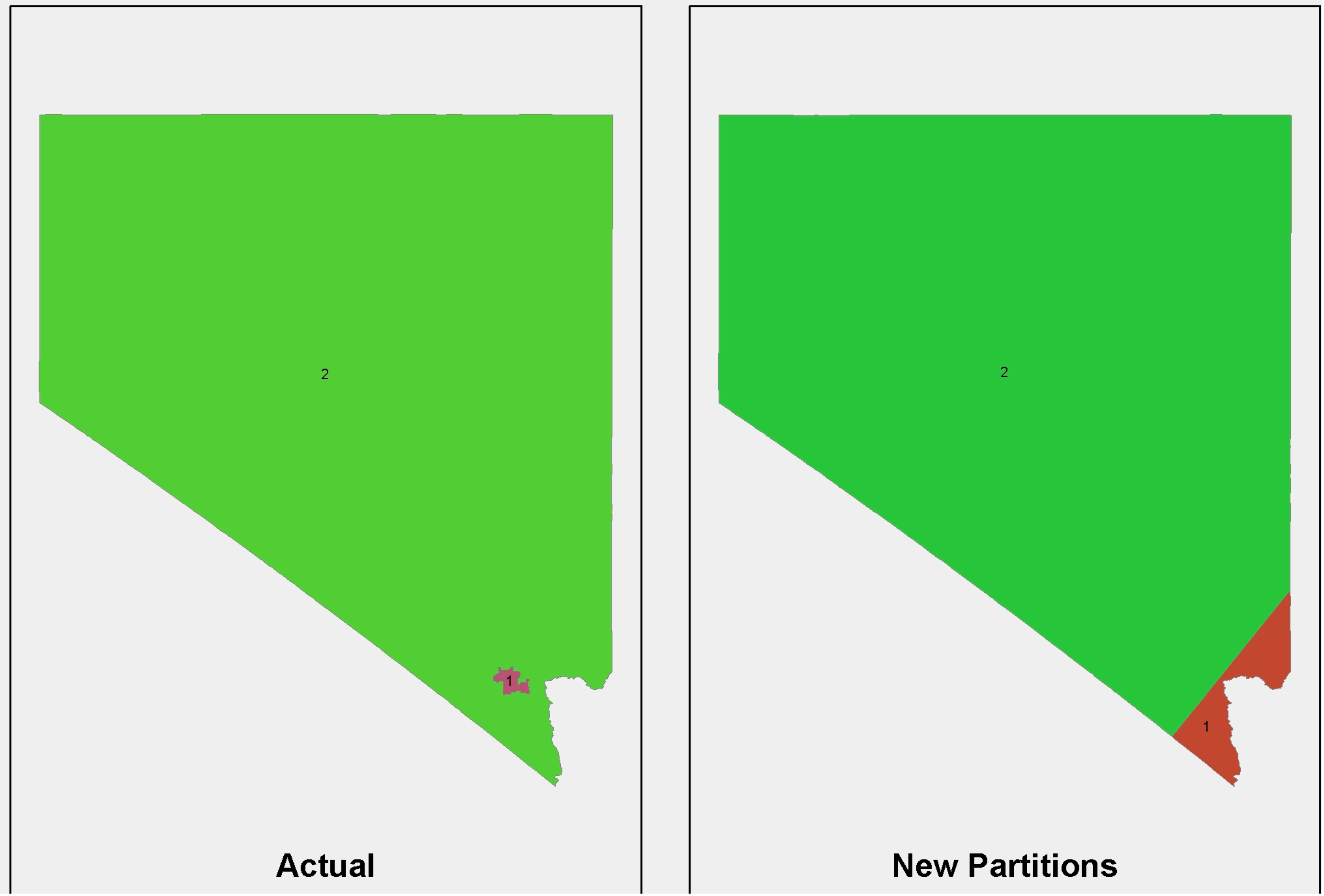

Figure 8: Nevada 106th Congress Districting Plans, Actual v. Algorithm

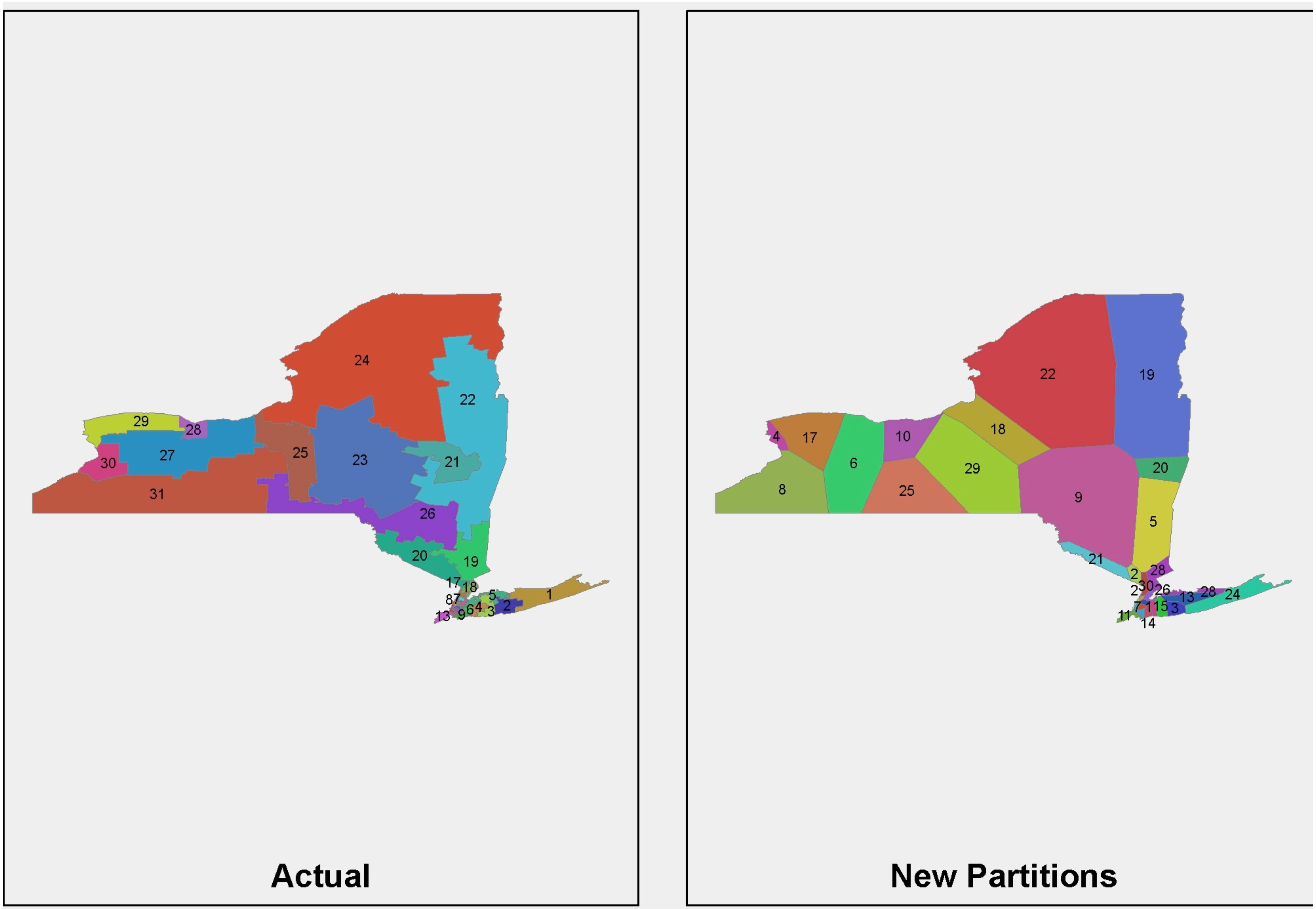

Figure 9: New York 106th Congress Districting Plans, Actual v. Algorithm

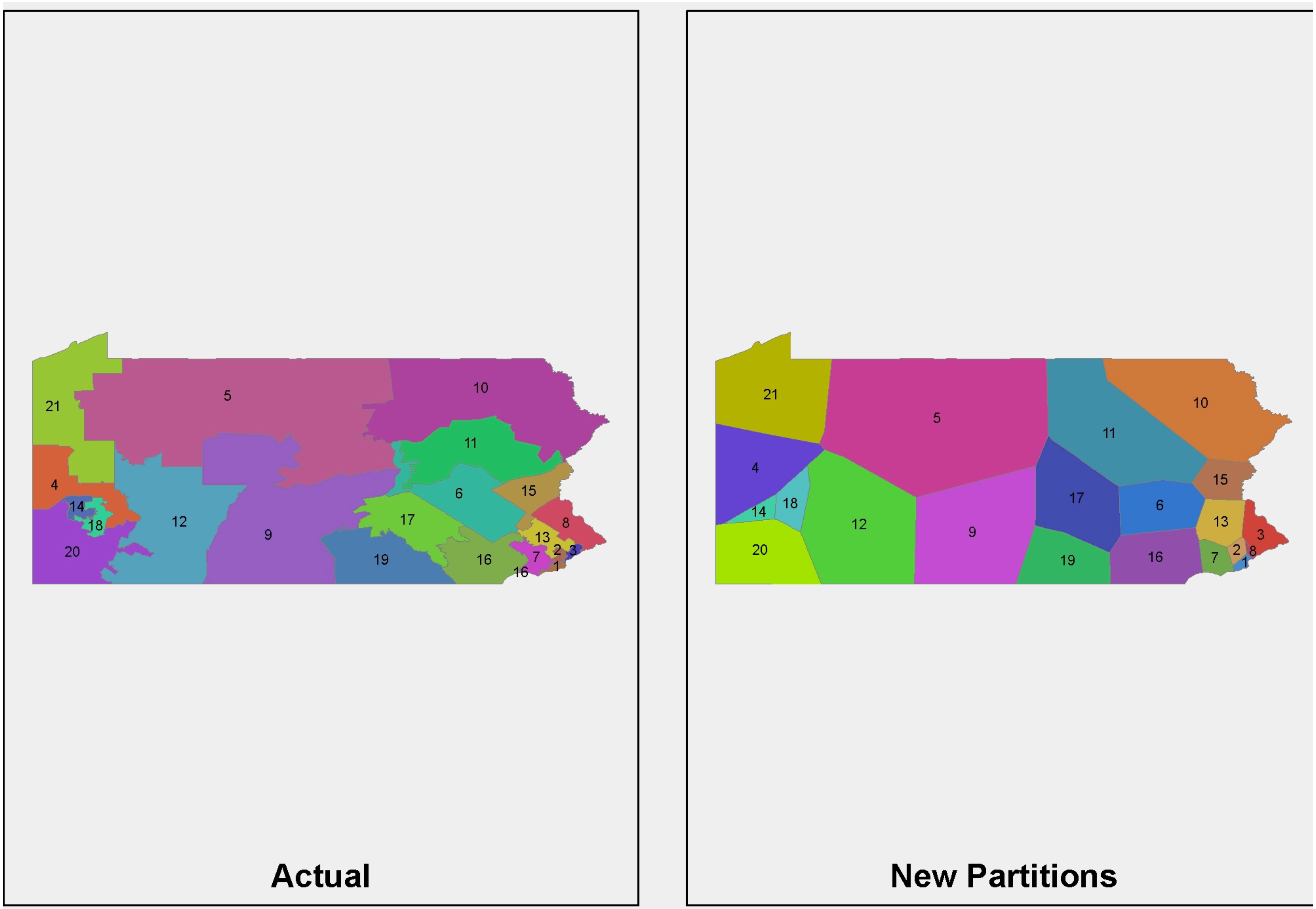

Figure 10: Pennslyvania 106th Congress Districting Plans, Actual v. Algorithm

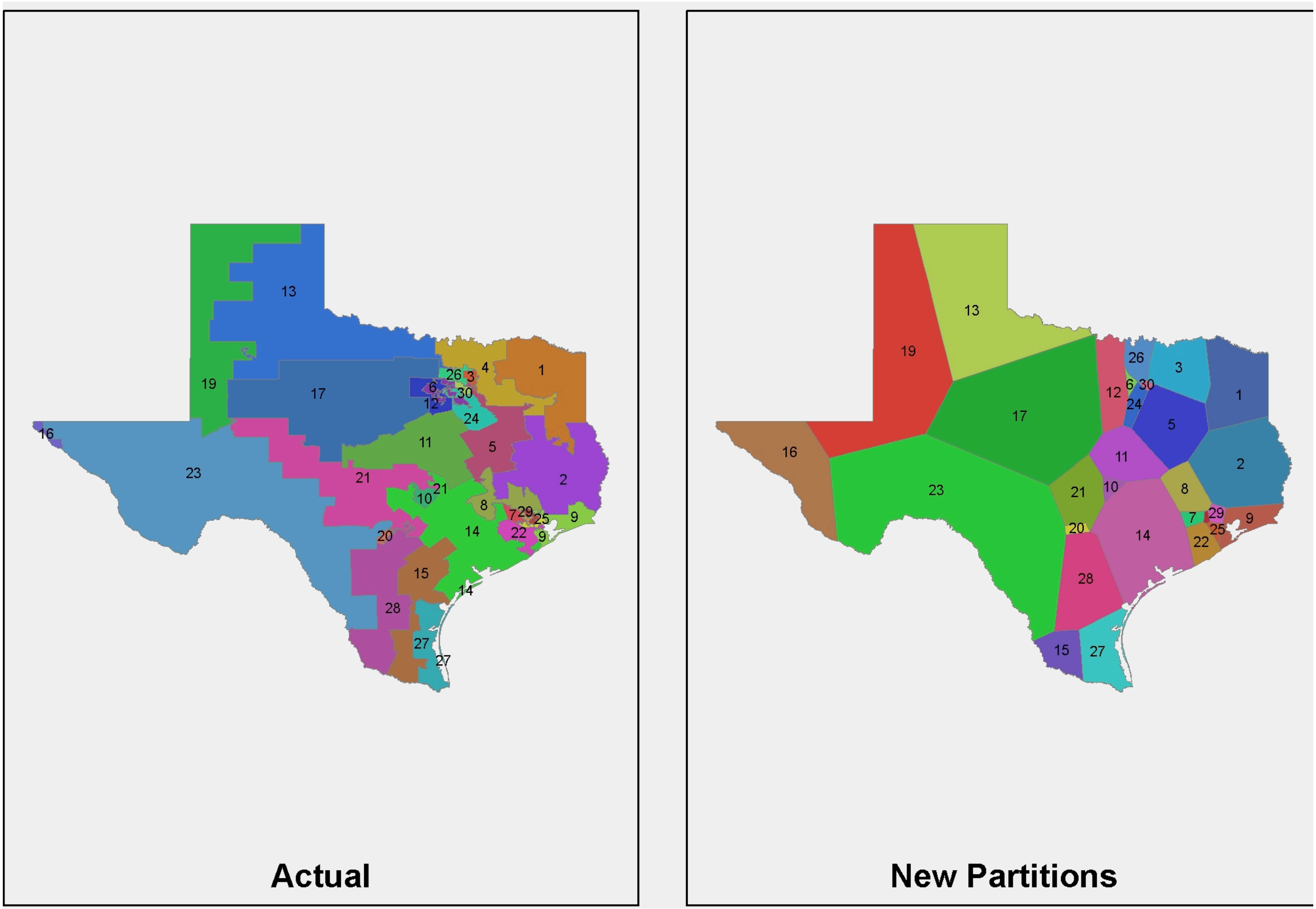

Figure11: Texas 106th Congress Districting Plans, Actual v. Algorithm

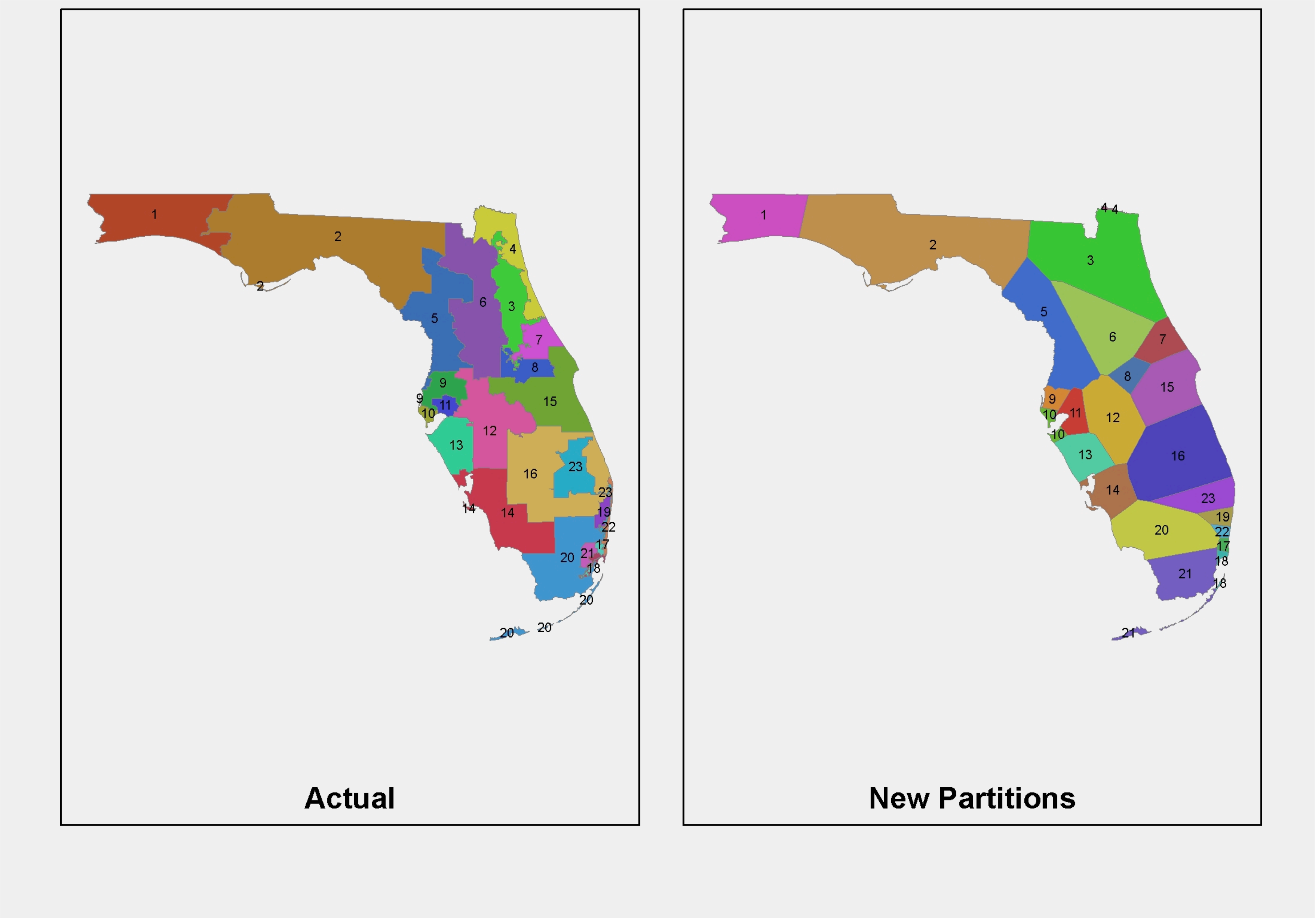

Figure 12: Florida 106th Congress Districting Plans, Actual v. Algorithm

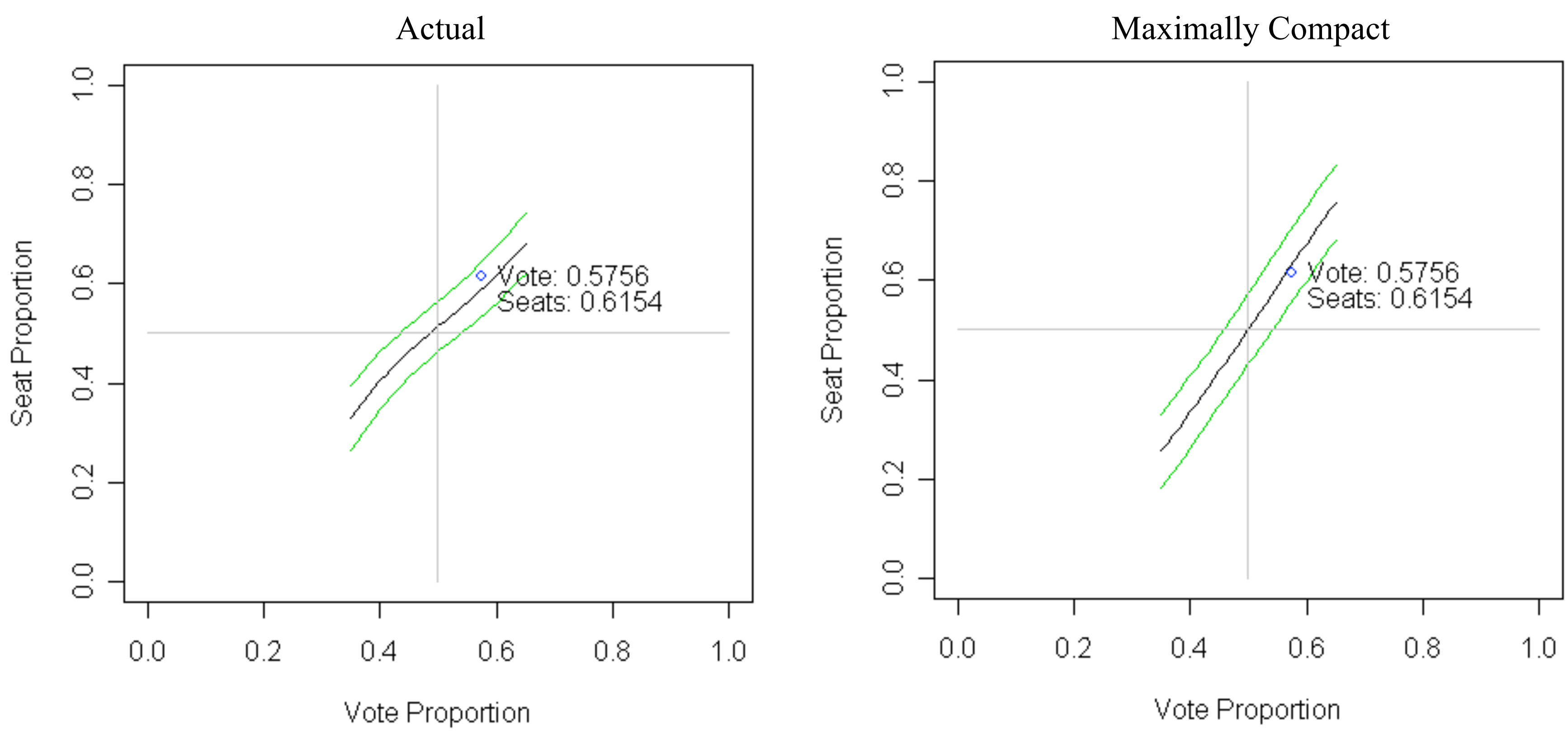

Figure 13: Seat-Vote Curves for California, Actual v. Maximally Compact

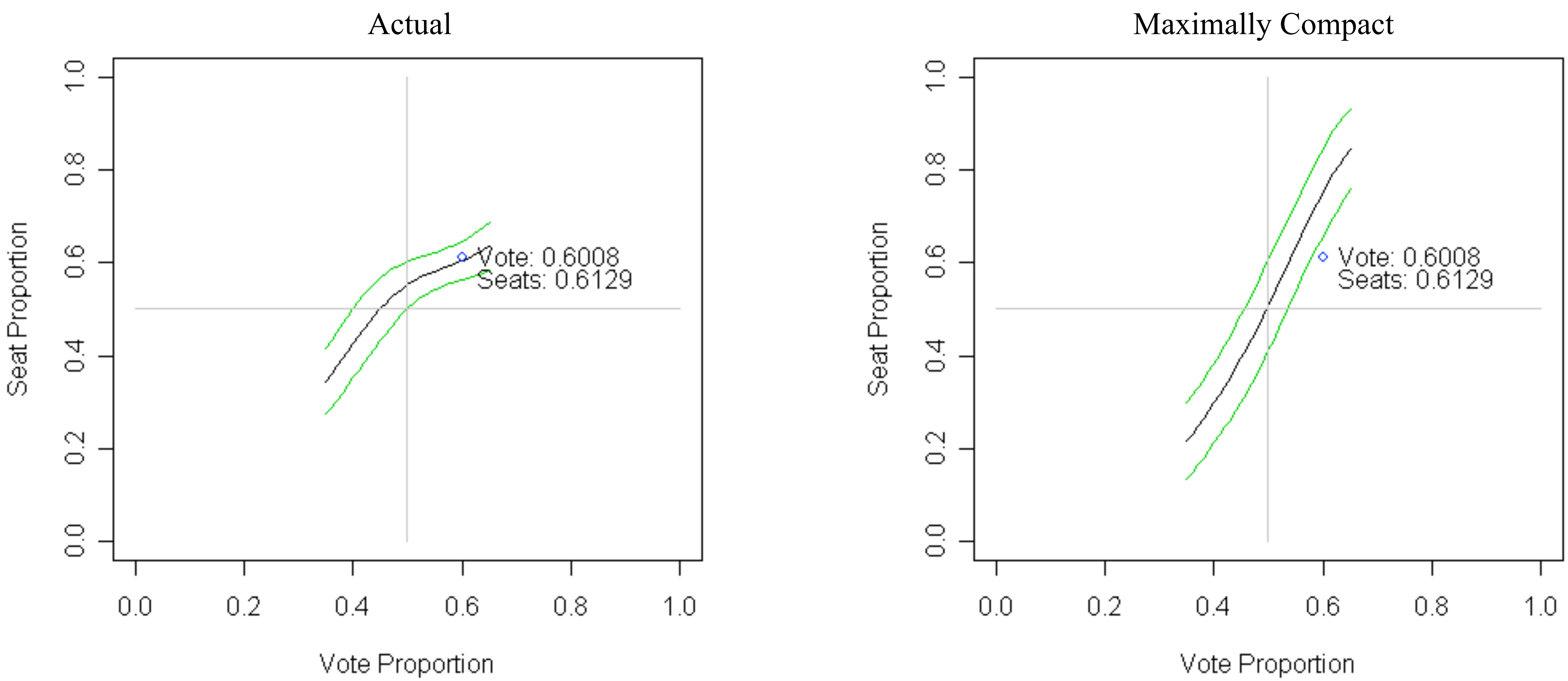

Figure 14: Seat-Vote Curves for New York, Actual v. Maximally Compact

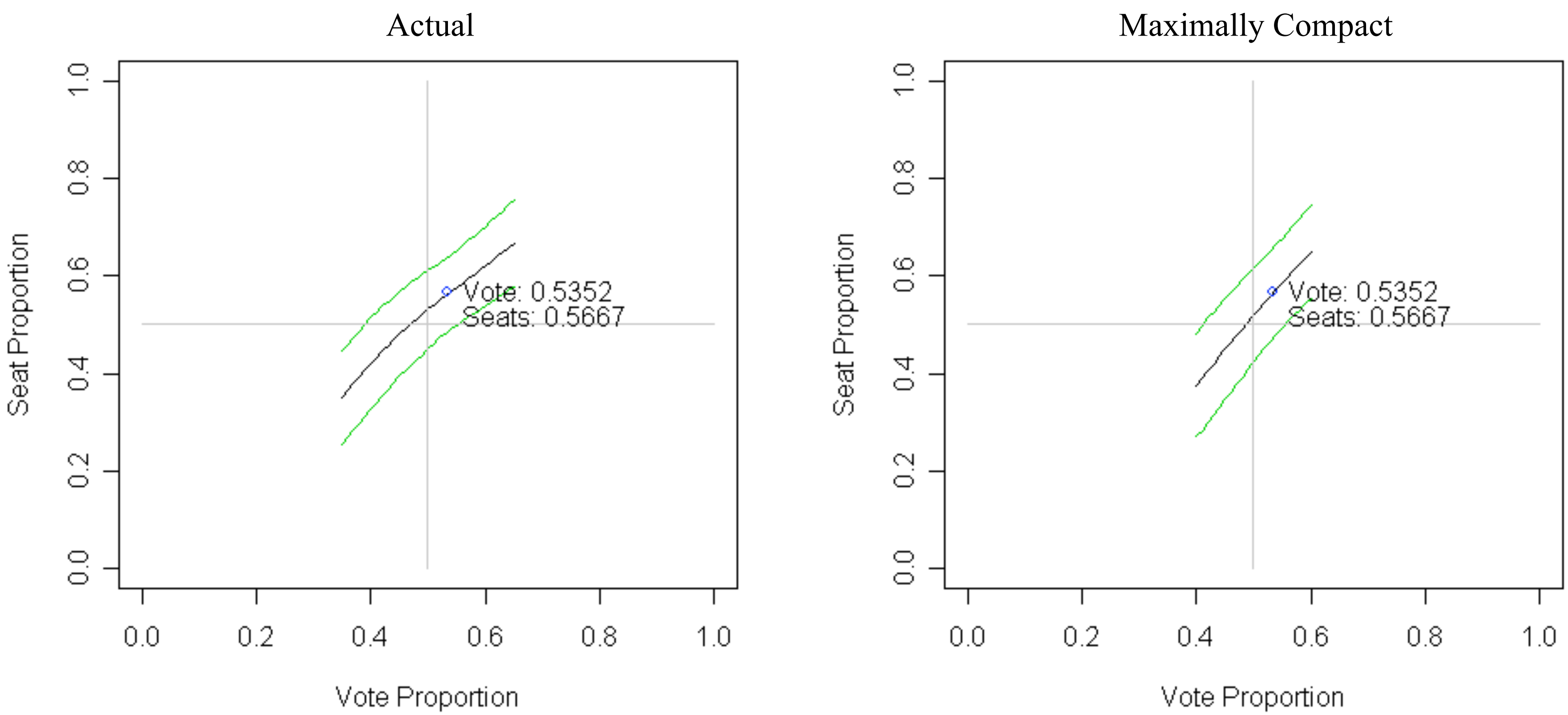

Figure 15: Seat-Vote Curves for Texas, Actual v. Maximally Compact

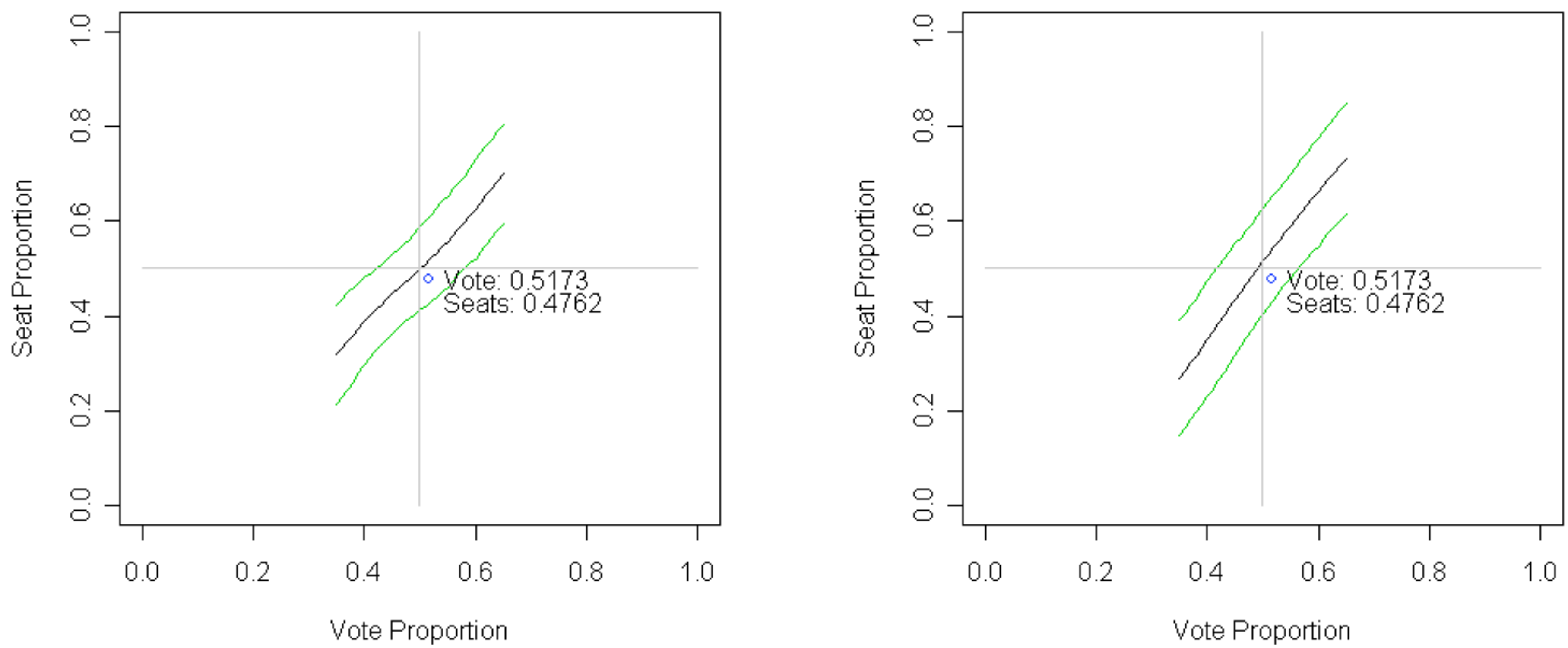

Figure 16: Seat-Vote Curves for Pennsylvania, Actual v. Maximally Compact

Table 1: The Relative Proximity Index, 2000

State Name	RPI	Max Deviation (Actual)	Max Deviation (Algorithm)	Mean RPI	Standard Deviation RPI	Percentile
Alabama	1.21	0.27	0.05	0.99	0.03	1.00
Arizona	1.34	0.20	0.15	1.27	0.04	0.97
Arkansas	1.08	0.14	0.05	0.78	0.01	1.00
California	1.49	0.17	0.04	0.96	0.03	1.00
Colorado	1.59	0.15	0.05	1.28	0.02	1.00
Connecticut	1.36	0.02	0.01	1.09	0.35	0.78
Florida	1.39	0.46	0.07	0.83	0.08	1.00
Georgia	1.24	0.14	0.09	0.90	0.01	1.00
Hawaii	1.59	0.09	0.04	1.48	0.02	1.00
Idaho	0.97	0.10	0.02	0.80	0.02	1.00
Illinois	1.43	0.29	0.11	0.98	0.07	1.00
Indiana	1.49	0.20	0.06	1.05	0.02	1.00
Iowa	1.38	0.06	0.05	1.29	0.01	1.00
Kansas	1.11	0.08	0.05	0.95	0.01	1.00
Kentucky	1.51	0.14	0.05	1.22	0.01	1.00
Louisiana	1.15	0.13	0.05	0.79	0.43	0.80
Maine	1.39	0.04	0.03	1.15	0.01	1.00
Maryland	1.52	0.22	0.04	1.25	0.02	1.00
Masschussetts	1.87	0.10	0.05	1.54	0.01	1.00
Michigan	1.24	0.13	0.04	0.99	0.02	1.00
Minnesota	1.05	0.16	0.05	0.90	0.02	1.00
Mississippi	1.02	0.18	0.05	0.87	0.01	1.00
Missouri	1.38	0.23	0.05	1.01	0.16	0.99
Nebraska	1.01	0.05	0.04	0.89	0.23	0.70
Nevada	1.38	0.08	0.05	1.19	0.01	1.00
New Hampshire	1.10	0.01	0.00	1.09	0.00	0.95
New Jersey	2.27	0.21	0.05	1.69	0.02	1.00
New Mexico	1.23	0.06	0.04	1.14	0.01	1.00
New York	1.83	0.21	0.10	1.45	0.45	0.80
North Carolina	1.33	0.28	0.04	1.15	0.09	0.97
Ohio	1.62	0.13	0.05	1.42	0.01	1.00
Oklahoma	1.24	0.09	0.05	1.42	0.36	0.31
Oregon	1.26	0.09	0.04	1.21	0.28	0.56
Pennsylvania	1.81	0.25	0.22	1.27	0.05	1.00
Rhode Island	1.18	0.03	0.02	1.18	0.01	0.55
South Carolina	1.22	0.21	0.04	1.27	0.02	0.00
Tennessee	2.91	0.25	0.04	2.59	0.04	1.00
Texas	1.90	0.30	0.22	1.24	0.07	1.00
Utah	1.46	0.06	0.04	1.40	0.01	1.00
Virginia	1.38	0.22	0.07	1.14	0.04	1.00
Washington	1.17	0.15	0.06	0.77	0.03	1.00
West Virginia	1.68	0.06	0.05	1.61	0.01	1.00
Wisconsin	1.40	0.11	0.08	1.22	0.58	0.62

Notes: RPI values were calculated using tract-level data from the 2000 Census. Max Deviation 1 minus the total population of the largest congressional district divided by the total population of the smallest congressional district. Mean RPI was calculated as the mean of 200 repititions of the RPI -- each having different starting values.

Table 2: Partisan Bias and Responsiveness, Actual versus Maximally Compact Districtings

State	Bias (Actual)	Bias (Algorithm)	t-statistic on Difference	Responsiveness (Actual)	Responsiveness (Algorithm)	t-statistic on Difference
California	.028	.007	.469	1.086	1.731	-4.327**
	(.010)	(.045)		(.069)	(.132)	
New York	.103	.018	1.051	0.482	2.51	-6.540**
	(.014)	(.080)		(.036)	(.308)	
Pennsylvania	-0.0027	.031	-.363	1.138	1.562	-1.800*
	(.021)	(.076)		(.128)	(.198)	
Texas	.062	.039	.334	0.8872	1.305	-1.717*
	(.024)	(.064)		(.103)	(.221)	

Notes: Estimates are based on voter tabulation district level election return data for the 105th and 106th congress.