The University of Chicago
The Booth School of Business of the University of Chicago
The University of Chicago Law School

Measuring the Compactness of Political Districting Plans
Author(s): Roland G. Fryer Jr., and Richard Holden
Reviewed work(s):
Source: *Journal of Law and Economics*, Vol. 54, No. 3 (August 2011), pp. 493-535
Published by: The University of Chicago Press for The Booth School of Business of the University of Chicago and The University of Chicago Law School
Stable URL: http://www.jstor.org/stable/10.1086/661511
Accessed: 17/05/2012 03:46

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The University of Chicago Press, The University of Chicago, The Booth School of Business of the University of Chicago, The University of Chicago Law School are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Law and Economics.*

http://www.jstor.org

Measuring the Compactness of Political Districting Plans

Roland G. Fryer, Jr. *Harvard University*
Richard Holden *University of New South Wales*

Abstract

We develop 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. Any compactness measure that 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, which requires us to solve for each state's maximal compactness—a problem that is nondeterministic polynomial-time hard (NP hard). The correlations between our index and the commonly used measures of dispersion and perimeter are −.37 and −.29, 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 its average vote increases is significantly greater under maximally compact districting plans relative to the existing plans.

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

We are grateful to Alberto Alesina, Roland Benabou, 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 University (applied math), Harvard University (labor economics), the National Bureau of Economic Research Summer Institute (law and economics), and the University of Vienna (math) for helpful discussions and suggestions. Shiyang Cao, Alexander Dubbs, Laura Kang, Eric Nielsen, and Andrew Thomas 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 for its hospitality.

[1] Article 1, section 4, of the U.S. Constitution provides that "[t]he 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."

[*Journal of Law and Economics*, vol. 54 (August 2011)]
© 2011 by The University of Chicago. All rights reserved. 0022-2186/2011/5403-0017$10.00

heterogeneous, altering the landscape of political districts can have large effects on the composition of elected officials. Prior to the 2003 Texas redistricting, the congressional delegation comprised 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 has led to fewer minority-friendly policies (see Shotts [2002] for an excellent overview and critique).

There are several factors that weigh on the constitutionality of districting plans: (1) equal population (the Supreme Court first established this principle for congressional districts in *Wesberry v. Sanders* (376 U.S. 1 [1964]), (2) contiguity (which is a requirement in 49 state constitutions), and (3) compactness. This last 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 Justice Antonin Scalia in *Vieth v. Jubelirer,* 541 U.S. 267 [2004]). There is no consensus on how to adequately measure compactness.[4]

In this paper, we propose a simple index of compactness based on the average physical distance between voters and show that this index has a number of attractive features. The index is the ratio of distance between voters in the same political district under a given plan and the minimal such distance achievable by any possible districting plan. The greater this ratio, which we call the relative

[2] In the United States, political boundaries are typically redrawn every 10 years, after the decennial census. The 2003 middecade redistricting in Texas is a notable exception. The Supreme Court recently held that this was not unconstitutional in *League of United Latin American Citizens v. Perry,* 548 U.S. 399 (2006).

[3] The apportionment acts of 1842, 1901, and 1911 contained a compactness requirement. In *Davis v. Bandemer* (476 U.S. 173 [1986]), Justices Lewis Powell and John Paul Stephens pointed to compactness as a major determinant of partisan gerrymandering, and Justices Byron White, William Brennan, Harold Blackmun, and Thurgood 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* (541 U.S. 267, 290), "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. 159 (O'Connor, J., concurring in judgment)." First, the courts use compactness as one of several criteria. Second, it is an open question whether more compact districting plans have a positive or negative effect on racial or political representation.

proximity index (RPI), the less compact a district.[5] The index satisfies three desirable properties: (1) voters are treated equally (anonymity), (2) increasing the distances between voters within a political district leads to a larger value of the index (clustering), and (3) the index is invariant to the scale, population density, and number of districts in a state (independence). In Appendix A, we show that any compactness index that satisfies these properties ranks districting plans identically to the relative proximity index.

The RPI has several advantages over existing measures of compactness. First, it is the only compactness index that 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 (using the perimeter length of political districts, for example) 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, although 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 U.S. 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] Existing algorithms to solve similar problems in computer science and computational biology work only for small samples (≈ 100) or do not require that partitions have the same size. We develop an algorithm for approximating this partitioning problem that 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; Richter-Gebert,

[5] For the empirical analysis and characterization of the optimally compact districting 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."

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

[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.

Sturmfels, and Theobald 2003), condensed matter physics, and toric geometry and string theory (Diaconescu, Florea, and Grassi 2002).[8]

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, 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 40 percent more compact than the typical districting plan. States that are more compact tend to be states with a larger share of minorities and a larger difference between the percentages 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 its districting plans to the Department of Justice (under section 5 of the Voting Rights Act) is also highly correlated with compactness. With only 43 observations, these estimates are not statistically significant. The rank correlations between the RPI and the most popular indexes of compactness, dispersion, and perimeter are $-.37$ and $-.29$, respectively.

We conclude our analysis by estimating a counterfactual of the 2000 congressional elections in California, New York, Pennsylvania, and Texas using optimally compact districts derived from our algorithm. To better understand the impact that 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. Seat-vote curves are a common tool that political scientists use to analyze the partisan consequences of districting plans. These curves are characterized by two things: bias and responsiveness. Bias reports, when the vote is split, twice the difference between the seat share the Democrats get and 50 percent. Responsiveness is the fraction of seats the Democrats get if the average vote goes up 1 percent. Responsiveness can be interpreted as a measure of the nature of democracy in the state. For instance, if Responsiveness is 1, then representation is proportional to the share of the vote. If it is greater than 1, it is majoritarian, and if it were to be infinity, then it would be winner take all.

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 show 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

[8] 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 is 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.

relative proximity index and provides a brief discussion of its properties. 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 with a discussion of potential extensions and generalizations of our approach. There are five appendixes. Appendix A contains an axiomatic derivation of the RPI, showing that any index that satisfies our three axioms will rank districting plans identically to the RPI. Appendix B provides further technical details, including a formal description of the algorithm used to compute maximally compact districts and proofs of all technical results. Appendix C provides a guide to programs to calculate the RPI, Appendix D contains figures comparing actual district maps and those obtained from our algorithm, and Appendix E contains figures comparing seat-vote curves.

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" (364 U.S. 340). 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.[9] As Justice Sandra Day O'Connor put it, "We believe that reapportionment is one area in which appearances do matter" (*Shaw v. Reno,* 509 U.S. 603, 647 [1993]).

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 (478 U.S. 173), and Justices Byron White, Brennan, Blackmun, and Marshall described it as an important criterion (106 S. Ct. 2797, 2815). In *Vieth,* the plurality acknowledged compactness as a traditional districting principle. Justice Anthony Kennedy, in his concurring opinion, stated 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

[9] In *Shaw v. Reno* (509 U.S. 630 [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]).

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 gerrymander is the same regardless of whether the group is identified as political or racial" (541 U.S. 127, 335). 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 that have been proposed (see also Young 1988).[10] Niemi et al. (1990) classify existing measures into four categories: (1) dispersion measures, (2) perimeter measures, (3) population measures, and (4) other miscellaneous measures.[11] The important takeaway is that 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.

One class of dispersion measures are based on length versus width of a rectangle that 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.[12] At least two moment-of-inertia measures have been suggested. Schwartzberg (1966) and Kaiser (1966) consider the variance of the distances from each point in the district to the district's 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 measure based on fractal dimensions was proposed by Knight (2004). Various authors have proposed measures that 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.[13]

There are three population-based measures. Hofeller and Grofman (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.

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

[11] We draw heavily on their summary and classification.

[12] Reock (1961) proposes a circle, Geisler (1985) a hexagon, Horton (1932) and Gibbs (1961) a circle with diameter equal to the district's longest axis, and still others use the smallest convex figure (see Young 1988).

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

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: the absolute deviation of district area from average area in the state (Theobald 1970), a measure based on the number of reflexive and nonreflexive 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.

3. The Relative Proximity Index

3.1. Basic Building Blocks

Let **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 $V_S = \{v_1^S, \ldots, v_n^S\}$ 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 and allow n to represent a generic integer. We restrict voting districts to be equal in size, up to integer rounding.[14,15] Let $\mathcal{V}_S$ denote the set of all partitions of S that satisfy this restriction. We say that 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

The RPI is the ratio of two components. The numerator sums the pairwise squared distance between voters within each district in a state, as given by the actual districting plan in the state. The denominator is that same sum but for the districting plan that minimizes the sum.

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

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

[14] This was first held as a requirement by the Court in *Baker v. Carr* (369 U.S. 186 [1962]) 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 percent difference in the populations of its smallest and largest districts.

[15] In symbols, $|v_i^S| \in \{\lfloor |S|/|V_S| \rfloor, \lceil |S|/|V_S| \rceil\}$ for all $v_i^S \in V_S$, where $\lceil x \rceil = \inf\{n \in \mathbb{Z} \mid x \leq n\}$ and $\lfloor x \rfloor = \sup\{n \in \mathbb{Z} \mid n \leq x\}$.

Similarly, let $V_S^\star = \arg\min_{V_S \in \mathcal{V}_S} \{\pi(V_S)\}$. The RPI, for a partition of state S, V_S, is given by

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

The RPI is well defined if $\pi(V_S^\star) \neq 0$, which holds so long as all voters are not located at the same point.

In the nondegenerate 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 three times less compact than a state's maximal compactness.

3.3. A Constructive Example

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 RPI. 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 the 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 $\{\{1, 2, 3\}, \{4, 5, 6\}\}$.[16] We perform the same calculation as above for each of those partitions and then take the minimum of these 10 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 $2(1^2 + 2 + 1^2 + 2 + 1^2 + 1^2) = 16$. The index is thus $24/16 = 3/2$.

The example provides a snapshot of the RPI 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.

[16] 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}}, and {{1, 5, 6}, {2, 3, 4}}.

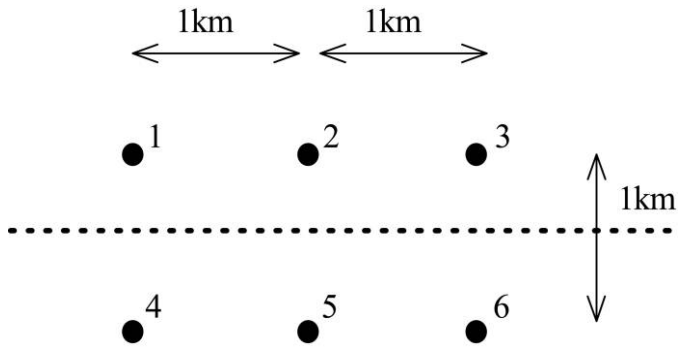

Figure 1. A simple example

3.4. Three Desirable Properties

Any desirable index of compactness should satisfy three properties. (Formal mathematical statements of these properties are provided in Appendix A.)

Anonymity. The index does not depend on the identity of any given voter.

Invariance. The index does not depend on a state's population density, physical size, or number of districts.

Clustering. If two states with the same number of voters, the same number of voting districts, and the same value for the minimum-partitioning problem have different total intradistrict distances, then the state with the larger value is less compact.

It is straightfoward to see from the above example that the RPI satisfies these properties. All voters are weighted equally, so anonymity is satisfied. The denominator of the RPI scales the index so that invariance is satisfied. Finally, clustering is satisfied because the numerator sums pairwise squared distances. In fact, we can say something much stronger:

Theorem 1. Any compactness index that satisfies anonymity, invariance, and efficient clustering ranks districting plans identically to the RPI.

Proof. See Appendix A.

The result is proved by noting that by transforming a given state (expanding the set of individuals and number of districts, for example) it can be compared to another state. Anonymity and independence ensure that this can be done in a way that does not alter the compactness index, and clustering then allows a comparison of two districting plans based on their total intracluster pairwise distances.

4. Implementing the Relative Proximity Index

In this section, we apply the RPI to the districting plans of the 106th Congress. The challenge with calculating our index is computing the denominator, which requires finding a districting plan that minimizes the distance between voters. This is a complex combinatorial problem for which existing algorithms are inadequate. We solve this problem by showing that optimal districting plans are akin to so-called power diagrams[17] and then modifying an algorithm presented in Aurenhammer, Hoffmann, and Aronov (1998) to create a power diagram. The key ingredient in the algorithm is the centroid, or geometric center, of existing districts,[18] a point that is provided in census data from the GeoLytics database. We apply our algorithm to the data from the 2000 census and calculate both the optimal districting plan following that census and the relative proximity index for the actual districting plans employed to elect 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 $\{(n-1)!/[(n/d-1)!\,(n-n/d)!]\}^{d-1}$. So, for California alone, using data at the tract level, $n = 6{,}800$ and $d = 53$. The cardinality of the set of feasible partitions is $78.4 \times 10^{59{,}351}$. Technically speaking, the problem is nondeterministic polynomial-time hard (NP hard).

Similar problems arise in fields such as applied mathematics (computer vision), computer science and operations research (the k-way equipartition 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 were simply distance, rather than distance squared, the problem would be precisely the k-way equipartition problem, which has received considerable attention in computer science and is related to a literature in computational biology employing minimum-spanning trees to partition sim-

[17] Power diagrams are a generalization of Voronoi diagrams due to Aurenhammer (1987). Voronoi diagrams are convex polygons with the important feature that each contains a so-called generator point such that that all other points within the polygon are closer to that generator point than to generator points of adjacent polygons.

[18] More precisely, a centroid is the intersection of all straight lines that divide the district into two parts of equal moment about the line.

ilar genes into clusters.[19] 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 $O[n\log(n')]$ time, where n' is the number of voters and n 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, which were introduced into computational geometry by 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, which is some arbitrary point, is given by the function $\text{pow}_\lambda(x, m_i) = \|x - m_i\|^2 - \lambda_i$, where $\|\cdot\|$ is the Euclidean norm.[20] 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 $\text{pow}_\lambda(x, m_i) < \text{pow}_\lambda(x, m_j)$ for all $j \neq i$. Roughly speaking, voters are placed in the district whose centroid they are closest to. 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_i's 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 $\|x - m_i\|^2 - \|x - m_j\|^2 = \lambda_i - \lambda_j$.

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 B.

This theorem follows from three lemmas that partially characterize an optimal

[19] Without the constraint that each district must 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 nondeterministic polynomial-time (NP) complete. An approximation for it in a general metric space that runs in $n^{O(1/\epsilon)}$ 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.

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

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 us 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 powers of the two centroids are equal. It represents all points at which one is indifferent between placing voters in one district or 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 is 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 were not perpendicular to the line connecting their centroids, then one could not be indifferent between points being in one district or 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.[21] Notice that, 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 caveats. First, we have not yet proven that there is a unique power diagram for every set of starting values. Second, we are able to map optimal districting plans into power diagrams only 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.

Figure 2A depicts the optimally compact districting plan for a hypothetical state. There are nine voters, arranged so that the state is a lattice. The stars

[21] Aurenhammer, Hoffmann, and Aronov (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 perpendicular- 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.

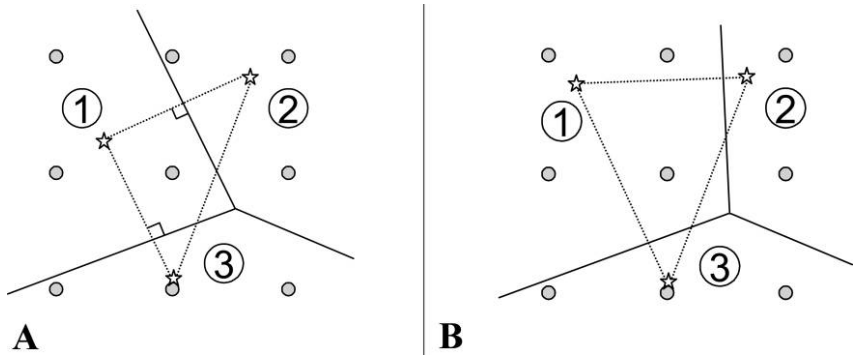

Figure 2. Good and bad generator points

represent the 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 for 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 that 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 Figure 2B, which chooses alternative generator points than those used to partition in Figure 2A. The generator point used for district 1 differs from that used in Figure 2A, resulting in four voters being placed in district 1 and only two in district 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, Hoffmann, and Aronov (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.

We provide a complete formal treatment in Appendix B and here give a heuristic description of the algorithm. The algorithm takes the centroids of existing districts as starting generator points and computes a power diagram. Power diagrams do not require partitions (cells) to be even roughly equal, so, after constructing the diagram, the algorithm adjusts the district boundaries until

the number of voters within each district is equal up to integer rounding. We then recalculate the centroids of the new districts and check to see if any pair of individuals can switch districts and reduce the objective function (total squared distances). Our modification of Aurenhammer, Hoffmann, and Aronov's algorithm continues to check until there are no more pairs that can be switched and reduce the objective function by a predetermined value. The algorithm then repeats itself—recalculating centroids, drawing power diagrams, adjusting boundaries, and so on—until it reaches a value within preset bounds for a stopping rule.

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 United States, its 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 U.S. census from the GeoLytics database, which contain the latitude and longitude of the geographic centroid of each tract, the political district each centroid is in, and its total population.[22] 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. Our main interest in using this level of aggregation (relative to blocks or block groups) is that census tracts are more likely to contain some notion of communities of interest.

An important consideration in the application of the RPI is how to handle tracts of different densities. 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; the burden will lessen 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 four smaller parts and assigned them to the same longitude and altered their latitude by .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,

[22] 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.

$\sum_{v \in V} \sum_{i,j \in v} (d_{ij})^2$, 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 maximum deviations from equal partitions in the actual data and those resulting from our algorithm are an indication of the degree to which the equal-size constraint holds. The bootstrapping technique that we used for the mean RPI is described 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, .46) is substantially larger than our largest deviation (California, .22).

Table 1 illustrates that the five states with the most compact districting plans are Idaho, Washington, Arkansas, Mississippi, and New Hampshire. The five most 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-partitioning problem is more than 40 percent more compact than the typical districting plan. The rank correlations between the RPI and the most popular indexes of compactness, dispersion and perimeter, are $-.37$ and $-.29$, respectively.

Axiom 3 (invariance to scale, population density, and number of districts; see Appendix A) ensures that the RPI can be compared across states, but it does not guarantee that the distribution of RPI values across states is the same. It is entirely plausible that it is easier (a lower percentile of the distribution of RPI values from feasible partitions) to obtain a given value of RPI for Texas than, say, Florida. Thus, gleaning an understanding of how sensitive RPI values are for a given state is difficult.

To try to address this issue, we calculated 200 RPI values for each state by randomly generating starting values for the algorithm. Table 1 reports the means and associated standard deviations from this process and in 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 the 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 nonoptimal. To obtain maximal compactness in these states, a significant restructuring is likely needed.

To understand what state demographic characteristics are correlated with compactness, we estimate a state-level ordinary least squares regression where the dependent variable is the RPI and the independent variables are the percentages

Table 1
The Relative Proximity Index, 2000

State	RPI	Max Deviation		Mean RPI	SD RPI	Percentile
		Actual	Algorithm			
Alabama	1.21	.27	.05	.99	.03	1.00
Arizona	1.34	.20	.15	1.27	.04	.97
Arkansas	1.08	.14	.05	.78	.01	1.00
California	1.49	.17	.04	.96	.03	1.00
Colorado	1.59	.15	.05	1.28	.02	1.00
Connecticut	1.36	.02	.01	1.09	.35	.78
Florida	1.39	.46	.07	.83	.08	1.00
Georgia	1.24	.14	.09	.90	.01	1.00
Hawaii	1.59	.09	.04	1.48	.02	1.00
Idaho	.97	.10	.02	.80	.02	1.00
Illinois	1.43	.29	.11	.98	.07	1.00
Indiana	1.49	.20	.06	1.05	.02	1.00
Iowa	1.38	.06	.05	1.29	.01	1.00
Kansas	1.11	.08	.05	.95	.01	1.00
Kentucky	1.51	.14	.05	1.22	.01	1.00
Louisiana	1.15	.13	.05	.79	.43	.80
Maine	1.39	.04	.03	1.15	.01	1.00
Maryland	1.52	.22	.04	1.25	.02	1.00
Massachusetts	1.87	.10	.05	1.54	.01	1.00
Michigan	1.24	.13	.04	.99	.02	1.00
Minnesota	1.05	.16	.05	.90	.02	1.00
Mississippi	1.02	.18	.05	.87	.01	1.00
Missouri	1.38	.23	.05	1.01	.16	.99
Nebraska	1.01	.05	.04	.89	.23	.70
Nevada	1.38	.08	.05	1.19	.01	1.00
New Hampshire	1.10	.01	.00	1.09	.00	.95
New Jersey	2.27	.21	.05	1.69	.02	1.00
New Mexico	1.23	.06	.04	1.14	.01	1.00
New York	1.83	.21	.10	1.45	.45	.80
North Carolina	1.33	.28	.04	1.15	.09	.97
Ohio	1.62	.13	.05	1.42	.01	1.00
Oklahoma	1.24	.09	.05	1.42	.36	.31
Oregon	1.26	.09	.04	1.21	.28	.56
Pennsylvania	1.81	.25	.22	1.27	.05	1.00
Rhode Island	1.18	.03	.02	1.18	.01	.55
South Carolina	1.22	.21	.04	1.27	.02	.00
Tennessee	2.91	.25	.04	2.59	.04	1.00
Texas	1.90	.30	.22	1.24	.07	1.00
Utah	1.46	.06	.04	1.40	.01	1.00
Virginia	1.38	.22	.07	1.14	.04	1.00
Washington	1.17	.15	.06	.77	.03	1.00
West Virginia	1.68	.06	.05	1.61	.01	1.00
Wisconsin	1.40	.11	.08	1.22	.58	.62

Note. Relative proximity index (RPI) values are calculated using tract-level data from the 2000 census. Max Deviation is calculated as 1 minus the total population of the largest congressional district divided by the total population of the smallest congressional district. Mean RPI is calculated as the mean of 200 repetitions of the RPI, each having different starting values.

of the populations that are black, Asian, or Hispanic; population density; difference in presidential vote shares between Democrats and Republicans; and whether the state is required to submit its districting plans to the Department of Justice under the preclearance provision of section 5 of the Voting Rights Act (not shown in tabular form).[23] States that are more compact tend to be states with a larger share of blacks and a larger difference between the percentages 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 its districting plans is also highly correlated with compactness. Consistent with axiom 2 (efficient clustering; see Appendix A), the RPI is uncorrelated with population density. It is important to note that none of these partial correlations are statistically significant because of small samples.

Beyond the technical considerations, perhaps the best evidence in favor of our approach can be illustrated visually. The figures in Appendix D present side-by-side comparisons of congressional district maps for actual districting plans and those obtained from our algorithm.[24] Figures D1 and D2 illustrate this comparison for the least and most compact states, Tennessee and Idaho, respectively. The districts in Tennessee, under the current plan, resemble the salamander-shaped districts drawn by Eldridge Gerry that gave rise to the name "gerrymandering." Under the algorithm, however, Tennessee is transformed into a neat set of convex polygons. Idaho is at the other extreme. Because the state need only be cut into two equal parts, the existing cut and our preferred cut are very similar. Further, our partition provides a more equal distribution of voters across the districts, which explains why the calculated RPI is slightly less than 1.

These figures illustrate three 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 that rank relatively high (low) in terms of the RPI appear to quite different (similar) to the partition resulting from our algorithm. Third, Figures D3 and D6 (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 another. The issues faced by residents of the outer islands might well be more similar to each other than they are to those of residents in Honolulu. This serves to highlight why compactness is only one factor that weighs on the redistricting question. The RPI in its current implementation ignores this consideration. An RPI with a more general notion of

[23] The states that are subject to the preclearance provision are Alabama, Alaska, Arizona, Georgia, Louisiana, Mississippi, South Carolina, Texas, and Virginia.

[24] For a complete set of maps, see Roland Fryer, Papers (http://www.economics.harvard.edu/faculty/fryer/papers_fryer).

distance or carefully selected starting values for the power diagram can address this issue.

4.5. How Good an Approximation?

One wonders how good an approximation our algorithm provides to an exact solution to the minimum-partitioning problem. We have two ways to address this question. The first is to note that the compter science literature on power diagrams and algorithms based on them (see, for example, Aurenhammer, Hoffmann, and Aronov 1998) shows that thse algorithms typically perform very well (to within a few percentage points of the actual optimum). This can be shown by taking hypothetical data sets to which the exact solution can be found (because they are sufficiently small) and then comparing the performance of the algorithm. Yet it is not clear how performance on these algorithms scales.

One might also wonder whether the use of tract-level data (rather than finer grained block-level data) leads to markedly less precision. To address this, we ran several smaller states at the block level. The average RPI calculated at the block level is slightly higher than in the tract-level analysis reported in Table 1. For instance, Nebraska has an RPI of 1.01 in the tract-level data and 1.33 using blocks. The key issue with block-level analysis is our inability to calculate RPI for medium or large states. On computers with eight high-speed processors and 16 gigabytes of RAM (such as the one we used in our analysis), we estimate that large states such as Texas and California would take several years each to finish.[25]

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 to the U.S. Congress and for electoral votes for the U.S. presidency, the winner 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, which yields an easy condition for an election win in a district.

Assuming that there are n districts, labeled $i \in [1, \ldots, n]$, let ϕ_i denote the proportion of the two-party vote received by the candidate from the first

[25] Currently, large clusters or supercomputers can run at above 1.5 petaflops (a petaflop is 10^{15} floating point operations per second), and the IBM Sequoia project is projected to run at 20 petaflops by 2011. That is roughly the power of 2,000,000 laptops, or around 11,000 times faster than the machine on which we conducted our analysis. Thus, analysis of our index at the block level will be feasible soon.

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: one for single-member districts or three or more for the electoral college, for example. Two important summary statistics are the average district vote, $\Phi = (1/n) \sum_{i=1}^{n} \phi_i$, and the seat share, $S = \sum_{i=1}^{n} s_i / \sum_{i=1}^{n} w_i$.

Many other statistics can be generated using the vote and seat outcomes directly, but we are particularly interested in partisan bias and responsiveness. Namely, Bias $= 2E(S|\Phi = .5) - 1$ estimates the deviation from the median share of seats if each side receives an identical average district vote, and Responsiveness $= (dS/d\Phi)|\Phi$ estimates how a small shift in the average district vote would translate into a shift in the share of seats. This estimate is taken at either the observed average district vote or the median vote. Bias measures the degree to which an evenly divided state would elect an uneven slate of representatives, and Responsiveness is the fraction of seats the Democrats get if the average vote goes up 1 percent.

5.1. Data and Statistical Framework

Our empirical strategy has four steps. First, we estimate a cross-sectional regression of Democratic vote shares on controls such as past election results and incumbency using the 2000 congressional districting plan. The regression is at the voter tabulation district (VTD) level, a subdivision of congressional districts. Second, using the optimally compact congressional districting plans we devised in Section 4, we reassign voter districts to new congressional districts. Not only will this change how voter district results are aggregated to the congressional district level, it will also change some of the controls for each voter district. Third, we use the coefficient estimates and the estimate of residual variance from the voter district regression to simulate outcomes under both the actual districting plan and the optimally compact districting plan. Finally, we aggregate VTD-level results up to the congressional districts in each simulation and compare the distribution of simulations across the two districting plans.

We use VTD-level election return data from U.S. elections for the 105th and 106th Congresses for four large states: California, New York, Pennsylvania, and Texas. These states were chosen because of their large numbers 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, although 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 contains 13,000 for 30.

The intuition behind our approach is straightforward. Consider Figure D7, 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 cho-

sen by aggregating the votes from the VTDs within each district. In Figure D7, this amounts to adding votes from roughly 433 voting centers in districts 27–31. Now suppose we want to estimate how the choice of representatives would 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 D7 this now amounts to districts 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 the candidate is an incumbent—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 terse synopsis of their approach below. 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 applied before the election campaign began and would be identical if the election were to be 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 that 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 $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.

There are two key assumptions in the Gelman-King method. First, errors are expressed 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; $\lambda = \sigma_\gamma^2/(\sigma_\gamma^2 + \sigma_\varepsilon^2)$. Second, the counterfactual assumes that the regrouping of voters into new districts will not have a systematic effect on voting behavior.

5.1.1. 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 (that is, in which the electoral map is identical to that of the previous election) and taking the mean, we have an estimator for λ.[26]

5.1.2. 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 $\mathbf{X}$ 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

$$\phi_{\text{hyp}} = \mathbf{X}_{\text{hyp}}\beta + \delta_{\text{hyp}} + \omega, \tag{3}$$

where β is the posterior distribution, with mean $\hat{\beta} = (X'X)^{-1}X'\phi$ 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:

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

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

5.1.3. 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 (for example, incumbency), either by aggregating and averaging the previous values in each precinct or by making sensible predictions for their value. 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 his or her district. Under the new districting plan, if there is one such incumbent per district, he or she becomes the incumbent used in the model. In the rare cases where there is more than one incumbent assigned to a district under a new districting plan, we break the tie by choosing the incumbent

[26] Ideally, one would have historical votes for many years to tease out the systematic error component. We have only 2 years of such data.

closest to the resulting centroid and moving the other incumbent to another district to keep the numbers constant. Finally, with our new map we simulate the model 1,000 times; deriving the relevant parameters is straightforward.

5.2. Analyzing Seat-Vote Curves

Using the methodology described above, the figures in Appendix E 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 axes depict the proportion of seats won by Democrats. The horizontal axes depict 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 is the fraction of seats the Democrats received in the election (not the hypothetical seat share). The dark lines represent our estimate of the seat-vote curve, and the two lines parallel to them are 95 percent confidence intervals. One can see that there is a marked difference between the seat-vote curves estimated 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 states. The slopes in 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, although 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 fivefold increase, from .482 to 2.51. In other words, under the current partition, a 1 percent increase in vote share for Democrats results in a .482 percent increase in seats. When districting is maximally compact, however, a 1 percent increase in vote share results in a 2.51 percent increase in seats. The next largest change is in California—increasing from 1.086 to 1.731. Pennsylvania and Texas show smaller increases, which are statistically significant at the 10 percent 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 states and time, a feature not found in existing measures, and our algorithm provides a way of approximating the most compact plan. Further, the impact that a maximally compact districting plan can have on the responsive of votes is encouraging. These are first steps

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

	Bias			Responsiveness		
State	Actual	Algorithm	*t*-Statistic	Actual	Algorithm	*t*-Statistic
California	.028	.007	.469	1.086	1.731	−4.327*
	(.010)	(.045)		(.069)	(.132)	
New York	.103	.018	1.051	.482	2.51	−6.540*
	(.014)	(.080)		(.036)	(.308)	
Pennsylvania	−.0027	.031	−.363	1.138	1.562	−1.800^{+}
	(.021)	(.076)		(.128)	(.198)	
Texas	.062	.039	.334	.8872	1.305	−1.717^{+}
	(.024)	(.064)		(.103)	(.221)	

Note. Estimates are based on voter tabulation district–level election return data for the 105th and 106th Congresses.
$^{+}$ Statistically significant at the 10% level.
* Statistically significant at the 5% level.

toward a more scientific understanding of districting plans and their effects. Extensions and generalizations abound.

Perhaps the most obvious extension is to consider higher dimensional spaces, generalized distance functions, and communities of interest. Aurenhammer and Klein (2000) provide a comprehensive survey of Voronoi diagrams and how to incorporate generalized notions of distance, including *p*-norms, convex and airlift distances, and nonplanar spaces. These extensions are not only mathematically interesting and elegant: they have real-world content. Consider the following thought experiment. Suppose there is a city on a hill.[27] On the west side is a mild, long incline toward the rest of the city, which is in a plane. On the east side is a steep cliff, either impassable or with just a narrow, winding road that very few people use. While the next residential center to the east is much closer to the hilltop on a horizontal plane, it is much farther in terms of all sorts of distances that we think might matter: transportation time, intensity of social interactions, sets of shared local public goods and common interests, and so forth. Thus, for all practical purposes, one probably wants to include the hilltop in a western district rather than an eastern one. More general notions of distance can handle this. A similar situation arises when there is a natural boundary (for example, a river or highway) that effectively segregates or reduces communication between two population centers that are geographically very close. Conversely, there could be something (such as a tunnel or subway) that makes two nonconnected regions effectively close to each other, or there may be other notions of communities and shared interest that lend themselves to a natural clustering. It is imperative to note that the derivation of our index assumed only a general metric space—many of these ideas fit squarely within our framework. The empirical application of the index, however, required us to only consider Euclidean

[27] We are grateful to Roland Benabou for this illustrative example.

distances. The challenge ahead is to incorporate more general notions of distance into an empirically tractable algorithm.

Appendix A

An Axiomatic Derivation of the Relative Proximity Index

A1. Three Properties

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

A1.1. Axiom 1: Anonymity

Axiom 1, 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, and so forth) of voters. Consider a state S with associated partition V and compactness index $c(V, S)$. For any bijection h: $S \rightarrow S$ and compactness index $c_h(V, S)$, $c_h(V, S) = c(V, S)$.

A1.2. Axiom 2: Clustering

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).[28] Our second axiom requires that if two states with the same number of voters and voting districts and the same value for the minimum-partitioning problem have different weighted intradistrict distances, then the state with the larger value is less compact.

Let $\gamma_k = \sum_{i,j \in \nu} \alpha_{ij} (d_{ij})^{\delta}$ for $k = \{1, \ldots, n\}$ and let $g(\gamma_1, \ldots, \gamma_n) : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 and the same number of districts, and

$$\min_{V \in \nu_{S_1}} g_{S_1}(\gamma_1, \ldots, \gamma_n) = \min_{V \in \nu_{S_2}} g_{S_2}(\gamma_1, \ldots, \gamma_n).$$

Then

$$g_{S_1}(\gamma_1, \ldots, \gamma_n) > g_{S_2}(\gamma_1, \ldots, \gamma_n) \Rightarrow c(V, S_1) > c(V', S_2).$$

A1.3. Axiom 3: Independence

Our final axiom requires that any measure of the 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. Before stating the property formally, we need some further notation. We say that a state $\hat{S}$ is

[28] 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).

an n replica of S if and only if $\forall i \in S$, $\exists j_1, \ldots, j_n \in \hat{S}$ such that $d_{ij} = 0$, $\forall i$ and $d_{j_j j_k} = 0$, $\forall i, k$. 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.

Consider S, $\hat{S} \in S$ with cardinality $|S|$ and $|\hat{S}|$, respectively.

Scale. If $d_{ij} = \lambda d_{ij}$ for all $i, j \in S$, $\hat{S}$ then $c(V, S) = c(V, \hat{S})$ for all V.

Density. If $|\hat{S}| = \lambda |S|$ and $\hat{S}$ is a λ replica of S, then $c(V, S) = c(V, \hat{S})$ for all V.

Number of Districts.

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

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, Mass., and New York City, for example) that 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.

A2. Uniqueness Result

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

Theorem 1.

1) The relative proximity index satisfies anonymity, clustering, and independence.

2) Suppose that $\delta = 2$ and that $g_{S_i}(\cdot)$ is symmetric for all i; then any compactness index that satisfies anonymity, clustering, and independence is ordinally isomorphic to the relative proximity index.

A2.1. Proof of Theorem 1.1

That the RPI satisfies the three axioms follows from five simple lemmas that 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 $h: S \to S$. The term $\sum_{v \in V_S} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2$ is unchanged since h is a bijection, and hence there are the same number of points 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

$$\sum_{v \in V_S^1} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 > \sum_{v \in V_{S'}^2} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 . \tag{A1}$$

Clustering requires that

$$c(V_S^1, S) > c(V_{S'}^2, S).$$

Suppose, by way of contradiction, that expression (A1) holds, and

$$c(V_1, S) < c(V_2, S). \tag{A2}$$

That is,

$$\frac{\sum_{v \in V_S^1} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2}{\min_{V \in \nu_S} \sum_{v \in V} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2} < \frac{\sum_{v \in V_{S'}^2} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2}{\min_{V \in \nu_S} \sum_{v \in V} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2} . \tag{A3}$$

The denominators are identical, and hence the supposition requires that

$$\sum_{v \in V_S^1} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 < \sum_{v \in V_{S'}^2} \sum_{i \in v} \sum_{j \in v} (d_{ij})^2 , \tag{A4}$$

a contradiction. Q.E.D.

Lemma 3. The relative proximity index satisfies density independence.

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

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

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

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

which is clearly equal to the left-hand side for any partition. Q.E.D.

Lemma 4. The relative proximity index satisfies scale independence.

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

$$\frac{\sum_{\nu \in V_S} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2}{\min_{V \in \nu_S} \sum_{\nu \in V} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2} = \frac{\sum_{\nu \in V_{\hat{S}}} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2}{\min_{V \in \nu_S} \sum_{\nu \in V} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2}$$

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

$$\frac{\sum_{\nu \in V_S} \sum_{i \in \nu} \sum_{j \in \nu} (\lambda d_{ij})^2}{\min_{V \in \nu_S} \sum_{\nu \in V} \sum_{i \in \nu} \sum_{j \in \nu} (\lambda d_{ij})^2} = \frac{\lambda^2 \sum_{\nu \in V_S} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2}{\lambda^2 \min_{V \in \nu_S} \sum_{\nu \in V} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2},$$

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

Lemma 5. The relative proximity index satisfies number-of-districts independence.

Proof. The proof follows immediately from the definition of independence with respect to number of districts. Q.E.D.

We can now prove theorem 1.2. 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 that does not alter the compactness index, and clustering then allows a comparison of two districting plans to be made based on their total intracluster pairwise distances.

A2.2. Proof of Theorem 1.2

From theorem 1.1 we have $\mathrm{RPI}(V, S_m) > \mathrm{RPI}(\hat{V}, S_n) \Rightarrow c(V, S_m) > c(\hat{V}, S_n)$ for any m, n. Suppose that theorem 1.2 is not true. This implies that

$$c(V, S_m) > c(\hat{V}, S_n) \quad \text{and} \quad \mathrm{RPI}(V, S_m) < \mathrm{RPI}(\hat{V}, S_n) \tag{A5}$$

or

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

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 combined with equality implies that g is additively separable in its arguments. Then by equality and clustering we have

$$\sum_{\nu \in V_{S_m}} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2 > \sum_{\nu \in \hat{V}_{S_n}} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2 \Rightarrow c(V, S_m) > c(\hat{V}, S_n),$$

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

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

we have

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

By clustering this implies that $c(V,\ S_m) < c(\hat{V},\ S_n)$, a contradiction. Identical reasoning rules out the case in which

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

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 λ replica of S_n and a μ replica of S_m so that the number of voters is the same as in the transformed state S_m. Note that $c(V,\ S_m)$ and $\text{RPI}(V,\ S_m)$ are unchanged because of independence. In a slight abuse of notation we will continue to use V and S_m in reference to the μ-replicated state. Second, expand or contract the state in the sense that the distance between any two points—say, d_{ij}—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

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

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

$$\min_{V \in \nu_{S_m}} \sum_{\nu \in V_{S_m}} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2 = \min_{V \in \nu^{\gamma_m}_{S_{n'}}} \sum_{\nu \in V_{S_{n'}}} \sum_{i \in \nu} \sum_{j \in \nu} (d_{ij})^2. \tag{A6}$$

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

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

and that

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

Hence,

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

By independence,

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

and

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

From expression (A5),

$$c(V,\ S_m) > \frac{\beta}{\theta} c(\hat{V}',\ S_{n'}) \quad \text{and} \quad \mathrm{RPI}(V,\ S_m) < \frac{\beta}{\theta} \mathrm{RPI}(\hat{V}',\ S_{n'}). \tag{A7}$$

But since S_m and $S_{n'}$ have the same number of voters, the same number of districts, and equation (A6) holds, it follows that expression (A7) implies that c violates clustering.

Identical reasoning rules out the case in which

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

and hence the proof is complete. Q.E.D.

Appendix B

Proofs and Description of the Algorithm

B1. 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 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. The function $\Psi(D_1, \ldots, D_d)$ 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.

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

Q.E.D.

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. It suffices to show that substituting m_i for z_i minimizes the expression on the right. Its first-order condition with respect to z_i is

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

Q.E.D.

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. 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 $D_1' = D_1 \bigcup \{v_2\} - \{v_1\}$ and $D_2' = D_2 \bigcup \{v_1\} - \{v_2\}$. By the optimality of $(D_1, \ldots, D_d)$ and the optimality lemma,

$$\begin{aligned}
\psi(D_1) + \psi(D_2) \le\ & \psi(D_1') + \psi(D_2') \le \psi_{m_1}(D_1') + \psi_{m_2}(D_2') \\
\Rightarrow\ & \|v_1 - m_1\|^2 + \|v_2 - m_2\|^2 \\
\le\ & \|v_1 - m_2\|^2 + \|v_2 - m_1\|^2 \\
\Rightarrow\ & -2v_1 \times m_1 - 2v_2 \times m_2 \\
\le\ & -2v_1 \times m_2 - 2v_2 \times m_1 \\
\Rightarrow\ & (v_2 - v_1) \times (m_1 - m_2) \le 0 \\
\Rightarrow\ & (x_2 - x_1) \times (-\xi) + (y_2 - y_1) \times 0 \le 0 \\
\Rightarrow\ & x_1 \le x_2.
\end{aligned}$$

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, which shows 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. Q.E.D.

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. Without loss of generality, we prove this lemma 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 that there exists $q \in \mathbb{R}^2$ 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 do so 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 zero 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 they are 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 $D_1' = D_1 \bigcup \{v_3\} - \{v_1\}$, $D_2' = D_2 \bigcup \{v_1\} - \{v_2\}$, and $D_3' = D_3 \bigcup \{v_2\} - \{v_3\}$. 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. Q.E.D.

Proof of Theorem 2. We prove that any optimal districting is a power diagram with sites 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 that 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. Q.E.D.

B2. Algorithm Details

The algorithm we propose is a modification of the second algorithm presented in Aurenhammer, Hoffmann, and Aronov (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, Hoffmann, and Aronov (1998) algorithm for finding a power diagram that 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

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

Aurenhammer, Hoffmann, and Aronov (1998) simplify 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:

$$\Lambda_{t+1} = \Lambda_t + \varepsilon_t \times \nabla \xi(\Lambda_t).$$

In our implementation we set Λ_0 to be the zero vector. It remains to pick the step sizes $\{\varepsilon_t\}_{t \geq 0}$. 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

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

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

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

The size 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 Section B1) 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 not all of the D_i's are 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 that the algorithm described in Aurenhammer, Hoffmann, and Aronov (1998) tends to fail when one of the districts is randomly set to zero. 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.

Appendix C

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 of the main programs below.

Main.m and Compute_Index.m are both shell programs that call District.m, the actual algorithm, and store its output in text files. Typing Compute_Index(File Name, 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 2 and 3, respectively, the federal information processing standards (FIPS) code of the state repeated in every entry of column 4, the current districts of all census tracts in column 5, and the populations of all census tracts in column 6. 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 RPIs, the second has the accuracy of the districting, and the third has the accuracy of the current districting. Accuracy is measured as

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

Compute_Index.m has the following hard-coded parameters that are passed to District.m: outside_tol_ratio, tol_ratio, outside_bail, and bail. The parameters tol_ratio and bail are the stopping criteria for the subroutine 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, respectively, .9 times the real accuracy, whichever is the lesser of .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(File Name) reads a list of states and iterations for each state to be run by Compute_Index.m. The file is of the following form:

states	bootstraps
alabama	4
arizona	7
arkansas	3
california	1

Names of states and numbers of iterations are separated by tabs. If "arizona" is written in this file, Compute_Index.m will open a file called "arizona.out." Main.m creates an additional file called "index.txt" that 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 greater than one 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.

Appendix D

Congressional District Map Comparisons for the 106th Congress

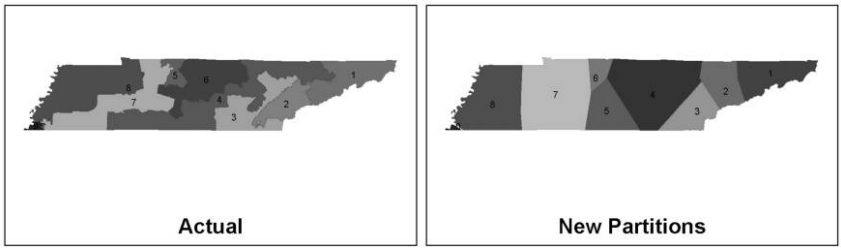

Figure D1. Tennessee

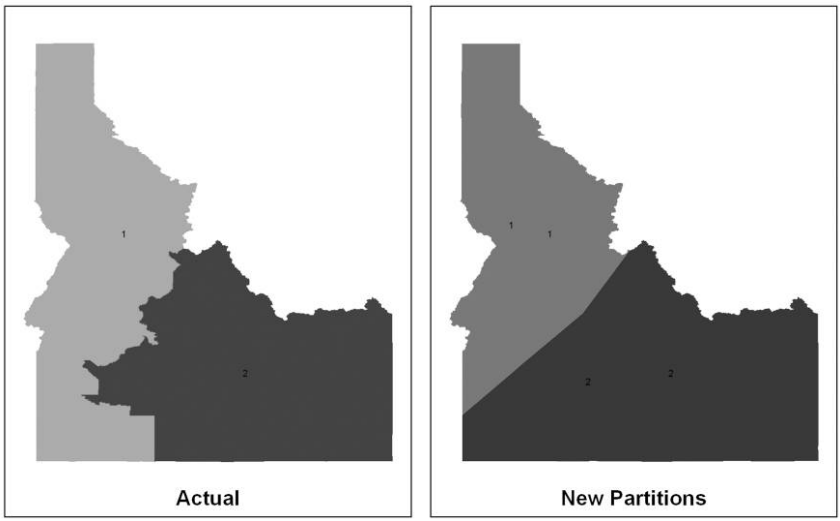

Figure D2. Idaho

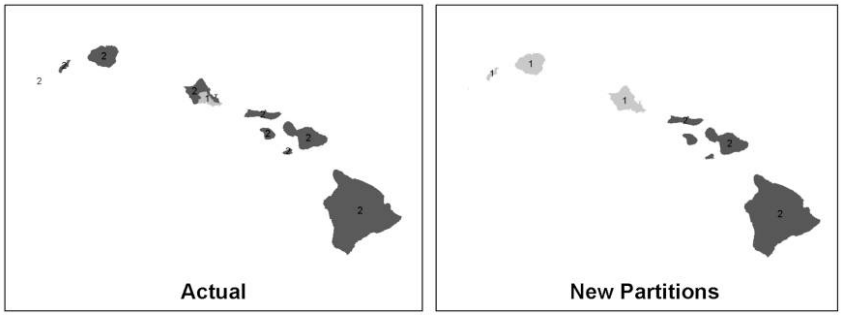

Figure D3. Hawaii

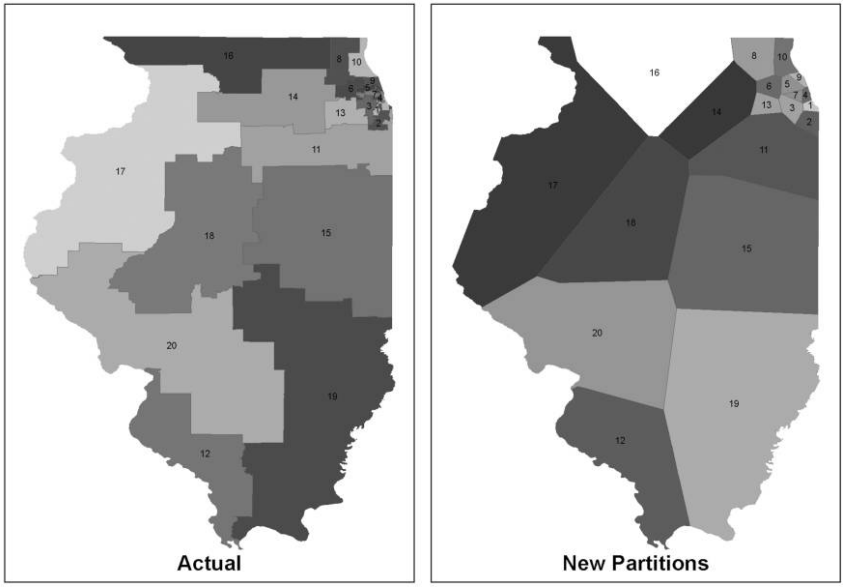

Figure D4. Illinois

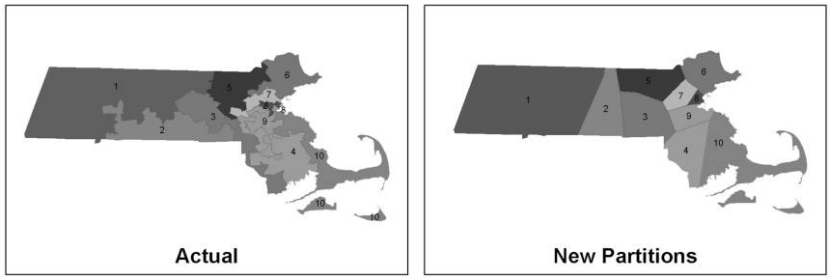

Figure D5. Massachusetts

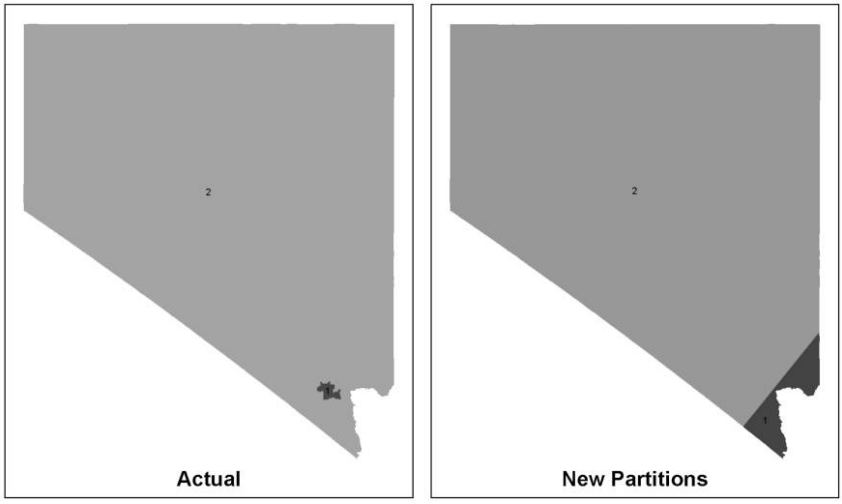

Figure D6. Nevada

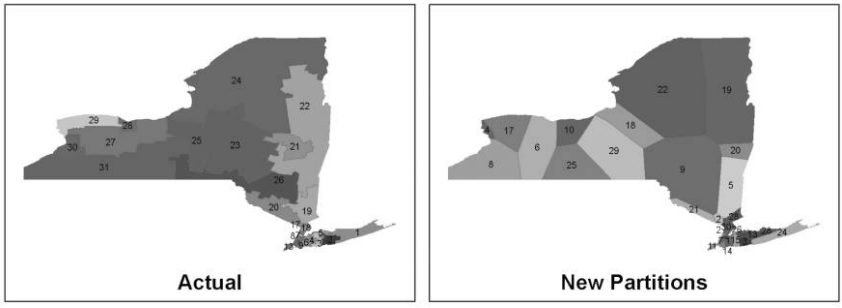

Figure D7. New York

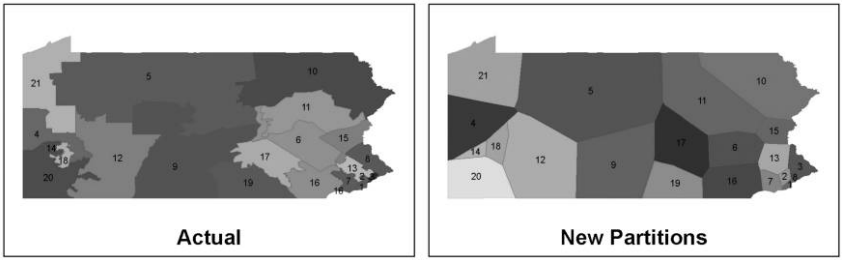

Figure D8. Pennsylvania

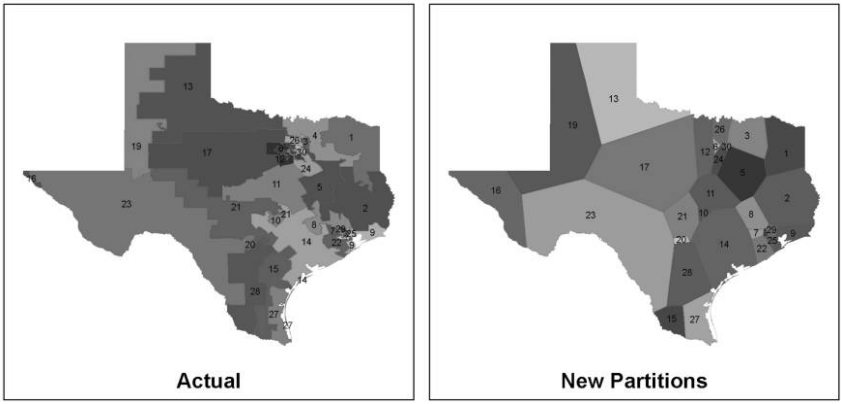

Figure D9. Texas

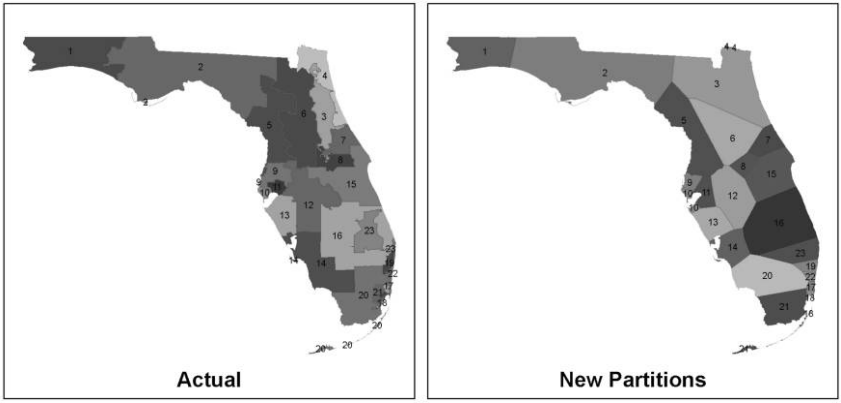

Figure D10. Florida

Appendix E

Comparison of Actual and Maximally Compact Seat-Vote Curves

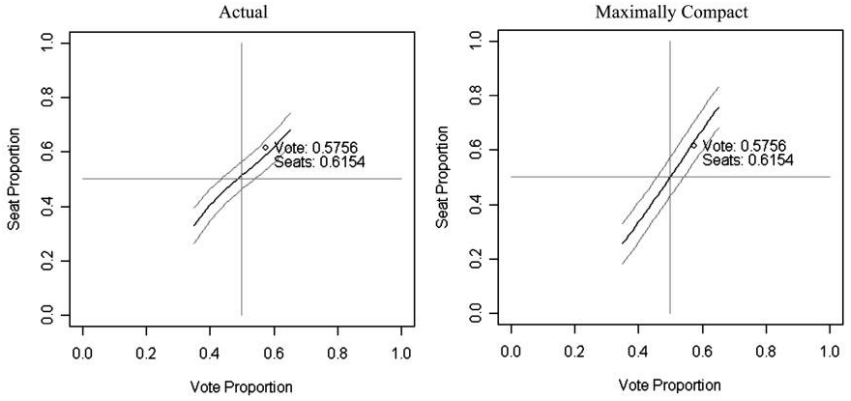

Figure E1. California

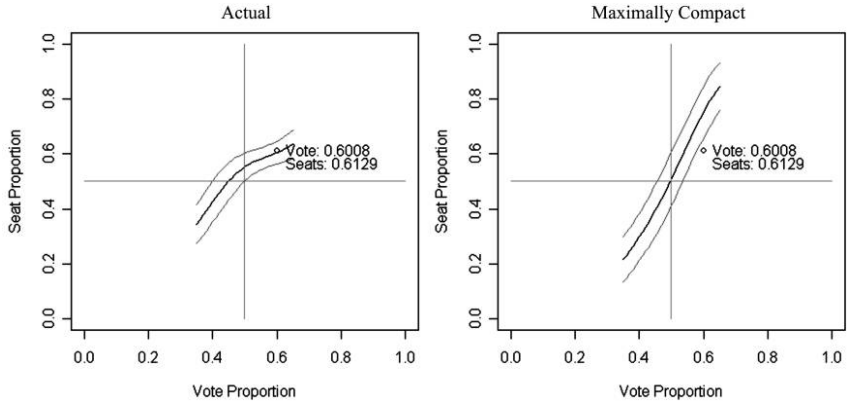

Figure E2. New York

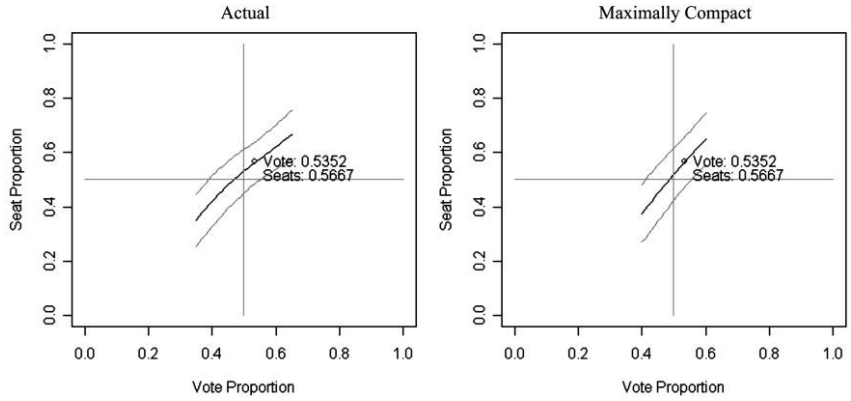

Figure E3. Texas

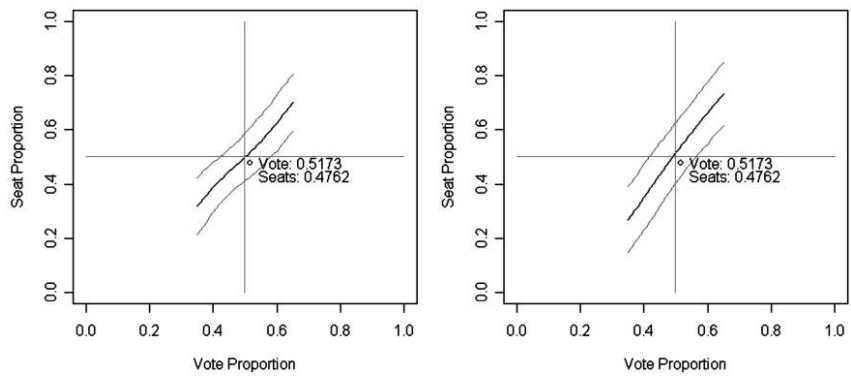

Figure E4. Pennsylvania

References

Adams, Bruce. 1977. A Model State Reapportionment Process: The Continuing Quest for "Fair and Effective Representation." *Harvard Journal on Legislation* 14:825–904.

Arrow, Kenneth. 1970. *Social Choice and Individual Values.* Cowles Foundation Monograph Series. New Haven, Conn.: Yale University Press.

Attneave, Fred, and Malcolm D. Arnoult. 1956. The Quantitative Study of Shape and Pattern Perception. *Psychological Bulletin* 53:452–71.

Aurenhammer, Franz. 1987. Power Diagrams: Properties, Algorithms and Applications. *SIAM Journal on Computing* 16:78–96.

Aurenhammer, Franz., F. Hoffmann, and B. Aronov. 1998. Minkowski-Type Theorems and Least-Squares Clustering. *Algorithmica* 20:61–76.

Aurenhammer, Franz, and Rolf Klein. 2000. Voronoi Diagrams. Pp. 201-90 in *Handbook*

of Computational Geometry, edited by J. Sack and G. Urrutia. Amsterdam: Elsevier Science.

Barabas, Jason, and Jennifer Jerit. 2004. Redistricting Principles and Racial Representation. *State Politics and Policy Quarterly* 4:415–35.

Bartal, Yair, Moses Charikar, and Danny Raz. 2001. Approximating min-sum k-Clustering in Metric Spaces. Pages 11–20 in *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing.* New York: Association for Computing Machinery.

Boyce, Ronald R., and W. A. V. Clark. 1964. The Concept of Shape in Geography. *Geographical Review* 54:561–72.

Cox, E. P. 1927. A Method of Assigning Numerical and Percentage Values to the Degree of Roundness of Sand Grains. *Journal of Paleontology* 1:179–83.

Diaconescu, D. E., B. Florea, and A. Grassi. 2002. Geometric Transitions, del Pezzo Surfaces and Open String Instantons. *Advanced Theoretical and Mathematical Physics* 6:643–702.

Eig, Larry M., and Michael V. Seitzinger. 1981. *State Constitutional and Statutory Provisions Concerning Congressional and State Legislative Redistricting.* Report No. 81–143A. Washington, D.C.: Congressional Research Service.

Geisler, R. Gene. 1985. Measuring Gerrymandering of Legislative Districts. Unpublished manuscript.

———. 1994. A Unified Method of Evaluating Electoral Systems and Redistricting Plans. *American Journal of Political Science* 38:514–54.

Gibbs, J. P. 1961. A Method for Comparing the Spatial Shape of Urban Units. Pp. 99–106 in *Urban Research Methods*, edited by J. P. Gibbs. New York: Van Nostrand.

Harris, Curtis C., Jr. 1964. A Scientific Method of Districting. *Behavioral Science* 9:219–25.

Hofeller, Thomas, and Bernard Grofman. 1990. Comparing the Compactness of California Congressional Districts under Three Different Plans: 1980, 1982, and 1984. Pp. 281–88 in *Political Gerrymandering and the Courts*, edited by Bernard Grofman. New York: Agathon.

Horton, R. E. 1932. Drainage Basin Characteristics. *Transactions of the American Geophysical Union* 13:350–61.

Ji, Xiaoyun, and John E. Mitchell. 2005. Finding Optimal Realignments in Sports Leagues Using a Branch-and-Cut-and-Price Approach. *International Journal of Operations Research* 1:101–22.

Kaiser, Henry F. 1966. An Objective Method for Establishing Legislative Districts. *Midwest Journal of Political Science* 10:200–213.

Knight, Jeffrey L. 2004. GIS Based Compactness Measurement Using Fractal Analysis. Department of Economics, Brown University, Providence, R.I.

Mandelbrot, Benoit B. 1967. How Long Is the Coast of Britain? Statistical Self-Similarity and Fractal Dimension. *Science* 156:636–38.

Mehrotra, Anuj, Ellis L. Johnson, and George L. Nemhauser. 1998. An Optimization Based Heuristic for Political Districting. *Management Science* 44:1100–14.

Mitchell, John E. 2003. Realignment in the National Football League: Did They Do It Right? *Naval Research Logistics* 50:683–701.

Mumford, David, and Jayant Shah. 1989. *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems.* Cambridge, Mass.: Harvard University Press.

Niemi, Richard G., Bernard Grofman, Carl Carlucci, and Thomas Hofeller. 1990. Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan and Racial Gerrymandering. *Journal of Politics* 52:1155–81.

Papayanopolous, L. 1973. Quantitative Principles Underlying Legislative Apportionment. *Annals of the New York Academy of Sciences* 19:181–91.

Passare, Mikael, and Hans Rullgard. 2004. Amoebas, Monge-Ampere Measures, and Triangulations of the Newton Polytope. *Duke Mathematics Journal* 121:481–507.

Pounds, Norman J. G. 1972. *Political Geography.* 2nd ed. New York: McGraw-Hill.

Reock, Ernest C., Jr. 1961. Measuring Compactness as a Requirement of Legislative Apportionment. *Midwest Journal of Political Science* 5:70–74.

Richter-Gebert, Jurgen, Bernd Sturmfels, and Thorsten Theobald. 2003. First Steps in Tropical Geometry. Unpublished manuscript. Technische Universität München, Zentrum Mathematik, Munich. http://arxiv.org/abs/math/0306366.

Sahni, S., and T. Gonzales. 1976. P-Complete Approximation Problems. *Journal of the ACM* 23:555–66.

Schwartzberg, Joseph E. 1966. Reapportionment, Gerrymanders, and the Notion of Compactness. *Minnesota Law Review* 50:443–52.

Shotts, Kenneth W. 2002. Gerrymandering, Legislative Composition, and National Policy Outcomes. *American Journal of Political Science* 46:398–414.

Taylor, Peter J. 1973. A New Shape Measure for Evaluating Electoral District Patterns. *American Political Science Review* 67:947–50.

Theobald, H. Rupert. 1970. Equal Representation: A Study of Legislative and Congressional Appointment in Wisconsin. Pp. 71–260 in *The State of Wisconsin Blue Book,* edited by H. Rupert Theobald and Patricia V. Robbins. Madison: Wisconsin Legislative Reference Bureau.

Weaver, James B., and Sidney W. Hess. 1963. A Procedure for Nonpartisan Districting: Development of Computer Techniques. *Yale Law Journal* 72:288–308.

Wells, David I. 1982. How to Inhibit Gerrymandering: The Purpose of Districting Criteria. *National Civic Review* 71:183–87.

Young, H. P. 1988. Measuring the Compactness of Legislative Districts. *Legislative Studies Quarterly* 13:105–15.