TOWARDS DISCRETE GEOGRAPHY

MOON DUCHIN AND BRIDGET EILEEN TENNER

ABSTRACT. We discuss the "compactness," or shape analysis, of electoral districts, focusing on some of the most popular definitions in the political science literature, which compare area to perimeter. We identify four problems that are present in these and all contour-based scores of district geometry. Instead, we set the stage for *discrete* versions of classical shape scores, laying out definitions, goals, and questions for a promising new fusion of discrete geometry with electoral geography.

1. Introduction

A variety of elections in the United States—for the House of Representatives, state legislatures, city councils, school boards, and many others—are conducted by partitioning a region into geographically-delimited *districts* and selecting one winner per district via a plurality election. A suitable partition of the region is called a *districting plan*, and the act of revising it is called *redistricting*. States may have their own regulations governing the redistricting process, often including specific requirements for valid plans, and the procedures and outcomes are the subject of endless debate and legal scrutiny.

There are two main principles commonly applied to the geometric forms of districts: districters should aspire to make cut jurisdictions into pieces that are contiguous and compact. Districts that are topologically connected (as opposed to being made up of separate connected components) are called *contiguous* in the redistricting literature, and this feature is a widespread and uncontroversial requirement of districting plans.*

In contrast, *compactness* gestures at the "reasonableness" of district shapes, but is rarely defined precisely, if at all, in state rules. Even when it is left vague, the notion of compactness is critical to any discussion of redistricting (and, in particular, to any discussion of abusive districting practices broadly known as *gerrymandering*) because compactness appears as an explicit requirement in many states and is nationally recognized as a traditional districting principle.† To date, more than thirty possible definitions of compactness as a shape quality metric have been proposed in the political science literature (see [1, 11, 4] and references).

The purpose of this paper is to call attention to a shared feature in nearly all of the existing definitions of compactness: they represent districts as regions enclosed by contours on a map, then base their numerical scores on lengths and areas from the map. This makes all the classical definitions of compactness susceptible to some common flaws, undermining the extent to which the definitions can be made precise and meaningful. In this paper we discuss the problems and limitations of contour-based compactness scores and we call for research towards creating a new generation of *discrete* scores that make more fundamental use of the population network that is at the heart of the redistricting problem.

2010 *Mathematics Subject Classification.* Primary: ; Secondary:

*This is largely unambiguous, but there are numerous examples where contiguity is achieved by connective tissue along a highway or through water.

†See `http://redistricting.lls.edu/where-state.php#compactness`, which describes compactness requirements of some kind in 37 of the 50 states.

Acknowledgements. This work began during the Geometry of Redistricting Workshop held at Tufts University in August 2017. We are grateful to the sponsors, speakers, and participants. MD is supported by NSF DMS-1255442. BT is partially supported by a Simons Foundation Collaboration Grant for Mathematicians and a DePaul University Faculty Summer Research Grant. Special thanks to Assaf Bar-Natan, Richard Barnes, Seth Drew, Lee Hachadoorian, Adriana Rogers, and Alejandro Velez-Arce for computing and data support.

2. Compactness and electoral geography

2.1. **Introducing compactness.**

The political relevance of requiring districts to be reasonably shaped—and not unnecessarily elongated or twisting—can be defended in several ways. Geometric eccentricity can signal a districting plan that has been engineered to produce an extreme outcome, often by exploiting demographics and geography in order to maximize representation for one group at the expense of another. Extreme outcomes can be constructed by *packing* the out-group into a small number of districts, with wastefully high vote share in those districts, and *cracking* their leftover population by diluting and dispersing it. Either strategy can require distended district shapes in order to achieve its purpose. A second, related argument for limitations on shapes is that any limitation placed on districters is a healthy check on their power. But third and maybe most fundamental is that being more compact should mean that districts are more surveyable, transitable, and cognizable, representing chunks of territory that have or can develop a meaningful unity. A historical overview and taxonomy of compactness definitions can be found in [4].

Key to various attempts at evaluating a district's shape by some notion of compactness is a measurement of area and perimeter. For instance, as discussed below, one of the most prominent definitions is a simple ratio of area to perimeter-squared. In this paper we argue that standard measures of area and perimeter are problematic in their application to redistricting. To the extent that districting plans are defined by how they partition the *people* in a jurisdiction, they rely on fundamentally *discrete* data, which is obscured in the focus of map contours.

2.2. **Districts and their building blocks.**

Political jurisdictions—even states themselves—have legal definitions, but the usable formats for communicating the definitions are usually based on discrete approximations to the legal definitions. In the age of GIS (geographic information systems), the most common data format for geographic units is called a *shapefile*. These files store a definition of each unit as a polygon, possibly with many thousands of vertices. For instance, Census geography comes in finest units called *blocks*, which nest into larger units called *block groups*, which in turn nest into *tracts*. These geographic units are supposed to partition the state that they belong to, meaning that the entire territory of the state (land and water) must be covered by the census units, and furthermore that the units at a given scale are disjoint from each other, except along their borders. Congressional districts are essentially always made out of whole census blocks (though they frequently cut across block groups and tracts), and within a given state each districts must have very nearly the same population as the others, counted by adding the Census population of their blocks.*

*There are a few subtleties here; for instance, large bodies of water such as Great Lakes may or may not be included in any district in a state. Except for issues like these, the Congressional districts form another partition of the state. The issue of which population to count is discussed further below.

The Census Bureau releases an updated *vintage* of its most precise shapefiles—so-called TIGER/Line Shapefiles [19]—every year, with a special release for congressional districts once they have stabilized after each decennial Census. It also releases Cartographic Boundary Shapefiles of districts, which are intended for the purpose of map-drawing rather than definition, for every Congress, at three levels of resolution [20].

3. Polsby-Popper and other contour-based scores

> "Compactness," unlike contiguity, is a continuous concept that concerns the geographical shapes of districts. There is no bright line test that determines whether a district is or is not compact, but districts may be considered more or less compact. While numerous quantitative measures of compactness have been proposed for this purpose, the two measures that are now referenced the most are a dispersion measure known as the Reock measure and a perimeter measure known as the Polsby-Po[p]per measure.
>
> –Dick Engstrom, *Martinez v. Bush*, expert report

We begin by introducing the most commonly cited compactness metric, the *Polsby-Popper score*, then we discuss its variants and a few alternatives in the legal literature. The motivating idea for Polsby-Popper and its cousins is that a "compact" region should have large area relative to its perimeter. This is an *isoperimetric* score, because it creates a ranking among regions with a given fixed ("iso" = same) perimeter.

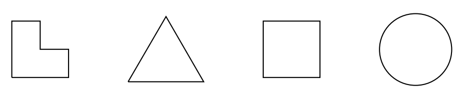

Figure 1. Four regions, each with the same perimeter, are shown from left to right in order of increasing area. The region with largest possible area relative to the fixed perimeter—the circle—is deemed the most "compact" by PP.

3.1. Perimeter versus Area.

It has been known (or guessed) since antiquity that

Circles have the most area among all shapes with a given perimeter.

In other words, all shapes satisfy $0 \le \frac{4\pi A}{P^2} \le 1$, where A stands for area and P stands for perimeter. Depending on the scope of the statement (i.e., on the generality of what counts as a "shape"), this can be credited to Jakob Steiner in the 1830s, although his methods are not fully general from the point of view of today's geometric analysis. Here is a modern statement of the phenomenon.

Theorem 1 (Isoperimetric theorem)**.** *Let Ω be a bounded open subset of $\mathbb{R}^2$ whose boundary $\partial\Omega$ is a rectifiable curve. Then the Lebesgue measure m and the length ℓ are related by the inequality*

$$4\pi \cdot m(\Omega) \le \ell(\partial\Omega)^2,$$

with equality if and only if Ω is a disk and $\partial\Omega$ is a circle.

In this generality, Theorem 1 is easily proved using the Brunn-Minkowski inequality [15]. Despite the pedigree of this theorem, it was a 1991 article by law scholars Polsby and Popper that led to their names on the associated formula in political science [12].

Definition 2. *The* Polsby-Popper score *of a region* Ω *is*

$$\mathsf{PP}(\Omega) := 4\pi \cdot \frac{\mathsf{area}(\Omega)}{\mathsf{perim}(\Omega)^2}.$$

The justification for using this score on voting districts is the idea referenced above, that tentacled or otherwise eccentric districts are red flags for possible gerrymandering. Shapes with skinny necks or long spurs will have much less area than could have been enclosed by the same boundary length around a plumper shape.

Higher Polsby-Popper scores are therefore deemed preferable to lower ones. By Theorem 1, the score satisfies $0 \leq \mathsf{PP}(\Omega) \leq 1$ for all shapes, with $\mathsf{PP}(\Omega) = 1$ realized only when Ω is a circle. The squaring of the perimeter in the denominator of the Polsby-Popper score makes the units of measurement cancel out, so that the score is (theoretically) scale-invariant. In other words, $\mathsf{PP}(k\Omega) = \mathsf{PP}(\Omega)$, meaning that if one were to enlarge an entire region by a factor of k, its Polsby-Popper score would not register the change. Thus this metric is designed to measure something about the shape, and not the size, of a district.

Polsby-Popper scores are cited in dozens of court cases on redistricting.*

Remark 3. *One easily verifies that* $\mathsf{PP}(\Omega)$ *is also equal to the ratio of the area of* Ω *to the area of a circle with the same perimeter.*

A cosmetic variant of the Polsby-Popper score, which in fact predates Polsby-Popper in the literature, is the *Schwartzberg score*, originally defined as the ratio of the perimeter of a district to the perimeter (circumference) of a circle having the same area, so that lower Schwartzberg scores are deemed preferable to higher ones [14]. This is expressed by

$$\mathsf{Schw}(\Omega) := \frac{\mathsf{perim}(\Omega)}{\sqrt{4\pi \cdot \mathsf{area}(\Omega)}} = \mathsf{PP}(\Omega)^{-1/2}.$$

From the power relationship between Schw and PP, it is clear that, although specific scores will differ, Schwartzberg and Polsby-Popper assessments must rank districts from best to worst in precisely the same way.† Because Schwartzberg worried that there was no way (with 1966 technology) to accurately measure perimeters of districts, he also proposed a notion of *gross perimeter*, which actually comes from a partial discretization [14]. As a result, software like Maptitude for Redistricting uses a different definition of perimeter in the computation of a Schwartzberg score than in the computation of a Polsby-Popper score, which of course can break their monotonic relationship. See [5] for an example of how these can differ.

The language used to formulate compactness scores is often imprecise in court documents, so what sound like different combinations of area and perimeter often point back to these two scores. For instance, there are numerous references to "perimeter-to-area" and "area-to-perimeter" scores in expert reports. These names suggest computations of P/A or A/P,

*See *Louisiana House of Representatives v. Ashcroft*, 539 U.S. 461 (2003); *Martinez v. Bush*, 234 F. Supp. 2d 1275 (S.D. Fla. 2002); *Perez v. Perry*, 835 F. Supp. 2d 209, 211 (W.D. Tex. 2011); *Vesilind v. Virginia State Board of Elections*, 15 F. Supp. 3d 657, 664 (E.D. Va. 2014); *Page v. Judd*, ***; *Sanders v. Dooly County*, 245 F. 3d 1289 (11th Cir. 2001); *Sessions v. Texas*, ***; *Session v. Perry*, 298 F. Supp. 2d 451 (E.D. Tex. Jan 6, 2004); *U.S. v. County of Los Angeles*, ***; *Harris v. McCrory*, 159 F. Supp. 3d 600, 611 (MDNC 2016); *Johnson v. Miller*, 922 F. Supp. 1552 (SD Ga. 1995); *Cromartie v. Hunt*, 526 U.S. 541 (1999); *Moon v. Meadows*, 952 F. Supp. 1141 (ED Va. 1997); and many more.

†This is because for positive values of x and y, we have $x > y \iff x^{-1/2} < y^{-1/2}$. Therefore a higher (and thus better) PP score corresponds to a lower (and thus better) Schw score.

but where we have been able to find definitions, the definitions refer again to Polsby-Popper or its reciprocal.*

3.2. **The landscape of compactness metrics.** Though Polsby-Popper scores are frequently cited, there is no consensus on how they should be used when determining the validity of a districting plan. To make matters more confusing, legal contexts often call for the reporting of more than one type of compactness score. Consider the recent litigation-driven Congressional redistricting in Pennsylvania. In the court orders of January 22 and 26, 2018, it is required that "[A]ny redistricting plan the parties or intervenors choose to submit to the Court for its consideration shall include ... [a] report detailing the compactness of the districts according to each of the following measures: Reock; Schwartzberg; Polsby-Popper; Population Polygon; and Minimum Convex Polygon."

It is left completely open whether any of these assessments might be more important than the others, or how two plans are to be compared. Besides Polsby-Popper and Schwartzberg, the court required reporting of three other scores:

- **Reock**: the area of a district divided by the area of its smallest circumscribing circle;
- **Population Polygon**: the population of a district divided by the population contained in its convex hull†; and
- **Minimum Convex Polygon** (also known as the Convex Hull score): the area of a district divided by the area of its convex hull.

A final notable variant is to simply report the total perimeter. For example, the state constitutions of Iowa and Colorado and at least one expert report‡ compare the total area of the jurisdiction (which is constant across alternative/contending districting plans for that jurisdiction) to the sum of all district perimeters.

Each of these metrics, including Polsby-Popper and Schwartzberg, requires drawing a district as a domain on a map of the state bounded by a contour, then using classical (Euclidean) geometry to make some sort of computation. Population Polygon stands out by taking population location into account, but it still relies on the contour in a fundamental way. But modern geometry is not limited to the planar domains of Euclid; in the 20th century, geometry has flourished in a *discrete* setting, where the objects (like *graphs* and *groups* and *complexes*) are made up of individual particles that you can count, rather than smoothly varying quantities. Electoral districts, being made up of census units, therefore lend themselves extremely well to discrete techniques, as we will argue below.

*A few selections from expert reports: (1) "Perimeter-to-Area (PTA) measure compares the relative length of the perimeter of a district to its area. It is represented as the ratio of the area of a circle with the same perimeter as the district to the area of the district." R. Keith Gaddie, *Sessions v. Texas* report; (2) "Perimeter to Area – Known as the Polsby-Popper test, this measure compares the area of the district with the area of a circle of the same perimeter, using the formula 4πArea/Perimeter^2." Todd Giberson, *Perez v. Perry* report; (3) "The Polsby-Popper measure, a perimeter-to-area comparison, calculates the ratio of the district area to the area of a circle with the same perimeter." M. V. (Trey) Hood III, *Vesilind v. Virginia State Board of Elections* report.

†A *convex body* is a region that contains the entire line segment between any two of its points. The *convex hull* of a region is the smallest convex body containing the region. This is sometimes picturesquely referred to as the "rubber-band enclosure."

‡Puerto Rican Legal Defense and Education Fund v. Gantt, 796 F. Supp. 677 (E.D.N.Y. 1992)

4. Problems with contour-based scores

4.1. **Four issues.** As we have seen, the most-cited compactness scores are all contour-based. We now articulate issues that are inextricably tied to the use of Euclidean geometry on map contours, and we offer illustrations for each issue using actual congressional districts.

> **Issue A: Coastline effects.** *Districts with boundaries produced by natural physical features such as coastlines are heavily penalized by compactness scores based on the perimeters of contours.*
>
> **Issue B: Resolution instability.** *Choice of map resolution can have a dramatic impact on contour-based compactness scores, both individually and in comparison to each other, and there is no finest canonical resolution at which data can be gathered.*
>
> **Issue C: Coordinate dependence.** *Choices of map projection and coordinate system can impose drastic changes on the contour-based compactness scores of districts, both individually and in comparison to each other.*
>
> **Issue D: Empty space effects.** *Although unpopulated regions have no impact whatsoever on electoral outcomes, contour-based compactness scores are sensitive to allocation of these areas to districts.*

4.2. **Discussion and examples.** Let us consider each of these issues in turn.

Issue A: Coastline effects. *Districts with boundaries produced by natural physical features such as coastlines are heavily penalized by compactness scores based on the perimeters of contours.*

Contour-based scores with a perimeter component do not recognize when a portion of a region's border may be necessitated by an pertinent geographical feature like a coastline or an irregular state boundary. In such a situation, the region's compactness score may incur a steep penalty for having an erratic border, even though that district border had not been chosen through any questionable or manipulative process. For example, Alabama's 1st congressional district is partly bounded by the Gulf of Mexico to the south, and the Tombigbee and Alabama Rivers to the north. As depicted in Figure 2, this creates sections of eccentric natural boundary. Accordingly, this district ranks in the 1st decile (that is, in the worst ten percent) of Polsby-Popper scores.

Consideration of the coastline issue leads naturally to a related worry about stability of scores to changes in resolution. A coastline border is irregular and, in a sense, unmeasurable. This is the well-known "coastline paradox" sometimes attributed to Benoit Mandelbrot: the length of the coast of Great Britain depends on the size of your ruler [9]. In this way, the quantities $\mathsf{area}(\Omega)$ and $\mathsf{perim}(\Omega)$ depend on the scale of precision used when mapping the region, and can change significantly at different resolutions. In fact, perim is notably more sensitive than area to changes in resolution: admitting finer wiggles in the boundary may only slightly affect a region's area, while causing a substantial increase in the region's perimeter. Indeed it is clear from consideration of curves within a unit square that arbitrarily long perimeter can exist within finite area.

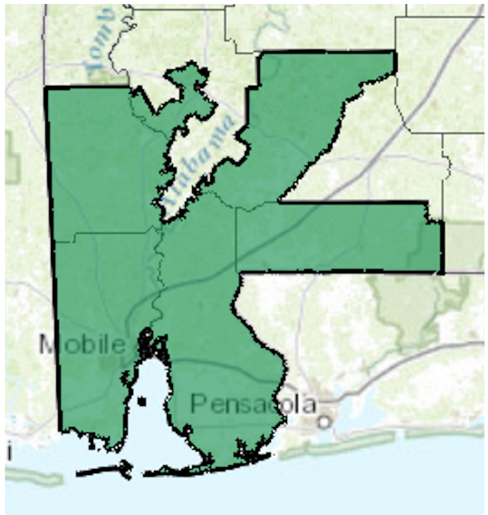

FIGURE 2. Alabama's 1st congressional district, with boundary partly defined by the Gulf of Mexico and the Tombigbee and Alabama rivers.

Issue B: Resolution instability. *Choice of map resolution can have a dramatic impact on contour-based compactness scores, both individually and in comparison to each other, and there is no finest canonical resolution at which data can be gathered.*

Census Bureau Cartographic Boundary files are available in three scales:

500K (1:500,000), 5M (1:5,000,000), 20M (1:20,000,000).

One would expect some variation in the perimeters and areas of districts because the 20M files are greatly simplified, and indeed the Census Bureau itself flags this issue, warning that "These boundary files are specifically designed for small scale thematic mapping... These files should not be used for... geographic analysis including area or perimeter calculation."[20] Nonetheless, we want to use those maps here as an extreme illustration of an issue that will be present whenever map resolution can vary: not only are area and perimeter themselves altered, but those changes are compounded by the way Polsby-Popper is calculated. Perimeter is typically more sensitive to resolution change, and because it is squared, the PP score may drop precipitously at higher resolutions.

For example, California's 53rd congressional district sees an 83.7% jump in perimeter when going from the 20M scale to the 500K scale. In the same transition, the district's area increases by 8.7%, notable both because this change is nonzero and because it is so different from the perimeter change. On average, when comparing data between the 20M scale and the 500K scale, congressional district perimeters increase by 23.2%, and district areas increase by 0.2%. Clearly both statistics are sensitive to resolution, and perimeter is markedly more so.

Pursuing this issue a step further, one might propose that the finest possible resolution will provide the best accuracy, and so the "shortest ruler," to borrow terminology from the coastline paradox, should be used. However, map scale can vary continuously, achieving arbitrarily high or low resolutions. Barnes and Solomon [2] explore this phenomenon fully in a 2018 preprint, where they do not rely on Census Cartographic maps but vary map resolution along a spectrum.

Even in practical terms, each year's Tiger/Line file release tends to offer a slightly modified version of state boundaries from the year before, thought of as achieving gradually improved accuracy. Therefore districting analysts would need to use maps not only from the same source but from precisely the same vintage in order to expect compatible results.

Issue C: Coordinate dependence. *Choices of map projection and coordinate system can impose drastic changes on the contour-based compactness scores of districts, both individually and in comparison to each other.*

The Earth's surface is roughly spherical, but most maps are planar renderings. That is, most maps are flat. It is well known that there are multiple competing methods for projecting regions from a sphere to a plane, and it is impossible to choose a map projection that is faithful to both shape/angle and to area of regions on the sphere. (This is because any smooth map that preserves area and angles must be a local isometry, so must preserve total curvature.)

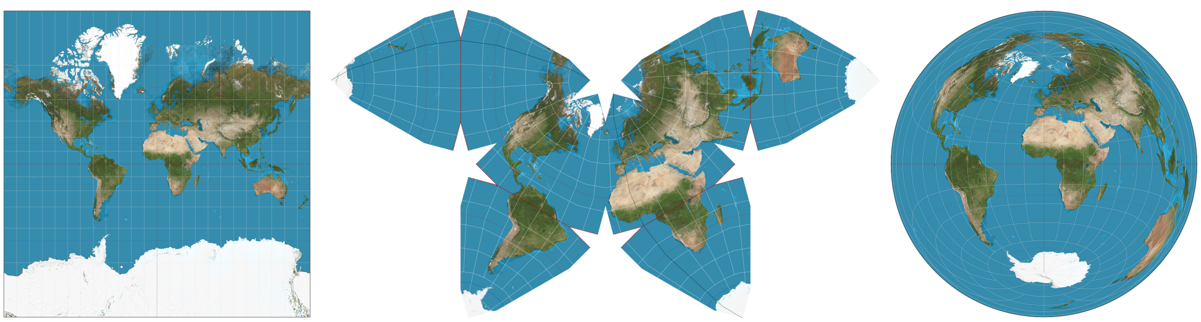

FIGURE 3. Three map projections. The first preserves angle, the last preserves area, and the middle map achieves a partial compromise between the two. [17, 8]

The Reock score, which is not as badly plagued by the first two issues as Polsby-Popper, shows up here as extremely problematic. Hachadoorian et al [7] show that Reock scores, can change by a factor of more than 3 just by choosing a different way to project the earth onto a flat plane. And even the Population Polygon score, which sounds promising because it is population-based, suffers from coordinate dependence. The geometric rendering of the district does not impact the population of the district, of course, but it heavily affects the form of the convex hull, and therefore the population enclosed by it.

Finally, we flag a fundamental problem with the use of area in compactness scores: above, we discussed several worries about whether area of districts is even well-defined, but even beyond that, surface area is of no particular relevance in redistricting.

Issue D: Empty space effects. *Although unpopulated regions have no impact whatsoever on electoral outcomes, contour-based compactness scores are sensitive to allocation of these areas to districts.*

Districts are to be equalized by population, not by acreage, and districts are intended to specify voter assignments. Consider an unpopulated geographical region—an uninhabitable mountain, say—with different districts to its north and south. The assignment of all or part or none of this unpopulated region to the northern district has no influence on voting, and should have no impact on the districts' quality. However, the choice of how to allocate this unpopulated region between the two districts can have a marked influence on their perimeters and areas.

One major source of unpopulated surface area is water. To see the impact that this can have on shapes, consider that by Census measurements, fifteen states are at least 10% water by area, with Michigan topping the list at 41.5%. Overall, water makes up 7% of the United States census geography.[21] More broadly, in the 2010 census, nearly 45% of census blocks had zero reported population.

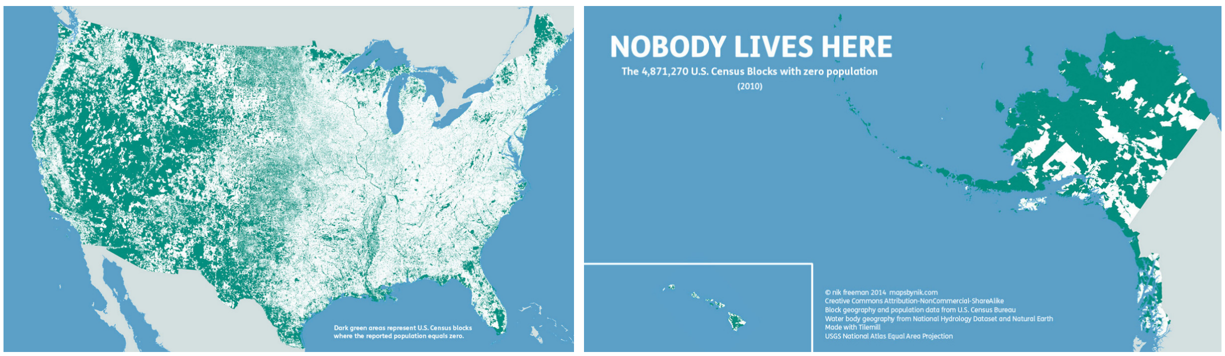

FIGURE 4. Unpopulated census blocks, excluding the Great Lakes, are depicted in dark green in these maps by Nik Freeman [6].

For every possible contour-based score, compactness can therefore be wildly skewed by the assignment of unpopulated surface area.

5. DISCRETE GEOMETRY

5.1. **Discreteness.** The mathematical term *discrete* refers to a set whose elements are distinguishably isolated from each other.* Any set with only finitely many elements is necessarily discrete.

In jurisprudence dating to 1964 and clarified as recently as 2016, the Supreme Court has affirmed that districts are to be created in a way that nearly equalizes their Census population.† The case law that has built up around this has led to the practice of "zeroing out" Census population in Congressional districts; that is, most states equalize population to within a difference of fewer than ten people across districts, and many get the deviation down to a single person. But the Census does not report the locations of individual people: its finest level of detail is the census block. The average number of people per block is about 28, and typical block size is from zero to 200 people. Blocks nest inside other Census-defined geographies such as block groups, tracts, and VTDs, making it possible to calculate

*The technical definition refers to a topological space, or a set with a notion of nearness. Discrete subsets are defined by the property that every point has a small neighborhood with no other set elements in it. By contrast, consider a solid region in the plane: there's no way to isolate a point from all others, no matter how closely you zoom in.

†Reynolds v Sims (1964) gave us the general dictum of "One Person, One Vote," which by common practice was taken to require the near-equalization of Census population across districts. In Evenwel v Abbott (2016), the court technically affirmed that Census population *may* be used for this purpose, leaving the door open for the use of other measurements such as voting-eligible population. However, the decision's deference to "history, precedent, and practice" has left intact the presumption that census-based population will continue to be the standard.

population precisely for these geographies. For these reasons, census blocks, and the larger elements they define, are the natural units for redistricting.

For redistricting purposes, it is also important to know which geographic units are adjacent to which other ones. The basic data type for recording finitely many elements and the adjacencies among them is called a *graph* or *network*.

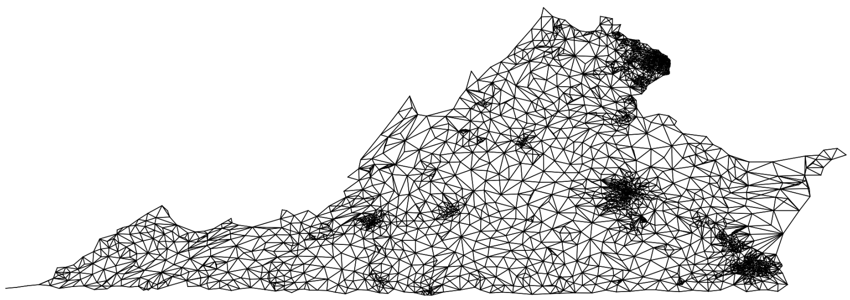

FIGURE 5. The 2372 VTDs in Virginia are shown as the vertices of a graph, with edges indicating geographic adjacencies. (A corresponding graph of the 285,762 census blocks would look much more complicated.)

We will discuss redistricting-specific ideas in the next section, but we first develop some terminology and notation for graphs.

5.2. **Graph basics.** A (simple, undirected) *graph* is defined by a set of *nodes* or *vertices* and a set of *edges* that records which vertices are to be regarded as *adjacent*. Formally, a graph $G = (V, E)$ consists of with vertex set V and edge set E, where each edge $e \in E$ is represented by a pair of distinct vertices $v, w \in V$. We may use the notation $e = \overline{vw}$ to refer to an edge between v and w. A *subgraph* G consists of a subset $\Omega \subseteq V$ of the vertices, together with all edges of G that connect those vertices. (We will sometimes abuse notation in a standard way and use the same symbol Ω to refer to the subgraph or its vertex set, so the size $|\Omega|$ of a subgraph is its number of vertices.)

Define the *internal boundary* $\partial_0\Omega$ of a subgraph Ω to be the set of vertices in Ω that have neighbors outside Ω:

$$\partial_0\Omega := \{v \in \Omega : \overline{vw} \in E \text{ for some } w \notin \Omega\}.$$

A *graph with boundary* $G = (V, E, \partial G)$ is a graph together with a subset of vertices $\partial G \subseteq V$ that are designated as *boundary* vertices. This allows us to define the *total boundary* of a subgraph Ω as $\partial\Omega := \partial_0\Omega \cup (\Omega \cap \partial G)$, which is the interior boundary together with any vertices from the boundary of the ambient graph G itself. Finally, a *vertex-weighted graph* $G = (V, E, w)$ is endowed with a weight function $w : V \to \mathbb{R}_{\geq 0}$ assigning positive numbers to the vertices.

For instance, if the graph in Figure 6 has vertex weights given by

v	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
$w(v)$	2	3	8	4	5	6	7	2	1	9

then one easily calculates that 5/7 of the vertices of Ω are in the boundary by count, but only 2/3 by weight.

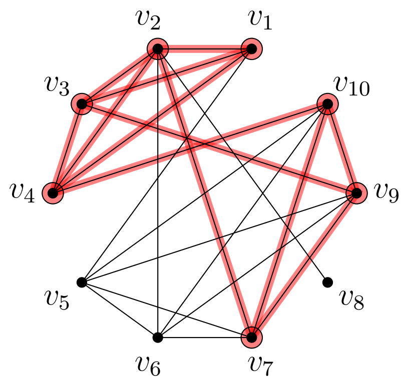

FIGURE 6. A graph G, with a subgraph Ω shown in pink. Here, Ω has five vertices in its interior boundary: $\partial_0\Omega = \{v_1, v_2, v_7, v_9, v_{10}\}$, because each of those is adjacent to at least one of v_5, v_6, or v_8.

5.3. **Motivation from group theory and topology.** In geometric group theory and algebraic topology, it is a standard technique to create combinatorial model spaces out of vertices, faces, and higher-dimensional cells in order to help understand the geometry of groups and manifolds. One can then compare the length of a loop in the edges of the model space (L) to the number of cells in a filling (N). The figure below shows four examples.

These kinds of filling inequalities, called *Dehn functions* in the world of groups, are discrete versions of the isoperimetric ratio, and they have been studied since the pioneering work of Dehn in the early 20th century. For an introductory-level treatment, see [10, Ch 8,9].

The inequalities depicted in the Figure can easily be converted to relate the size of a subgraph to the size of its boundary, with the same exponents in the bounds and only the coefficients changing.*

6. Discretizing compactness

6.1. **Using the census data graph.** We now have notation and terminology from the last section to discuss the representation of a state (or other jurisdiction) as a vertex-weighted graph, where the vertices are units of census geography (blocks, block groups, tracts, VTDs, or counties). Adjacencies are recorded when two units share part of their boundary, and the user can choose whether to include so-called *queen adjacency* in addition to *rook adjacency* (that is, whether to record an edge when two units meet at a corner rather than along a shared boundary of positive length). The weight function on the vertices of this graph G is given by the census population of the corresponding unit. And the boundary ∂G of the graph is the subset of vertices whose units are on the outer boundary of the state. (Here, the user might choose to separately mark vertices according to whether they border land or water, in order to possibly build definitions that counteract coastline effects).

The goal of this short paper is to give new tools for quantitative assessment of the compactness of districts, so we let Ω be the subgraph of the state's graph that corresponds to a district whose compactness is to be assessed. We note that if the units are census blocks,

*We can observe that the number of edges in a simple loop is equal to its number of vertices. Furthermore, if the complex has *bounded geometry* (there is some bound k so that every vertex has degree $\leq k$ and every face has $\leq k$ sides) then the number of faces in a filling, the number of edges, and the number of vertices are all mutually bounded.

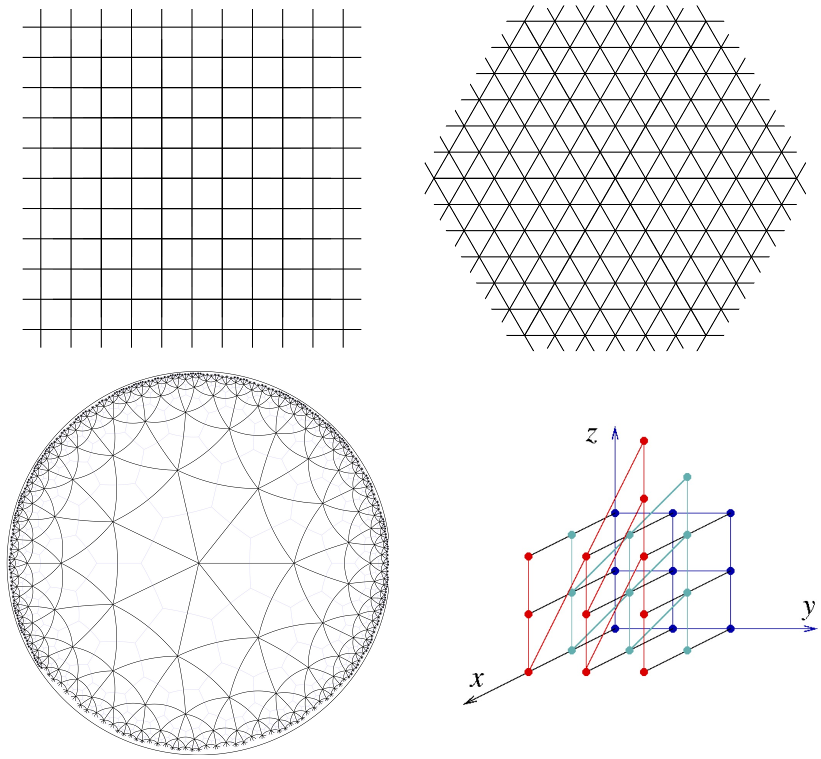

Figure 7. It is instructive to consider edge loops in the graphs depicted here and to compare their length, L, to the number of cells they can enclose, N. Any loop in the square grid satisfies the inequality $N \le \frac{1}{16}L^2$, while the triangular grid has $N \le \frac{1}{6}L^2$. The hyperbolic triangulation at bottom left has $N \le L$ for any loop, no matter how large, while the Heisenberg graph, of which a portion is shown at bottom right, has no inequality better than $N \le L^3$.

then a unit is typically completely in or completely out of a district, but as noted above, districts can cut across the larger geographic units. Therefore if bigger units are chosen, an allocation system is needed, such assigning each unit to the district in which the largest part of its land area or population lies.[*]

Then the principle of discrete compactness scores is to use the geometry of this graph. For instance, taking a cue from the filling inequalities in §5.3, it is natural to let the *discrete area* be given by counting the vertices in the district subgraph and the *discrete perimeter* be the number of vertices in its boundary, possibly choosing to weight both of them by population. We discuss formulas, other ingredients, and directions for future research in the next sections.

6.2. **Ideas and questions for discrete compactness.** First, note that any score based on a census data graph will automatically be independent of map projection or choice of coordinates (Issue C above), since coordinate data is not recorded in the graph.

We begin with several immediate discrete analogs of the Polsby-Popper score of a district. We can let the compactness be measured by:

[*]Other attractive options include an areal threshold for membership, such as including a vertex in the subgraph associated to a district if, say, 10% or 50% of its area lies in the district.

- Discrete area divided by the square of discrete perimeter; or
- The same calculation, but weighted by population.

These scores are, respectively,

$$\frac{|\Omega|}{|\partial\Omega|^2}\ ; \qquad \text{and} \qquad \frac{\sum_{v\in\Omega} w(v)}{\left[\sum_{u\in\partial\Omega} w(u)\right]^2}.$$

To defend the decision to square the perimeter, consider the lattice examples in Figure 8.

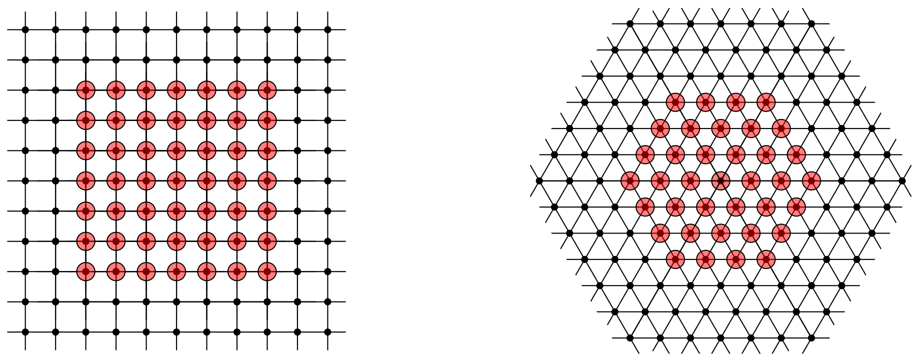

FIGURE 8. Square and triangular lattices with square and hexagonal "districts" Ω_7 and Ω'_4, respectively.

The square-shaped subgraph Ω_n in the square lattice G and the hexagon-shaped subgraph Ω'_n in the triangular lattice G' have isoperimetric ratios

$$\frac{|\Omega_n|}{|\partial\Omega_n|^2} = \frac{n^2}{16(n-1)^2} \qquad \text{and} \qquad \frac{|\Omega'_n|}{|\partial\Omega'_n|^2} = \frac{3n^2-3n+1}{36(n-1)^2},$$

respectively. These tend to definite limits as n gets large (1/16 for the square case and 1/12 for the hexagons), whereas if any other power of perimeter had been used, the calculations would degenerate to zero or infinity for large n. We interpret this to say that if a grid has underlying geometry that is roughly Euclidean, then the squaring of perimeter makes these measurements stable under refinement. (Compare this to the other filling exponents found in the examples from Figure 7.)

We note that there is no need for a coefficient to play the role of 4π from the classical formula. This is because 4π was chosen in order to scale the continuous value PP to lie in the unit interval, whereas these discrete variants can take arbitrarily large or small values.* We think this is a feature, not a bug: it reminds the responsible user to only compare a compactness score to other ones collected at the same resolution, which is good practice anyway with classical scores.

How well do these initial ideas defuse Issues A,B, and D? We conclude this section by studying these in turn, posing questions for any proposed definitions of discrete compactness scores that we hope will drive future research.

At first glance, discretization seems also to dispatch coastline effects (Issue A), since the census geography along the coast will absorb any wiggling perimeter into a fixed (and not unduly large) number of units. But this is an empirical question.

*To see this, note that with a uniform weighting function, a path of length n would have isoperimetric ratio approaching zero; on the other hand, we can produce scores tending to infinity if we begin with a fixed district Ω and successively subdivide interior cells while leaving the boundary fixed.

Question 1. *Does the proposed definition of discrete compactness mitigate coastline issues? To measure this, one should examine a state with ocean or lake coast and determine whether districting plans tend to systematically score the coastal districts worse than inland districts, controlling for other features such as population density. Does the proposed definition introduce new coastline-like effects? One should study the correlates of good or bad performance in any new proposed score to be sure that no irrelevant features are being unduly penalized.*

We note that a possible strategy to inoculate discrete isoperimetry scores from many coastline-type effects entirely is to disregard exterior boundary in the calculation of district perimeter, say by computing $|\Omega|/|\partial_0\Omega|^2$ or its population-weighted counterpart. More generally, the precise definition of district boundary might be profitably expanded in some more delicate way, such as by declaring a vertex v to belong to the boundary if if there exists an edge $\overline{uv}$ such that u is not in the district and $F(u, v) = \text{TRUE}$, where F is some boolean function on the vertices u and v. (It could be $w(u) > 0$, or $w(u) > k$, or a function of both $w(u)$ and $w(v)$, etc.)

Next, some insulation from resolution instability (Issue B) is guaranteed for discrete compactness scores because the only parameter that plays the role of map resolution is the choice of geographic units, and that varies among only finitely many choices over a census cycle rather than varying arbitrarily. However, we should be on the lookout for other kinds of instability, such as excessive sensitivity to the choices made when assigning units to a district.

Question 2. *Does the proposed definition of discrete compactness either have a preferred unit of census geography, or can it be shown to be stable under the passage from blocks to block groups to tracts?*

Blocks seem like a good choice of standard—a better choice than tracts or VTDs in this regard—because districts are generally wholly composed of them, leaving fewer choices to the user for how to define and weight district membership.

Finally, discrete compactness scores that do not use population weighting are still subject to empty space effects (Issue D), but this can be cured entirely by use of the weighted scores, which do not "see" unpopulated areas. At first this seems like a powerful commendation for population-weighted scores like the one sketched in this section, but it raises worries that an abusively drawn district can make its gerrymander invisible by adding a buffer of unpopulated units around its border, thereby dropping the weighted perimeter to near zero. This is where ideas to complicate the definition of boundary may come especially in handy.

Question 3. *For the proposed definition of discrete compactness, what features of districts are incentivized? Would adoption of the definition promote fairer districting practices, by the lights of traditional districting principles and other civil rights goals?*

6.3. **Future research.** In the last section we introduced some opening ideas for discrete isoperimetry and toured through some of their features and the questions they raise. We feel that these questions can only be addressed through a mix of theory and data analysis. It is also possible to conceive of graph-based analogs of convex hull and dispersion-based scores, among others. We hope that future research will employ the data collected by the Voting Rights Data Institute and made public at [18] to analyze novel discrete compactness scores—against data drawn from historical and new districting plans—in order to better understand their properties and the extent to which they avoid the problems that contour-based scores have been shown to share.

References

[1] M. Altman, Modeling the effect of mandatory district compactness on partisan gerrymanders, *Political Geography*, Vol. 17, No. 8, 1998, 989-1012.
[2] Barnes and Solomon, *Gerrymandering and Compactness: Implementation Flexibility and Abuse*, `https://arxiv.org/abs/1803.02857`
[3] Bridson and Haefliger, Metric spaces of nonpositive curvature.
[4] Moon Duchin, *Compactness, Then and Now.* From Political Geometry: Interdisciplinary Investigations in Redistricting. Birkhäuser 2018.
[5] Moon Duchin, *Outlier Analysis for Pennsylvania Redistricting.*
[6] Nik Freeman, *Nobody lives here: The nearly 5 million Census Blocks with zero population.* `http://tumblr.mapsbynik.com/post/82791188950/nobody-lives-here-the-nearly-5-million-census`
[7] Lee Hachadoorian et al, (projection paper)
[8] By Justin Kunimune - Own work, CC BY-SA 4.0, `https://commons.wikimedia.org/w/index.php?curid=65635456`
[9] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, 1982.
[10] Dan Margalit and Matt Clay, eds. Office Hours with a Geometric Group Theorist. Princeton University Press, 2017.
[11] Niemi, Grofman, Carlucci, Hofeller, *Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan and Racial Gerrymandering.* The Journal of Politics, Vol 52, No 4 (Nov 1990), 1155–1181.
[12] Polsby, Daniel D and Robert D Popper (1991): *The third criterion: Compactness as a procedural safeguard against partisan gerrymandering.* In: Yale Law & Policy Review, no. 2, vol. 9, pp. 301–353.
[13] Reock, Ernest C (1961): *A note: Measuring compactness as a requirement of legislative apportionment.* In: Midwest Journal of Political Science, no. 1, vol. 5, pp. 70–74.
[14] Schwartzberg, Joseph E, *Reapportionment, gerrymanders, and the notion of "compactness"*. Minn. Law Rev. Vol. 50 (1966), 443–452.
[15] Stein and Shakarchi, Princeton Lectures in Analysis Vol 3: Real Analysis.
[16] K. Stephenson, Circle packing: a mathematical tale, *Notices of the American Mathematical Society*, vol. 50, 1376–1388.
[17] By Strebe - Own work, CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=16115242`
[18] Voting Rights Data Institute, *Discrete Compactness Data Repository*, `http://github.com/gerrymandr`
[19] `https://www.census.gov/geo/maps-data/data/tiger-line.html`
[20] `https://www.census.gov/geo/maps-data/data/cbf/cbf_description.html`
[21] `https://www.census.gov/geo/reference/state-area.html`
[22] `https://www2.census.gov/geo/pdfs/reference/GARM/Ch11GARM.pdf`

Department of Mathematics, Tufts University, Medford MA 02155
E-mail address: `moon.duchin@tufts.edu`

Department of Mathematical Sciences, DePaul University, Chicago, IL 60614
E-mail address: `bridget@math.depaul.edu`