DISCRETE GEOMETRY FOR ELECTORAL 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. To address these issues, we set the stage for *discrete* versions of classical shape scores, laying out definitions, goals, and questions for a promising new fusion of combinatorics and 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 considerable debate and legal scrutiny.

There are two main principles commonly applied to the shape (that is, the geometric form) of districts: jurisdictions should be cut into pieces that are "contiguous" and that are "compact." Those evaluating a plan, whether during its construction and approval or during subsequent challenges, need tools for assessment in the context of these two aims, among others.

The first of these guidelines, *contiguity*, refers to topological connectedness: a district should not be made up of multiple connected components. This is a widespread and uncontroversial requirement of districting plans.* In contrast with that clarity, *compactness* gestures at the reasonableness of district shapes, and is rarely defined precisely, if at all. Even on such unsteady footing, the notion 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, 5, 12] and references).

The purpose of this paper is to call attention to a shared feature in nearly all of the existing definitions of compactness scores: they represent districts as regions enclosed by contours on

2010 *Mathematics Subject Classification.* Primary: 91D20; Secondary: 05C90, 52C99, 90C35.

*Note that while contiguity is largely unambiguous, there are numerous examples of it being achieved by connective tissue along a highway or through water. For example, Illinois's 4th Congressional District uses a stretch of Interstate 294 to connect its northern and southern components, producing a shape often described as "earmuffs."

†See `http://redistricting.lls.edu/where-state.php#compactness` (accessed August 15, 2018), which describes compactness requirements of some kind in 37 of the 50 states.

a map, then base their numerical scores on lengths and areas from the map. This makes all of the classical definitions of compactness susceptible to a common set of flaws, undermining the extent to which the definitions can be made precise and meaningful. In contrast, the kind of discretization we propose here will make fundamental use of the population network that is at the heart of the redistricting problem. We call for research towards creating a new generation of *discrete* scores and we set some benchmarks and goals in that direction.

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. A mapmaker can tilt outcomes 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 dispersing it, thus diluting those voters' influence. Either strategy can induce distended district shapes in order to achieve its purpose. A second, related argument for shape guidelines 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 represent chunks of territory that have a meaningful cohesion and can be traveled efficiently. A historical overview and taxonomy of compactness definitions can be found in [5].

Many compactness definitions, when they are attempted, rely on the measurement of area and perimeter. For instance, as discussed below, one of the most prevalent definitions of compactness takes the 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. This reality is obscured in the focus on map contours.

2.2. **Districts and their building blocks.** Political jurisdictions—even states themselves—have legal definitions. The usable formats for communicating those 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, the finest units of census geography are called *blocks*, which nest into larger units called *block groups*, which in turn nest into *tracts*. For each category, the geographic units in that category partition the state that they belong to, meaning that the entire territory of the state (land and water) is covered by the census units at that given scale, and furthermore that those units are disjoint from each other, except along their borders. Blocks also nest into *voting tabulation districts* (VTDs), which can be thought of as the Census Bureau's recommended precincts, and are the closest match in census geography for the level at which election results are reported.

Congressional districts form another partition of the state.* They are essentially always made out of whole census blocks, whereas district lines frequently cut across block groups, tracts, and VTDs.† Within a given state, every district must have very nearly the same population, counted by adding the census population of each district's blocks.‡

The Census Bureau releases an updated *vintage* of its most precise shapefiles—so-called *TIGER/Line Shapefiles* [20]—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 [21], which are intended for the purpose of map-drawing rather than definition. These Cartographic maps are prepared for every Congress, at three levels of resolution, discussed further below.

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 in litigation, the *Polsby-Popper score*, and 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 ("iso" = same) perimeter.

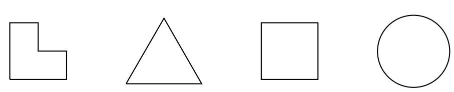


Figure 1. Four regions, each with the same perimeter, are shown from left to right in order of increasing area. The region with largest possible area relative to the fixed perimeter—the circle—is deemed the most "compact" by Polsby-Popper scoring.

*There are a few subtleties in the definition of congressional districts that may make them not quite a well-defined partition; for instance, large bodies of water such as the Great Lakes may or may not be included in any district in a state.

†Only Kentucky has districts that split census blocks, according to the Census Bureau (`https://www.census.gov/rdo/pdf/CD113_BlockSplits.pdf`, accessed August 15, 2018).

‡The issue of how to count population—and which population gets counted—is discussed further on page 10.

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 $0 \leq \frac{4\pi A}{P^2} \leq 1$, 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*

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

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 [16]. Despite the pedigree of this theorem, it was a 1991 article by law scholars Polsby and Popper that led to their names being attached to the associated formula in political science [13].

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 reason that experts frequently apply this score to 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 given by the ratio of the area of Ω to the area of the circle having circumference equal to the perimeter of Ω.*

A cosmetic variant of the Polsby-Popper score, which in fact predates Polsby-Popper in the literature, is the *Schwartzberg score.* This was originally defined as the ratio of the perimeter of a district to the perimeter (circumference) of the circle having the same area, so

*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 Virginia State Board of Elections*, 15 F. Supp. 3d 657 (E.D. Va. June 5, 2015); *Sanders v Dooly County*, 245 F. 3d 1289 (11th Cir. 2001); *Session v Perry*, 298 F. Supp. 2d 451 (E.D. Tex. Jan 6, 2004); *Garza v. County of Los Angeles*, Cal., 756 F. Supp. 1298 (C.D. Cal. 1991); *Harris v McCrory*, 159 F. Supp. 3d 600, 611 (M.D.N.C. 2016); *Johnson v Miller*, 922 F. Supp. 1552 (S.D. Ga. 1995); *Cromartie v Hunt*, 526 U.S. 541 (1999); *Moon v Meadows*, 952 F. Supp. 1141 (E.D. Va. 1997); and many more.

that lower Schwartzberg scores are deemed preferable to higher ones [15]. This is expressed by

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

Since one score is simply the other score raised to a power, it is clear that, although specific numerical values will differ, Schwartzberg and Polsby-Popper assessments must rank districts from best to worst in precisely the same way.* Because Joseph 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 comes from a partial discretization [15]. 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 the scores' monotonic relationship. See [6] for an example of how the resulting scores can differ.

The language used to formulate compactness scores is often imprecise in court documents. This means that what sound like different combinations of area and perimeter frequently 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, 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.** Despite the fact that experts frequently cite Polsby-Popper scores, there is no consensus on how these scores 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."‡

- **Reock**: the area of a district divided by the area of its smallest circumscribing circle [14];
- **Population Polygon**: the population of a district divided by the population contained in its convex hull;§ and

*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.

†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.

‡*Turzai v League of Women Voters of Pennsylvania*, 17A909 (2018).

§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."

- **Minimum Convex Polygon** (also known as the Convex Hull score): the area of a district divided by the area of its convex hull.

The court orders do not specify whether any of these assessments might be more important than the others, nor how two plans are to be compared.

A final notable variant to these scores is to simply report the total perimeter involved in a districting plan. 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. This domain is bounded by a contour, and then classical (Euclidean) geometry is invoked 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. Modern geometry, however, is not limited to the planar domains of Euclid. Since the twentieth century, geometry has flourished in a *discrete*, *combinatorial* setting. The objects in this framework (such as *graphs*, *groups*, and *complexes*) are made up of individual elements that one can enumerate, rather than the smoothly varying quantities of classical geometry. Electoral districts, being made up of census units, therefore lend themselves well to discrete techniques, as we will argue below.

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 problems with the scores that are inextricably tied to the use of Euclidean geometry on map contours. We offer illustrations for each issue in the next section, 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 highly sensitive to their assignment to districts.*

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

4.2. **Discussion and examples.** In the following discussion, individual congressional districts and district data refer to the vintage from the 113th Congress (2013), unless specified otherwise.

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 register when a portion of a district's border may be explained by a pertinent geographical feature like a coastline or an irregular state boundary. In such a situation, the districting plan might incur a steep penalty for having this erratic perimeter, 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. Some shapefiles mitigate the effects from Gulf boundary by extending into the water, but this is problematic in other ways—and of course nothing similar can be done for the river boundary. Accordingly, AL-1 has a fairly low Polsby-Popper score (approximately .162, ranking 318th out of 435) in the TIGER/Line Shapefiles (shown at left in the figure), but scores significantly worse (.111, ranking 367th) in the Cartographic maps (shown at right).* There is no standard on whether to include water when reporting compactness scores.†

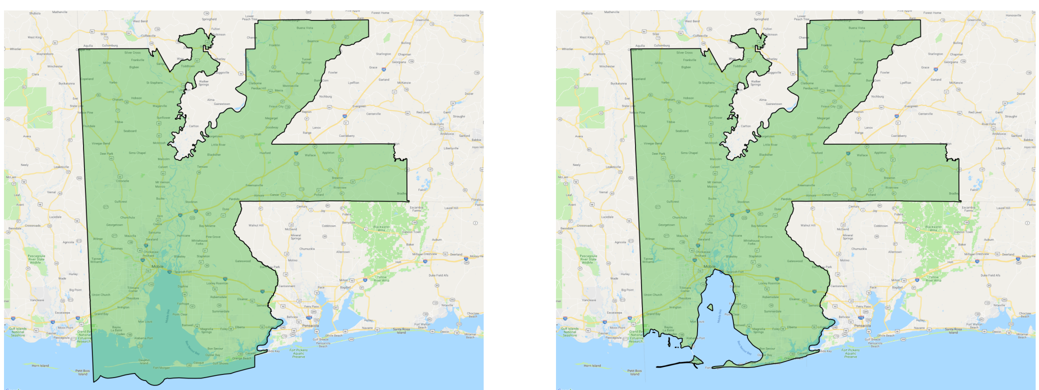


Figure 2. Alabama's 1st district has boundary partly defined by the Gulf of Mexico and the Tombigbee and Alabama rivers. The TIGER/Line (left) and Cartographic 500K (right) maps are shown here, illustrating that the Polsby-Popper quantification of compactness leaves the modeler caught in an unpleasant choice between a map subject to coastline effects (Issue A) or to empty space effects (Issue D).

Consideration of the coastline issue leads naturally to a related worry about stability of scores under 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 one's ruler [11].

*The reader can find code, data, and documentation for area and perimeter statistics at [18].

†For instance, some districting plans filed with the court in Pennsylvania's 2018 redistricting included portions of Lake Erie in the northwest of the state, and others did not.

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 a fixed finite area.

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:

$$\text{500K (1:500,000)}, \quad \text{5M (1:5,000,000)}, \quad \text{20M (1:20,000,000)}.$$

One would expect some variation in the perimeters and areas of districts because the 20M files are greatly simplified. 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" [21]. Nonetheless, we 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 Polsby-Popper score may drop precipitously at higher resolutions.

For example, California's 53rd Congressional District sees an 81% jump in perimeter when going from the 20M scale to the 500K scale. In the same transition, the district's area increases by less than 9%, notable both because this change is nonzero and because it is so different from the perimeter change. This has an enormous effect on its relative ranking among the 435 congressional districts: from ranking 61st at the coarsest zoom, it drops to 191st and then to 292nd at the finest zoom. That means that the district's assessed shape quality goes from being in the best third, to the middle third, to the worst third, inviting completely different qualitative assessments. On average, when comparing data between the 20M scale and the 500K scale, congressional district perimeters increase by about 23%, while 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 explore this phenomenon fully in a 2018 preprint [2], where they do not rely on census cartographic maps but vary map resolution along a spectrum.

Even in practical terms, if we tried to treat census-provided geography as canonical, issues of measurement precision do not disappear. Each year's TIGER/Line file release tends to offer a slightly modified version of state boundaries from those released the year before, sometimes with adjustments on the order of inches from one year to the next, reflecting what the Bureau regards as improved accuracy. Redistricting analysts would need to use maps not only from the same source, but also 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 that it is impossible to choose a map projection which faithfully preserves both shape and area of regions on the sphere. (This is because any smooth map that preserves area and angles must be a local isometry, and so must preserve total curvature.)

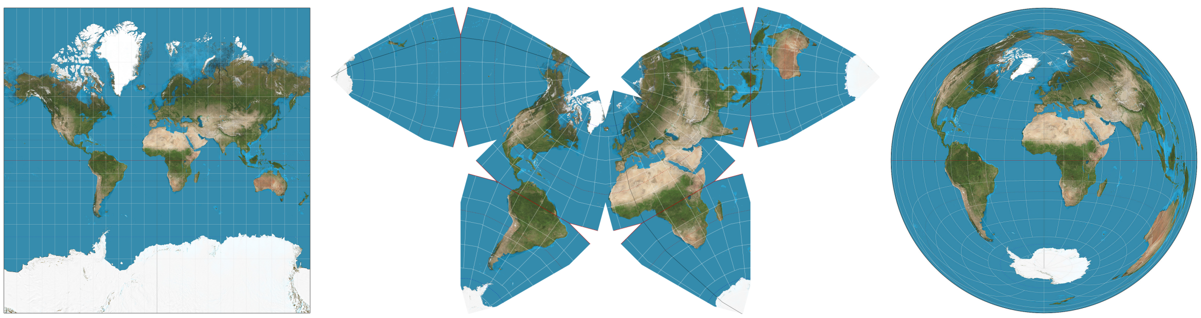


FIGURE 3. Three map projections. The first preserves angle measurements from the sphere; the last preserves area; and the middle map is an attempted compromise between the two, sacrificing some accuracy in each [10, 17].

The Reock score, which is not as badly plagued by the first two issues as Polsby-Popper, emerges as extremely problematic in its coordinate dependence. Hachadoorian et al. [8] show that Reock scores can change dramatically just by choosing a different way to project the earth onto a flat plane. For instance, out of 18 districts that they selected for comparison, 8 had changes of 24% or more in their Reock scores among three projections (locally parametrized Albers equal-area; World Mercator; and plate carrée lat-long). Even worse, the changes were in unpredictable directions, with some scores increasing and some decreasing over a given shift in map projection.

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 well-defined, but even beyond that, area itself 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 highly sensitive to their assignment 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 not have a great 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 [22]. More broadly, in the 2010 census, nearly 45% of census blocks had zero reported population. Indeed, under every contour-based score, compactness can be wildly skewed by assignments of unpopulated surface area to districts.

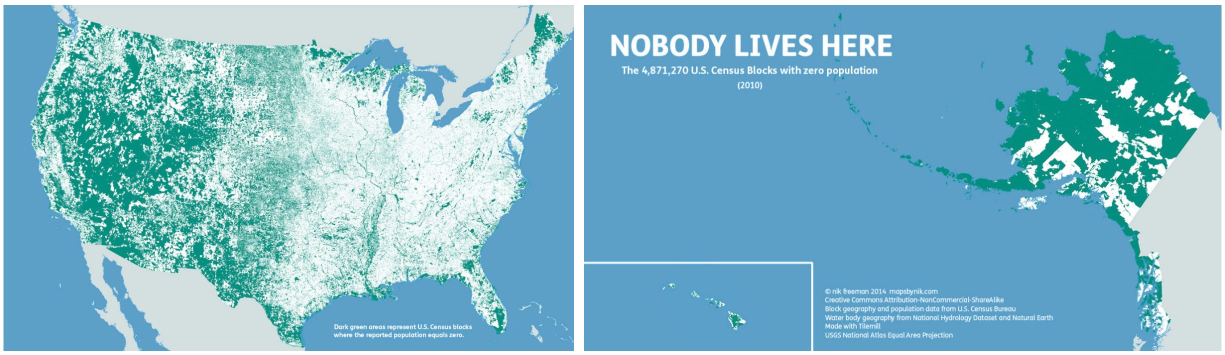


Figure 4. Unpopulated census blocks, excluding the Great Lakes, are depicted in dark green in these maps by Nik Freeman [7].

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 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 in the 2010 Census is about 28. As discussed above, blocks nest inside of other census geographies (block groups, tracts, VTDs, counties) making it possible to calculate the population of these units. For these reasons, census blocks, and the larger elements they define, are the natural units for redistricting.

*The technical definition is as follows: given a topological space $(X, \mathcal{T})$, a subset $S \subseteq X$ is *discrete* if for each element $x \in X$, there exists a neighborhood U_x containing x and no other element of S; that is, $U_x \cap S = \{x\}$. By contrast, consider a disk in the plane: there is no way to isolate a single point from all other points, no matter how closely one zooms in.

†*Reynolds v Sims* (1964) gave the general dictum "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.

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

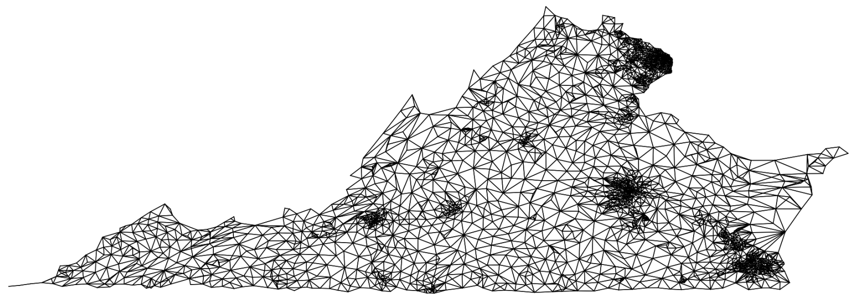


FIGURE 5. The 2372 VTDs in Virginia are shown as the vertices of a graph, with edges indicating geographic adjacencies. One could also make the corresponding graph of the 285,762 census blocks, and it would look substantially more complicated.

We will discuss redistricting-specific ideas in Section 6, but we first develop terminology and notation for graphs.

5.2. **Graph basics.** A graph is defined by a set of vertices, sometimes called nodes, and a set of edges that record adjacencies between vertices. Here we will introduce only as much terminology as is needed for this discussion, and refer the reader to numerous graph theory texts (for example, [4]) for more information. Throughout this work, all graphs will be assumed to be simple and undirected.

A *graph* $G = (V, E)$ consists of a *vertex set* V and an *edge set* E, where each edge is an unordered pair of distinct vertices, and those vertices are said to be *adjacent* to each other. We may use the notation $e = \overline{uv} = \overline{vu}$ to refer to an edge e joining vertices u and v, and we say that u and v are *incident* to e. A graph may be endowed with a *(vertex) weight function*, $w : V \to \mathbb{R}_{\geq 0}$, assigning nonnegative values to the vertices.

An *induced subgraph* of G consists of a subset of the vertices, together with all edges of G that are present among those vertices; that is, vertices in the subgraph are considered to be adjacent if and only if they were adjacent in the original graph G. The induced subgraph in G formed by the vertices $\Omega \subseteq V$ is denoted $G[\Omega]$. The *order* of a graph is the number of its vertices. Thus the order of G is $|V|$ and the order of the induced subgraph $G[\Omega]$ is $|\Omega|$. The *complement* of a subset $\Omega \subseteq V$ is the set $\Omega^c := \{v : v \in V \text{ and } v \notin \Omega\}$. Thus any Ω and its complement form a partition of the vertex set V. The set of edges $\overline{uv}$ for which $u \in \Omega$ and $v \in \Omega^c$ is a *cut-set*, or sometimes a *cut*, of G. Observe that this cut-set is entirely defined by the partition formed by Ω and its complement, and hence entirely defined by Ω.

All of the above terms are standard graph theoretic concepts. We now set out a few specialized definitions that will be helpful below.

Definition 4. *Given a graph $G = (V, E)$ and a subset $\Omega \subseteq V$, the* internal boundary $\partial_0\Omega$ *of the induced subgraph $G[\Omega]$ is the set of vertices in Ω that have neighbors outside of Ω:*

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

Equivalently, the internal boundary $\partial_0\Omega$ is the set of vertices in Ω that are incident to edges in the cut-set of G defined by Ω.

Definition 5. *A* graph with boundary *is a graph $G = (V, E)$ together with a designated subset of vertices $\partial G \subseteq V$ that are designated as* boundary *vertices. The* total boundary *of an induced subgraph $G[\Omega]$ is $\partial\Omega := \partial_0\Omega \cup (\Omega \cap \partial G)$, which is the interior boundary together with any vertices of Ω that belong to the boundary of the ambient graph G.*

Example 6. *Let $G = (V, E)$ be the graph depicted in Figure 6, and define its boundary to be $\partial G := \{v_1, v_2, v_3, v_4, v_5\}$. Consider the subset of vertices $\Omega := \{v_1, v_2, v_3, v_4, v_7, v_9, v_{10}\} \subset V$. In Figure 6, the boundary ∂G is shaded and the induced subgraph $G[\Omega]$ is colored red. The internal boundary of $G[\Omega]$ is*

$$\partial_0\Omega = \{v_1, v_2, v_4, v_7, v_{10}\},$$

because each of those vertices is adjacent to at least one element of $\Omega^c = \{v_5, v_6, v_8\}$, meaning that it is incident to an edge in the cut-set. The total boundary of $G[\Omega]$ is

$$\begin{aligned}\partial\Omega &= \{v_1, v_2, v_4, v_7, v_{10}\} \cup (\Omega \cap \{v_1, v_2, v_3, v_4, v_5\}) \\ &= \{v_1, v_2, v_3, v_4, v_7, v_{10}\}.\end{aligned}$$

Endow G with vertex weights as displayed in the table.

v	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
$w(v)$	2	7	8	1	5	6	0	2	10	2

One easily calculates that $6/7$ of the vertices of Ω are in its total boundary, while the boundary contains only $2/3$ of the vertex weight:

$$\frac{\sum_{v\in\partial\Omega} w(v)}{\sum_{v\in\Omega} w(v)} = \frac{2+7+8+1+0+2}{2+7+8+1+0+10+2} = \frac{20}{30} = \frac{2}{3}.$$

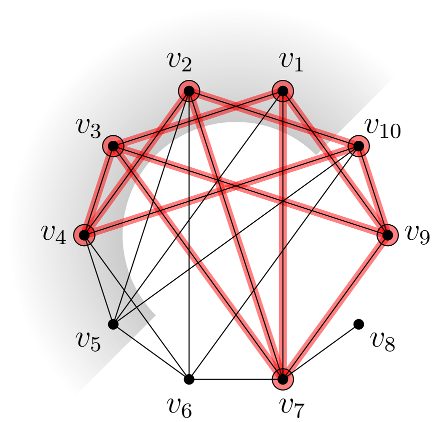


FIGURE 6. The graph G from Example 6. The boundary vertices of the graph, ∂G, are shaded grey, the vertices in Ω are circled, and the induced subgraph $G[\Omega]$ is colored red.

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 (L) of a loop in the edges of the model space to the number (N) of cells in a *filling* (the faces enclosed by the loop). In the context of our previous discussion, these values L and N present abstract analogs of perimeter and area, respectively. Figure 7 shows four examples.

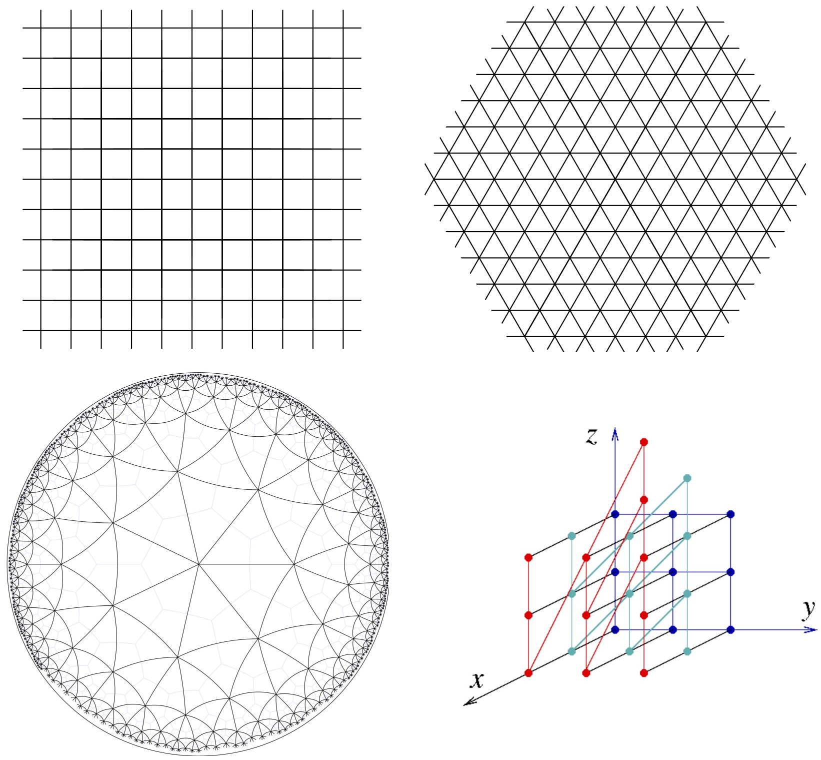


FIGURE 7. It is instructive to consider edge loops in the graphs depicted here and to compare their length, L, to the number of cells they can enclose, N. Any loop in the square grid satisfies the inequality $N \leq \frac{1}{16}L^2$, while the triangular grid has $N \leq \frac{1}{6}L^2$. The hyperbolic triangulation at bottom left has $N \leq L$ 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 $N \leq L^3$.

Filling inequalities of the kind discussed in the caption of Figure 7, called *Dehn functions* in the setting of groups, are discrete versions of the isoperimetric ratio. These have been studied since the pioneering work of Dehn in the early twentieth century. For an introductory-level treatment, see [3, Ch. 8 and 9].

The inequalities depicted in Figure 7 can easily be rewritten to relate the order of an induced subgraph to the order of its boundary. The resulting bounds may differ in their coefficients, but will have the same exponents.* That is, for any subgraph Ω whose boundary

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

$\partial\Omega$ lies on a loop, the inequalities relating N and L can be converted to inequalities relating $|\Omega|$ and $|\partial\Omega|$ that are all of the form

$$|\Omega| \le c \cdot |\partial\Omega|^k.$$

This is just as in the classical Isoperimetric Theorem, in which $k = 2$ and $c = 1/(4\pi)$. This motivates our move below to view the discrete area of a district in terms of the vertices of its defining graph, and the discrete perimeter in terms of boundary vertices.

6. Discretizing compactness

We will lay out some direction below for how to import isoperimetric ideas from discrete geometry into the study of electoral geography. In doing so, we will flag the decision junctures at which the modeler—who may be working as part of the creation and approval process for new districts, or may be working to assess a districting plan after the fact—is making choices that might impact the qualitative features of the scores.

6.1. **Creating the census data graph.** We now have notation and terminology needed to discuss the representation of a state (or other jurisdiction) as a vertex-weighted graph. We record one vertex for each unit of census geography (such as blocks, block groups, tracts, VTDs, or counties). An edge is added between two vertices when the corresponding units share part of their boundary, and the modeler can choose whether to use so-called *rook adjacency* or *queen adjacency*, where the names are drawn from the movement of the corresponding chess pieces. This is a standard construction called the *dual graph* of a partition of the plane, and Figure 8 gives examples.

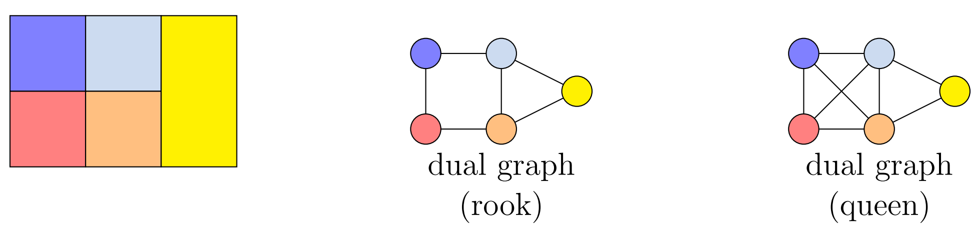


Figure 8. On the left is a partition of a region into five units. The middle and righthand figures represent dual graphs of this partition, where the middle figure has used rook adjacency and the righthand figure uses queen adjacency.

The weight of the vertices in the dual graph G is given by the census population of the corresponding unit. The boundary ∂G of the graph is the subset of vertices whose units are on the outer boundary of the state. Here, the modeler might choose to separately mark vertices according to whether they border water or other significant topographical features, in order to engineer definitions that most effectively counteract coastline-type effects. (In that case, one might modify the definition of $\partial\Omega$ accordingly.)

The goal of this paper is to give new tools for quantitative assessment of the compactness of districts, and so we will consider the induced subgraph that corresponds to a district. As we have seen, census block inclusion in districts is typically all or nothing, but districts frequently split the larger census units. Therefore, if bigger units are being used, then an allocation system is needed in order to decide which units belong to which districts. For example, the modeler might assign each unit to the district in which the largest part of its land area lies, or the largest portion of its population. Alternatively, there could be an area threshold for membership, such as including a vertex in the induced subgraph for a district if, say, 10% or 50% of its area lies in the district; taking this logic a step further, a vertex for a split unit could be included with fractional weight. We note that for any new definition of compactness, the impact of making different choices at this stage will have to be studied.

6.2. **Ideas and questions for discrete compactness.** We now discuss ideas for compactness based on dual graphs built from census units. The aim of our work here is to establish a framework for redesigning compactness scores in a discrete setting, to which the data is better suited and in which the issues outlined above have been defused. We will consider these issues in turn, posing challenges for any proposed definitions of discrete compactness scores that we hope will drive future research.

First, note that coordinate data is not recorded in any of the variants of the census data graph discussed above. Thus any compactness score based on this graph—in contrast with all contour-based scores—will automatically be independent of map projection or choice of coordinates. This completely addresses Issue C above. The remaining three issues require more care.

The principle of discrete compactness scores is to use the combinatorial geometry of these graphs. Taking a cue from the filling inequalities in §5.3, we have a clear starting point suggested by the mathematical developments of the last several decades. We start by taking the *discrete area* to be the order (that is, the number of vertices) of the district subgraph and the *discrete perimeter* to be the order of its boundary, possibly choosing to weight both of these calculations by population, as demonstrated in Example 6. This immediately suggests two discrete analogs of the Polsby-Popper score of a district. We can let the compactness be measured by

(1) discrete area divided by the square of discrete perimeter; or
(2) the same calculation, but weighted by population.

These scores are, respectively,

$$\frac{|\Omega|}{|\partial\Omega|^2} \tag{1}$$

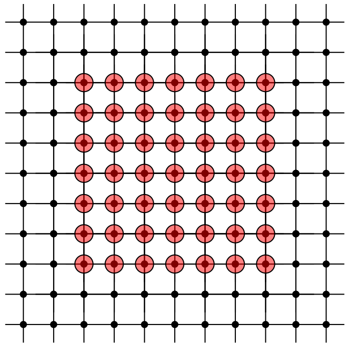
and

$$\frac{\sum\limits_{v\in\Omega} w(v)}{\left(\sum\limits_{v\in\partial\Omega} w(v)\right)^2}. \tag{2}$$

To defend the decision to square the perimeter, consider the lattice examples in Figure 9. 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

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

respectively. These tend to positive, finite limits as n gets large (1/16 for the square case and 1/12 for the hexagon), whereas if any other power of perimeter had been used, the limits would be zero or infinity. 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 of the grid. (Compare this to the other filling exponents found in the non-Euclidean examples from Figure 7.)



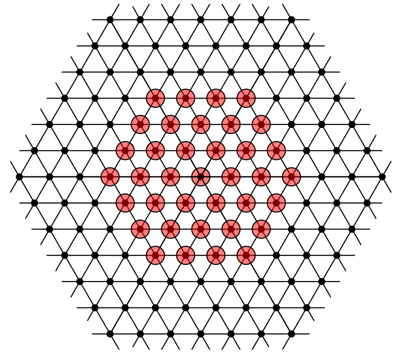


FIGURE 9. Square and triangular lattices with square and hexagonal "districts" Ω_7 and Ω_4', respectively.

We note that there is no need for a coefficient in the discrete calculation 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 consider this to be a feature, not a bug: it reminds the responsible modeler to only compare a compactness score to others that have been collected at the same resolution, which is a good practice, whether scores are contour-based or discrete.

We now turn our attention to Issues A, B, and D from Section 4. 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.

Benchmark 1. *Does the proposed definition of discrete compactness mitigate coastline issues? To measure this, one should examine a state with ocean, lake, or river coast and determine whether the assessment tends to systematically score the coastal districts worse than others, controlling for 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.*

One possible strategy to inoculate discrete isoperimetry scores from many coastline-type effects could be 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 definition of a district's internal boundary might be profitably revised in some more delicate way, such as by declaring a vertex u to belong to the boundary if it is incident to an edge $\overline{uv}$ in the cut-set (that is, $\overline{uv}$ is an edge and $v \in \Omega^c$) such that

$$F(u, v) = \text{TRUE},$$

where F is a particular boolean function on the vertices u and v. For example, one could choose to only include vertices v that represent a census unit with nonzero population (and so F would be "$w(v) > 0$"), or a census unit with population above a certain threshold (making F be "$w(v) > k$" for some k).

*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.

Next, some degree of insulation from resolution instability (Issue B) is guaranteed for discrete compactness scores. This is 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 as in the contour-based setting. 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.

Benchmark 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 transition from blocks to block groups to tracts? To study this stability, one could rank districts by their discrete compactness scores and compare those ranked lists at different resolutions. The lists would surely not be identical at each resolution, but a metric on the space of permutations (such as Kendall's τ metric* [9]*) can assess whether the rankings are close.*

Blocks seem like a good choice of standard—a better choice than tracts or VTDs in this regard—because blocks generally do not straddle multiple districts, leaving fewer choices to make in defining and weighting 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 population-weighted scores, which do not "see" unpopulated areas. At first this seems like a powerful commendation for population-weighted scores like in Expression (2) above. However, this raises the concern 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. To deal with this tradeoff, ideas to refine the definition of boundary—such as by use of the boolean function F described above—may be particularly relevant.

Benchmark 3. *For the proposed definition of discrete compactness, what district features are incentivized? Would adoption of the definition promote fairer districting practices, by the lights of traditional districting principles and other civil rights goals? To study this, one might for instance use algorithmic generation to create large numbers of alternative districting plans for the same jurisdiction and look for common geographical or demographic features of the districts that score high or low.*[*]

6.3. **Future research.** In the previous 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 will also be interesting to push beyond isoperimetry and work to construct discrete (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—applied to data drawn from historical and new districting plans—in order to better understand their properties and the extent to which they avoid the critiques presented here, which are fundamentally applicable to all contour-based scores.

References

[1] M. Altman, Modeling the effect of mandatory district compactness on partisan gerrymanders, *Political Geography* **17** (1998), 989–1012.

[*]Numerous algorithms exist for generating random districting plan. An open-source implementation can be found at [19].

[2] R. Barnes and J. Solomon, Gerrymandering and compactness: implementation flexibility and abuse, preprint, 2018. https://arxiv.org/abs/1803.02857

[3] M. Clay and D. Margalit, eds. *Office Hours with a Geometric Group Theorist.* Princeton University Press, 2017.

[4] R. Diestel, *Graph Theory.* Graduate Texts in Mathematics **173**. Springer, 2005.

[5] M. Duchin, Compactness, then and now. *Political Geometry: Interdisciplinary Investigations in Redistricting* [working title]. Under contract with Birkhäuser.

[6] M. Duchin, Outlier analysis for Pennsylvania redistricting. arXiv, forthcoming.

[7] N. 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. Accessed August 15, 2018.

[8] L. Hachadoorian et al, Comparison of Reock scores across map projections. https://github.com/gerrymandr/compactr. Accessed August 15, 2018.

[9] M. Kendall, A new measure of rank correlation, *Biometrika* **30** (1938), 81–89.

[10] J. Kunimune, Own work, CC BY-SA 4.0, 18 January 2018. https://commons.wikimedia.org/w/index.php?curid=65635456. Accessed August 15, 2018.

[11] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, 1982.

[12] R. G. Niemi, B. Grofman, C. Carlucci, T. Hofeller, *Measuring compactness and the role of a compactness standard in a test for partisan and racial gerrymandering. The Journal of Politics* **52** (1990), 1155–1181.

[13] D. D. Polsby and R. D. Popper, The third criterion: compactness as a procedural safeguard against partisan gerrymandering. *Yale Law & Policy Review* **9** (1991), 301–353.

[14] E. C. Reock, A note: measuring compactness as a requirement of legislative apportionment. *Midwest Journal of Political Science* **5** (1961), 70–74.

[15] J. E. Schwartzberg, Reapportionment, gerrymanders, and the notion of "compactness." *Minnesota Law Reviwe* **59** (1966), 443–452.

[16] E. M. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces.* Princeton Lectures in Analysis **3**. Princeton University Press, 2005.

[17] D. R. Strebe, Own work, CC BY-SA 3.0, 15 August 2011. https://commons.wikimedia.org/w/index.php?curid=16115242. Accessed August 15, 2018.

[18] Voting Rights Data Institute, *Discrete Compactness Data Repository.* http://github.com/gerrymandr/DiscreteCompactness. Accessed August 15, 2018.

[19] Voting Rights Data Institute, *Markov chain Monte Carlo Redistricter.* http://github.com/gerrymandr/RunDMCMC. Accessed August 15, 2018.

[20] https://www.census.gov/geo/maps-data/data/tiger-line.html. Accessed August 15, 2018.

[21] https://www.census.gov/geo/maps-data/data/cbf/cbf_description.html. Accessed August 15, 2018.

[22] https://www.census.gov/geo/reference/state-area.html. Accessed August 15, 2018.

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