SIAM Review
Vol. 66, No. 3, pp. 481–500

Persistent Homology for Resource Coverage: A Case Study of Access to Polling Sites*

Abigail Hickok†
Benjamin Jarman‡
Michael Johnson‡
Jiajie Luo‡
Mason A. Porter§

Abstract. It is important to choose the geographical distributions of public resources in a fair and equitable manner. However, it is complicated to quantify the equity of such a distribution; important factors include distances to resource sites, availability of transportation, and ease of travel. We use persistent homology, which is a tool from topological data analysis, to study the availability and coverage of polling sites. The information from persistent homology allows us to infer holes in a distribution of polling sites. We analyze and compare the coverage of polling sites in Los Angeles County and five cities (Atlanta, Chicago, Jacksonville, New York City, and Salt Lake City), and we conclude that computation of persistent homology appears to be a reasonable approach to analyzing resource coverage.

Key words. persistent homology, topological data analysis, resource coverage, voting access

MSC codes. 55N31, 91D20, 91B18

DOI. 10.1137/22M150410X

1. Introduction.

The geographical distribution of resources such as polling sites (i.e., locations where people vote), hospitals, COVID-19 vaccination sites, Department of Motor Vehicles (DMV) locations, and Planned Parenthood clinics is a major factor in the equitability of access to those resources. Consequently, given the locations of a set of resource sites, it is important to quantify their geographical coverage and to identify underserved geographical regions (i.e., "holes in coverage").

*Received by the editors June 21, 2022; accepted for publication (in revised form) January 5, 2024; published electronically August 8, 2024. AH, BJ, MJ, and JL contributed equally.
https://doi.org/10.1137/22M150410X
Funding: The work of the first author was supported by a Eugene V. Cota-Robles fellowship. The work of the second author was supported by National Science Foundation grant 2011140. The work of the fourth author was supported by National Science Foundation grant 1829071. The work of the fourth and fifth authors was supported by National Science Foundation grant 1922952 through the Algorithms for Threat Detection (ATD) program.

†Department of Mathematics, Columbia University, New York, NY 10027 USA (ah3966@columbia.edu). AH worked on this project predominantly as a member of the Department of Mathematics at the University of California, Los Angeles.
‡Department of Mathematics, University of California, Los Angeles, CA 90095 USA (bjarman@math.ucla.edu, mcjcard@math.ucla.edu, jerryluo8@math.ucla.edu).
§Department of Mathematics, University of California, Los Angeles, CA 90095 USA; Department of Sociology, University of California, Los Angeles, CA 90095 USA; Santa Fe Institute, Santa Fe, NM 87501 USA (mason@math.ucla.edu).

A naive approach to quantifying resource coverage is to consider the geographical distances from resource sites by simply calculating the percentage of people who reside within some cutoff distance D of the nearest resource site. This naive approach is common in policy. For example, in March 2021, United States President Joseph Biden announced a goal to ensure that at least 90% of the adult U.S. population is within 5 miles (i.e., $D = 5$ miles) of a COVID-19 vaccination site [38]. As another example, it is required by Indian law that 100% of voters live within 2 km of a polling site [35] (i.e., $D = 2$ km). However, such an approach poses at least two issues:

(1) it requires choosing an arbitrary cutoff distance D; and
(2) using only geographical distance fails to account for many other factors, such as population density and the availability (and facility) of public transportation, that affect the ease of access to a resource.

These issues severely limit the utility of this naive approach.

In the present paper, we use *topological data analysis* (TDA) to study holes in resource coverage. One of the main tools in TDA is *persistent homology* (PH), which uses ideas from algebraic topology to (1) identify clusters and holes in a data set and (2) measure their persistences at different scales. We use PH to analyze data in the form of a *point cloud*, which is a finite collection $X = \{x_i\}_{i=1}^n$ of points in a metric space (M, d).[1] In this paper, X is a collection of resource sites, with specified latitudes and longitudes, and $M = \mathbb{R}^2$ with a non-Euclidean distance function d (see section 3). Given a point cloud X and a scale parameter $r > 0$, one can consider the *r-coverage* $C_r := \bigcup_{i=1}^n B(x_i, r)$. As the scale parameter r grows, holes arise and subsequently fill in. PH tracks the formation and disappearance of these holes. When a point cloud is a collection of resource sites, one can interpret holes that persist for a large range of r as holes in coverage. Our TDA approach gives a way to measure and evaluate how equitably a resource is distributed geographically.

Our approach using PH addresses both of the issues (see points (1) and (2)) of the naive approach that we discussed above. First, PH eliminates the need to choose an arbitrary cutoff distance because one can study holes in coverage at all scales. Second, instead of employing geographical distance as our metric, we construct a distance function d that is based on travel times. We also incorporate the waiting time at each resource site by constructing a weighted Vietoris–Rips (VR) filtration (see section 2). In a city with a high population density or a poor transportation system, the time that is spent waiting at or traveling to a resource site can be a much higher barrier to access than geographical distance [20,22]. We estimate waiting times using Global Positioning System (GPS) ping data from mobile phones at the resource sites, and we estimate travel times using street-network data, per capita car-ownership data, and the Google Maps application programming interface (API) [21]. Using these estimates, we construct a weighted VR filtration. We weight vertices by our estimates of waiting times, and we define the distance between two vertices to be the estimated round-trip travel time between them. Because the weighted VR filtration is stable, small errors in our estimates cause only small errors in the resultant PH [2].

In this paper, we examine polling sites as a case study of using PH to study the coverage of resource sites. We restrict our attention to six cities:[2] Atlanta, Chicago,

[1] One can weaken the requirement that d is a metric. In this paper, we use a distance function d that is not technically a metric because it does not satisfy the triangle inequality.

[2] Although we frame our discussion in terms of cities, some organizations instead use counties when considering the coverage of polling sites.

Jacksonville (in Florida), Los Angeles,[3] New York City (NYC), and Salt Lake City. We use these cities in part because data about them (e.g., car-ownership data) is widely available. Additionally, these cities differ considerably in their demographics and infrastructures, and we can thus compare a variety of different types of cities. Atlanta and New York City are both infamous for long waiting times at polling sites, especially in non-White neighborhoods [19,26]. In 2020, some counties in the Atlanta metropolitan area had a mean of 3,600 voters per polling site; the number of polling sites had been cut statewide in Georgia by 10% since 2013 [19]. In New York City, each polling site had a mean of 4,173 voters in 2018. As a comparison, in 2004, Los Angeles County and Chicago had only an estimated 1,300 and 725 voters per polling site, respectively [26]. However, Los Angeles is infamous for its traffic [36], which can affect voters' travel times to polling sites. Los Angeles and Chicago also differ in the quality of their public transportation, which also affects travel times to polling sites. In our investigation, we seek both to compare the coverage of polling sites in our six focal cities and to identify underserved areas within each city.

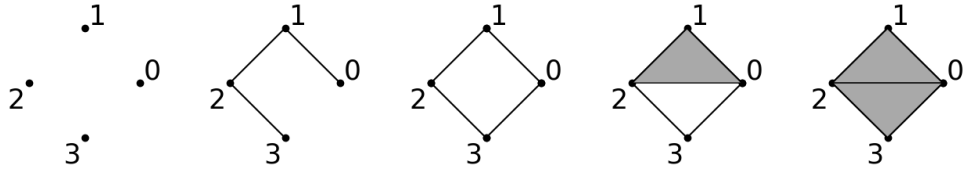
1.1. Related Work. One can use tools from geography to study resource accessibility. Pearce, Witten, and Bartie [34] used a geographical-information-systems (GIS) approach to examine the accessibility of community resources and how it affects health. Hawthorne and Kwan [23] used a GIS approach and a notion of perceived distance to measure healthcare inequality in low-income urban communities. Brabyn and Barnett [6] illustrated that there are regional variations in geographical accessibility to general-practitioner doctors in New Zealand and that these regional variations depend on how one measures accessibility.

Another motivation for our study of resource-site coverage is the related problem of sensor coverage. Given a set S of sensors in a domain $\Omega \subseteq \mathbb{R}^2$, one seeks to determine whether every point in Ω is within sensing range of at least one sensor in S. Typically, each sensor has a fixed, uniform sensing radius r_s. In this case, the problem is equivalent to determining whether or not the domain Ω is covered by balls of radius r_s around each $s \in S$. In [12,13], de Silva and Ghrist gave homological criteria for sensor coverage. Approaches to studying sensor coverage that use computational geometry (specifically, approaches that involve the Voronoi diagram of S and the Delauney triangulation of S) were discussed in [29,30].

Our problem is also a coverage problem, but there are important differences. The key conceptual difference is that we consider neighborhoods whose sizes depend on a filtration parameter, rather than neighborhoods with a fixed size. Additionally, we do not seek to determine whether or not balls of any particular radius cover a domain; instead, our goal is to quantify the coverage at all choices of radius and to determine how the holes in coverage evolve as we increase the filtration parameter. Another difference between the present paper and sensor-coverage problems is that our point cloud represents a set of resource sites (in particular, polling sites), rather than a set of sensors. In a sensor network, pairwise communication between sensors can play a role in whether or not the sensors are fully "connected" to each other (in a graph-theoretic sense) and in determining whether or not a domain is covered [40]. By contrast, communication between resource sites does not play a role in access to those resource sites.

Several studies include applications of PH to geospatial data [10]. Feng and

[3]For Los Angeles, we actually study Los Angeles County. We discuss the reasons for this choice in section 3.5.



(a) $\mathcal{K}_0$ (b) $\mathcal{K}_1$ (c) $\mathcal{K}_2$ (d) $\mathcal{K}_3$ (e) $\mathcal{K}_4$

Fig. 1 *An example of a filtration. The simplicial complex $\mathcal{K}_i$ has the associated filtration-parameter value i. [This figure appeared originally in* [24]*.]*

Porter [18] developed two methods to construct filtrations—one that uses adjacency structures and one that uses the level-set method [32] of front propagation—and applied their approaches to examine geospatial distributions of voting results in the 2016 United States presidential election. They identified "political islands" (i.e., precincts that voted more heavily for a candidate than their surrounding precincts). In [17], Feng and Porter used their approaches to study spatial networks. Stolz, Harrington, and Porter [37] used PH to examine the geospatial distribution of voting results in the "Brexit" referendum. Hickok, Needell, and Porter [24] used PH to study geospatial anomalies in COVID-19 case-rate data (see also [16]) and vaccination-rate data. Corcoran and Jones [10] used PH to perform (1) a point-pattern analysis of pubs across different cities in the United Kingdom (UK) and (2) a spatiotemporal analysis of rainfall in the UK. Kauba and Weighill [27] used PH to examine demographic patterns in the Black and Hispanic populations of 100 U.S. cities.

1.2. Paper Organization. Our paper proceeds as follows. We present background information about PH in section 2, describe our approach in section 3, present and examine persistence diagrams in section 4, and conclude and discuss implications, limitations, and potential future directions of our work in section 5. Our code is available at https://bitbucket.org/jerryluo8/coveragetda/src/main/.

2. Background. We briefly review relevant mathematical background from TDA and PH. See [14,15,33] for more thorough discussions. To compute PH, we begin by constructing a "filtered simplicial complex" (which we will call a "filtration") from a point cloud X. A *simplicial complex* is a combinatorial description of a topological space. It is a collection of vertices, edges, triangles, and higher-dimensional *simplices* with certain requirements on simplex boundaries and pairwise simplex intersections. A *filtration* is a nested sequence $\mathcal{K}_{\alpha_0} \subseteq \mathcal{K}_{\alpha_1} \subseteq \cdots \subseteq \mathcal{K}_{\alpha_n}$ of simplicial complexes, where $\alpha_0 < \alpha_1 < \cdots < \alpha_n$. We show an example of a filtration in Figure 1.

Given a point cloud $X = \{x_1, \ldots, x_n\}$ in a metric space (M, d), there are several ways to construct a filtration that approximates the shape of X. Two of the most common constructions are the Čech filtration and the Vietoris–Rips (VR) filtration [33]. For $r > 0$, the *Čech complex* $\check{\mathrm{C}}_r(X, M, d)$ *at filtration parameter* r is the simplicial complex that has a simplex with vertices $[x_{i_0}, \ldots, x_{i_k}]$ if the intersection $\bigcap_j B(x_{i_j}, r)$ is nonempty, where $B(x, r) := \{y \in M \mid d(x, y) \leq r\}$. That is, $\check{\mathrm{C}}_r(X, M, d)$ is the nerve of the closed balls $\{B(x_i, r)\}_{x_i \in X}$. By the Nerve Theorem [5], $\check{\mathrm{C}}_r(X, M, d)$ is topologically equivalent (more precisely, it is homotopy-equivalent) to the union $\bigcup_i B(x_i, r)$ of balls (i.e., the "r-coverage" of X) in M whenever the balls $B(x_i, r)$

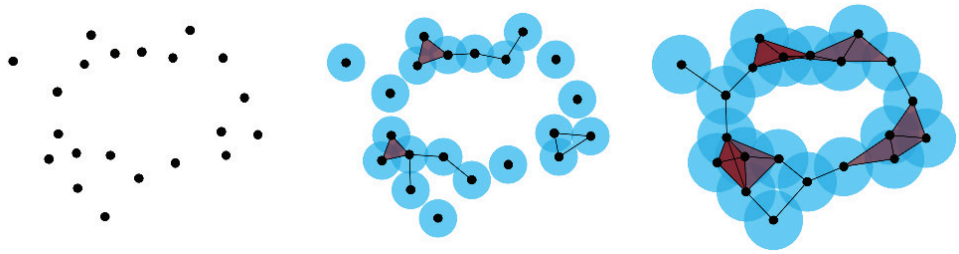


Fig. 2 *Illustration of a Čech filtration for a point cloud X that we sample from an annulus. [We generated this figure using* [1]*.]*

are convex.[4] This implies that $\bigcup_i B(x_i, r)$ and $\check{\mathrm{C}}_r(X, M, d)$ have the same homology (i.e., the same set of holes). A *Čech filtration* is a nested sequence of Čech complexes for increasing filtration parameter r. In Figure 2, we show an example of a Čech filtration.

In practice, it is uncommon to use Čech filtrations because they are difficult to compute. A *Vietoris–Rips (VR) complex* $\mathrm{VR}_r(X, M, d)$ is an approximation of a Čech complex that is faster to compute because it is only necessary to calculate pairwise distances between points. The VR complex at filtration parameter r has a simplex with vertices $[x_{i_0}, \ldots, x_{i_k}]$ if $d(x_{i_j}, x_{i_\ell}) < 2r$ for all j and ℓ. A *VR filtration* is a nested sequence of VR complexes for increasing filtration parameter r. A VR filtration "approximates" a Čech filtration in the sense that

$$\check{\mathrm{C}}_r(X, M, d) \subseteq \mathrm{VR}_r(X, M, d) \subseteq \check{\mathrm{C}}_{\sqrt{2}r}(X, M, d) \tag{2.1}$$

for all r. The complexes $\mathrm{VR}_r(X, M, d)$ and $\check{\mathrm{C}}_r(X, M, d)$ have the same set of edges for all r.

Weighted versions of the Čech and VR filtrations were described in [2]. Given a point cloud $X = \{x_1, \ldots, x_n\}$ in a metric space (M, d) and associated weights $\{w_1, \ldots, w_n\}$, the *radius function* at x_i is

$$r_{x_i}(t) := \begin{cases} -\infty\,, & t < w_i \\ t - w_i\,, & \text{otherwise}\,. \end{cases} \tag{2.2}$$

The closed ball $B(x_i, r_{x_i}(t))$ has no points until time $t = w_i$; at that time, the radius starts growing linearly with t, which is the filtration parameter. The *weighted Čech complex* $\check{\mathrm{C}}_t^{\mathrm{weighted}}(X, M, d, \{w_i\})$ at filtration parameter t is the simplicial complex that has a simplex with vertices $[x_{i_0}, \ldots, x_{i_k}]$ if the intersection $\bigcap_j B(x_{i_j}, r_{x_{i_j}}(t))$ is nonempty. That is, $\check{\mathrm{C}}_t^{\mathrm{weighted}}(X, M, d, \{w_i\})$ is the nerve of $\{B(x_i, r_{x_i}(t))\}_{x_i \in X}$. Like the unweighted Čech complex, the weighted Čech complex is homotopy-equivalent to the union $\bigcup_i B(x_i, r_{x_i}(t))$ of balls by the Nerve Theorem whenever the balls $B(x_i, r_{x_i}(t))$ are convex for all x_i. Much like an unweighted Čech complex, it takes too much time to compute a weighted Čech complex in practice, so researchers usually instead compute a *weighted VR complex* $\mathrm{VR}_t^{\mathrm{weighted}}(X, M, d, \{w_i\})$. This is the

[4]This condition is satisfied for all r when (M, d) is Euclidean, but it is not always satisfied for non-Euclidean metric spaces.

simplicial complex whose vertices are $\{x_i \mid w_i < t\}$ and whose simplices $[x_{i_0}, \ldots, x_{i_k}]$ satisfy $d(x_{i_j}, x_{i_\ell}) + w_{i_j} + w_{i_\ell} < 2t$. The sequence $\{\mathrm{VR}_t^{\mathrm{weighted}}(X, M, d, \{w_i\})\}_t$ for increasing t is the *weighted VR filtration*. Analogously to (2.1), the weighted VR filtration "approximates" the weighted Čech filtration in the sense that

$$\check{\mathrm{C}}_r(X, M, d, \{w_i\}) \subseteq \mathrm{VR}_r(X, M, d, \{w_i\}) \subseteq \check{\mathrm{C}}_{\sqrt{2}r}(X, M, d, \{w_i\}) \tag{2.3}$$

for all r.

Given a filtration $\mathcal{K}_{\alpha_0} \subseteq \cdots \subseteq \mathcal{K}_{\alpha_n}$, one can compute the *homology* of each simplicial complex $\mathcal{K}_{\alpha_i}$. A *homology class* represents a hole that exists in a filtration for some range of filtration-parameter values α_i. A 0-dimensional (0D) homology class represents a connected component, and a 1-dimensional (1D) homology class represents a hole that is bounded by a closed path. To see why 0D homology classes are "holes," we note that one can also view a 0D homology class as representing the empty space between connected components. Therefore, in identifying "holes," it is important to consider both 0D and 1D homology classes.

As the filtration parameter α_i grows, holes form and subsequently fill in. The information that is given by the birth and death of the homology classes of a filtration is called the *persistent homology* (PH) of the filtration. We say that a homology class is *born* at α_i if i is the minimum index such that the homology class appears in K_{α_i}. Its *birth simplex* is the simplex that creates the homology class. For example, in Figure 1, a 1D homology class is born at filtration-parameter value 2. Its birth simplex is the edge with vertices 0 and 3. A homology class that is born at α_i subsequently *dies* at α_j, with $j \geq i$, if j is the minimum index such that the homology class becomes trivial (i.e., the corresponding hole fills in) in K_{α_j}. We refer to α_i as the homology class's *birth value* and to α_j as its *death value*. Its *death simplex* is the simplex that destroys the homology class. For example, the homology class that is born at filtration-parameter value 2 in Figure 1 subsequently dies at filtration-parameter value 4. Its death simplex is the triangle with vertices 0, 2, and 3. In our application to polling sites, we interpret homology classes as holes in coverage and we interpret the death simplices as the locations of the holes.

One can summarize PH in a *persistence diagram* (PD). A PD is a multiset of points in the extended plane $\overline{\mathbb{R}}^2$. For a homology class with birth value b and death value d, the PD includes a point with coordinates (b, d). In Figure 3, we show the PD for the PH of the filtration in Figure 1.

3. Our Construction of Weighted VR Complexes. For each city, we construct a weighted VR filtration in which the point cloud $X = \{x_i\}$ is the set of polling sites in $\mathbb{R}^2$ and the weight w_i of a point x_i is an estimate of the waiting time at the corresponding polling site. Instead of computing a weighted VR filtration with respect to Euclidean distance, we define a distance function that estimates the mean amount of time that it takes to travel to and from a polling site. With respect to this distance function, the union $\bigcup_i B(x_i, r_{x_i}(t))$ (see (2.2)) is the set of points y such that the estimated time for an individual at y to vote (including waiting time and travel time[5] in both directions) is at most t. The weighted Čech complex $\check{\mathrm{C}}_t^{\mathrm{weighted}}(X, \mathbb{R}^2, d, \{w_i\})$ is an approximation of $\bigcup_i B(x_i, r_{x_i}(t))$. When the balls $B(x_i, r_{x_i}(t))$ are convex, the weighted Čech complex is homotopy-equivalent to $\bigcup_i B(x_i, r_{x_i}(t))$, so these two

[5]Incorporating information (such as waiting times) other than travel times is sensible both in principle and in practice. In our computational experiments, using only travel times yields results that differ drastically from those that we present in this paper.

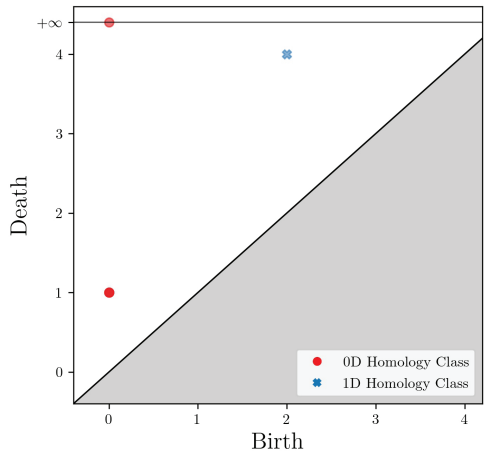


Fig. 3 *The persistence diagram for the* 0*D and* 1*D PH of the filtration in Figure* 1.

complexes have the same homology (i.e., the same set of holes). The weighted VR complex $\mathrm{VR}_t^{\mathrm{weighted}}(X, \mathbb{R}^2, d, \{w_i\})$ is an approximation of the weighted Čech complex.

We construct our distance function as follows. Let x and y be two polling sites. We estimate the expected time for an individual to travel from x to y and back to be

$$\begin{aligned}\tilde{d}(x,y) := C(Z(x)) \min\{t_{\mathrm{car}}(x,y), t_{\mathrm{pub}}(x,y), t_{\mathrm{walk}}(x,y)\} \\ + [1 - C(Z(x))] \min\{t_{\mathrm{pub}}(x,y), t_{\mathrm{walk}}(x,y)\},\end{aligned}$$

where $Z(x)$ is the zip code that includes x (a polling site), $C(Z(x))$ is an estimate of the fraction of voting-age people in $Z(x)$ who can travel by car to a polling site, and $t_{\mathrm{car}}(x,y)$, $t_{\mathrm{pub}}(x,y)$, and $t_{\mathrm{walk}}(x,y)$ are estimates of the expected travel times from x to y and back by car, public transportation, and walking, respectively. We calculate $C(Z(x))$ by dividing an estimate of the number of personal vehicles in $Z(x)$ by an estimate of the voting-age population in $Z(x)$; see section 3.3. We discuss how we calculate t_{car}, t_{pub}, and t_{walk} in section 3.1.

Our definition of $\tilde{d}(x,y)$ captures the cost (in time) to vote. In particular, $\tilde{d}(x,y)$ is an estimate of the mean travel time for an individual who resides in zip code $Z(x)$ to travel from x to y and back. We assume that all individuals choose the fastest mode of transportation that is available to them. Therefore, individuals who can travel by car choose the fastest option between driving, taking public transportation, and walking. Their travel time is $\min\{t_{\mathrm{car}}(x,y), t_{\mathrm{pub}}(x,y), t_{\mathrm{walk}}(x,y)\}$. Likewise, we assume that individuals who do not have access to a car choose the fastest option between taking public transportation and walking. Their travel time is $\min\{t_{\mathrm{pub}}(x,y), t_{\mathrm{walk}}(x,y)\}$. Our estimate of the fraction of a population with access to a car is $C(Z(x))$, so the fraction without a car is $1 - C(Z(x))$. Therefore, $\tilde{d}(x,y)$ is the (estimated) mean time for an individual who resides in zip code $Z(x)$ to travel from x to y and back.

The function $\tilde{d}(x,y)$ is not symmetric (i.e., $\tilde{d}(x,y) \neq \tilde{d}(y,x)$) because $C(Z(x)) \neq C(Z(y))$. However, we need a symmetric function to construct a weighted VR filtration. To construct a symmetric distance function that is based on $\tilde{d}(x,y)$, we define the distance between x and y to be a weighted average of $\tilde{d}(x,y)$ and $\tilde{d}(y,x)$, where we determine the weights from the populations of the zip codes that include x and y.

More precisely, we define the distance between x and y to be

$$d(x,y) := \frac{1}{P}[P_{Z(x)}\tilde{d}(x,y) + P_{Z(y)}\tilde{d}(y,x)]\,, \tag{3.1}$$

where $P_{Z(x)}$ and $P_{Z(y)}$ are the populations of zip codes $Z(x)$ and $Z(y)$, respectively, and $P := P_{Z(x)} + P_{Z(y)}$ is the sum of the populations of $Z(x)$ and $Z(y)$. With respect to this distance function, the ball $B(x,r)$ is the set of points y such that the expected time for an individual to travel back and forth between x and y is at most r, where the individual starts randomly at x or y with probabilities that are weighted by the populations of their associated zip codes. Although our distance function is not technically a metric (because it does not satisfy the triangle inequality), we can still construct a weighted VR filtration using the definition in section 2.

3.1. Estimating Travel Times. To compute our distance function (see (3.1)), we need to estimate the pairwise travel times by car, public transportation, and walking between each pair of polling sites. We measure these times in minutes.

We estimate the time that it takes to walk between each pair of polling sites using street networks, which are available through the OpenStreetMap tool [31], for each of our cities. Using OpenStreetMap, we calculate a shortest path (by geographical distance) between each pair of polling sites. In Figure 4, we show an example of a shortest path between two polling sites in Atlanta.

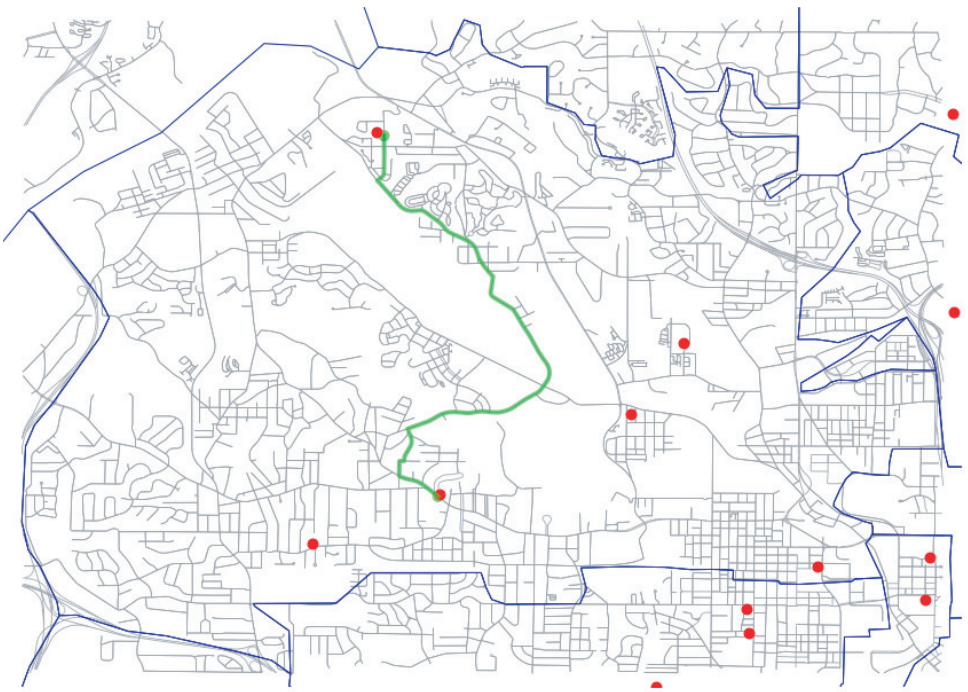


Fig. 4 *A shortest path (by geographical distance) between two polling sites in zip code* 30314 *in Atlanta.*

Let $L(x,y)$ denote the length (which we measure in meters) of a shortest path (by geographical distance) between polling sites x and y. Our estimate of the walking time

(in minutes) from x to y and back is $t_{\text{walk}}(x, y) := 2L(x, y)/v_{\text{walk}}$, where $v_{\text{walk}} = 85.2$ meters per minute is an estimate of the mean walking speed of an adult human [8].

To estimate the travel times by car and public transportation, we use the Google Maps Distance Matrix API [21]. Because of budgetary constraints (and the cost of five dollars per thousand API queries), we use this API to estimate only the travel times between each polling site and its 25 geographically closest polling sites. We refer to these sites as a polling site's 25 nearest neighbors.

For each of the 25 nearest neighbors, we separately calculate both the time from a polling site to each neighbor and the time to a polling site from each neighbor. These two travel times are often different because of different traffic conditions or other factors. We estimate the remaining pairwise travel times as follows. Let G be the unweighted, undirected graph whose vertices are the polling sites and whose edges connect each vertex to its 25 nearest neighbors.[6] Let G_{car} and G_{pub} be the weighted, directed graphs whose vertices and edges[7] are those of G and whose weights are the travel times (by car and public transportation, respectively) that we compute using the Google Maps API. The weight of the directed edge from vertex x to vertex y is the travel time from x to y. Therefore, the weight of the edge from x to y may differ from the weight of the edge from y to x. For any two polling sites x and y, let the travel times $\tilde{t}_{\text{car}}(x, y)$ and $\tilde{t}_{\text{pub}}(x, y)$ be the length of a shortest weighted path from x to y in the graphs G_{car} and G_{pub}, respectively. The corresponding symmetrized travel times $t_{\text{car}}(x, y)$ and $t_{\text{pub}}(x, y)$ are

$$t_{\text{car}}(x, y) := \tilde{t}_{\text{car}}(x, y) + \tilde{t}_{\text{car}}(y, x)\,,$$
$$t_{\text{pub}}(x, y) := \tilde{t}_{\text{pub}}(x, y) + \tilde{t}_{\text{pub}}(y, x)\,.$$

3.2. Estimating Waiting Times. Our weighted VR filtrations have weights at each vertex (i.e., polling site) that are given by an estimate of the mean time that a voter spends (i.e., the mean waiting time) at that polling site. In a nationwide study of waiting times at polling sites during the 2016 U.S. presidential election [9], Chen et al. used smartphone data of hundreds of thousands of voters to estimate waiting times. They also examined potential relationships between waiting times and racial demographics.

We construct our waiting-time estimates using the congressional district-level estimates in [9, Table C.2]. For each polling site x, we calculate the mean of the waiting-time estimates for each congressional district that overlaps with the zip code $Z(x)$ that contains x. This averaging procedure yields estimates of waiting times at the zip-code level. (We transform our waiting-time data to the zip-code level because the rest of our data is at the zip-code level.)

3.3. Estimates of Demographic Information. We obtain estimates of demographic data at the zip-code level from 2019 five-year American Community Survey data [39]. We use voting-age population data from their Table ACSDT5Y2019.B29001 and vehicle-access data from their Table ACSDT5Y2019.B25046.

3.4. Polling-Site Zip Codes. Much of our data is at the zip-code level, and we treat a polling site's zip code as representative of its local area. Certain polling sites (predominantly government buildings) have their own zip codes, despite their

[6] The relation of being one of a vertex's 25 nearest neighbors is not symmetric. Therefore, the degrees of some vertices are larger than 25.

[7] We view each undirected edge (x_i, x_j) of G as a bidirectional edge, and we include both of the associated directed edges in the directed graphs G_{car} and G_{pub}.

populations of 0. We adjust the zip codes of such polling sites to match the zip codes of the directly surrounding areas.

3.5. Special Treatments of Our Cities. The city of Atlanta does not include the suburbs of the Atlanta metropolitan area, so we use the entire area that is served by the Atlanta Regional Commission.

Chicago's boundary is not convex (especially in the northwest), so we include all areas of all zip codes, even when only a small portion of a zip code lies within the city of Chicago.

Because of the oddly shaped city boundaries of Los Angeles, which surrounds several exclaves, we use the entirety of Los Angeles County (except for its islands).

Because of the disconnected nature of New York City, we subdivide it into three regions (Queens and Brooklyn, Manhattan and the Bronx, and Staten Island) and treat each region separately. We then combine our results for the three regions into a single presentation. For example, we combine the PDs into a single PD for all of New York City.

See the file "readme.txt" in our repository https://bitbucket.org/jerryluo8/coveragetda/src/main/ for more information.

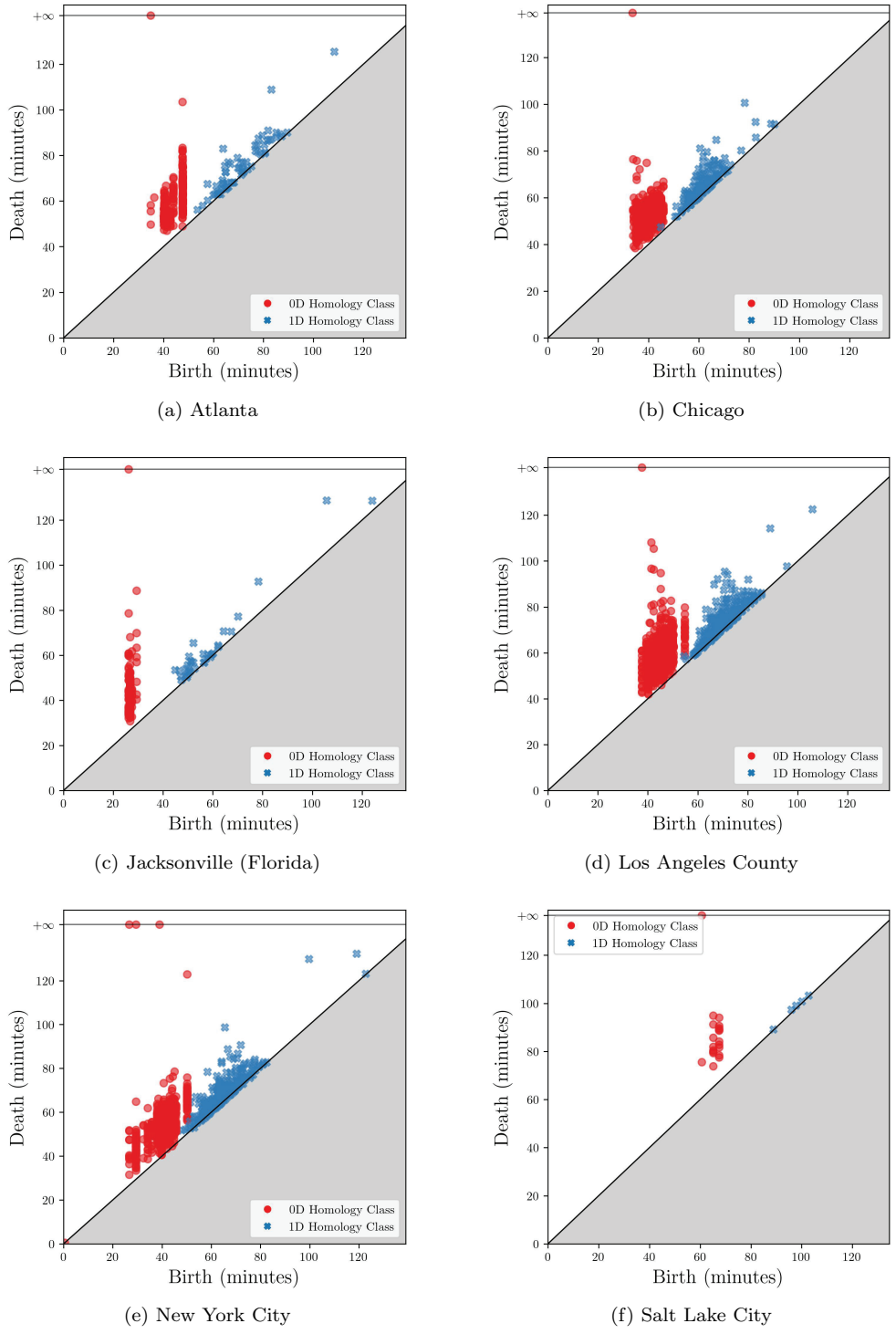
4. Results. We compute the PH of the weighted VR filtrations of section 3 for Atlanta, Chicago, Jacksonville, Los Angeles County, New York City, and Salt Lake City. We show their PDs in Figure 5. We examine 0D and 1D homology classes. The 0D homology classes represent holes between different connected regions of coverage, and the 1D homology classes represent holes in coverage that are bounded by closed paths. A homology class that dies at filtration-parameter value t represents a hole in coverage that persists until time t. An individual who lives in a hole in coverage that dies at time t needs t minutes (including both waiting time at a polling site and travel time back and forth to the site) to cast a vote.

In our analysis, we emphasize homology-class death values. We view homology-class birth values as largely irrelevant to our application. A homology-class birth value indicates the filtration-parameter value at which a coverage hole materializes. We use birth values only in the following way. If the death value divided by the birth value (i.e., the "death/birth ratio") of a homology class is very small (i.e., it is close to 1), then it is possible that this class is merely an artifact of using a VR approximation of a Čech complex. We thus focus on homology classes whose death/birth ratios are at least 1.05.[8] Beyond this, we use only the homology-class death values and death simplices.

Larger homology-class death values suggest that a city may have worse coverage, and a wider distribution of death values suggests that there may be more variation in polling-site accessibility within a city. In Figure 6, we show a box plot of the distribution of homology-class death values for each city. In Table 1, we show the medians and variances of the 0D and 1D homology-class death values for each city.

We compare the coverages of the cities by examining the death values in the PDs. For example, in the PDs for Atlanta and Chicago in Figure 5, we see that Atlanta's homology classes tend to die later than Chicago's homology classes. We also see this in the box plots in Figure 6 and in Figure 7, in which we plot the distributions of death

[8]Interested readers can explore thresholds other than 1.05 using our data, which is available at https://bitbucket.org/jerryluo8/coveragetda/src/main/. We describe the data in detail in the file "readme.txt".



(a) Atlanta (b) Chicago

(c) Jacksonville (Florida) (d) Los Angeles County

(e) New York City (f) Salt Lake City

Fig. 5 *The PDs for each city for the PH of our weighted VR complexes.*

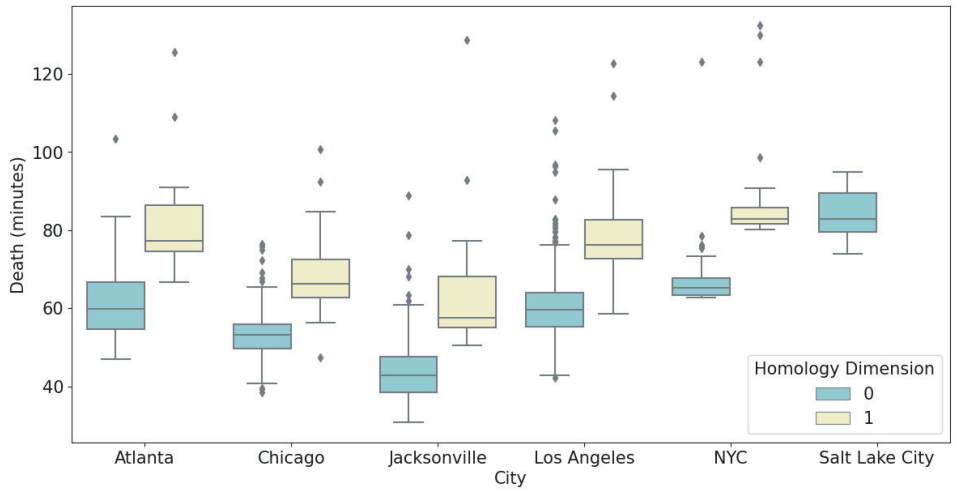


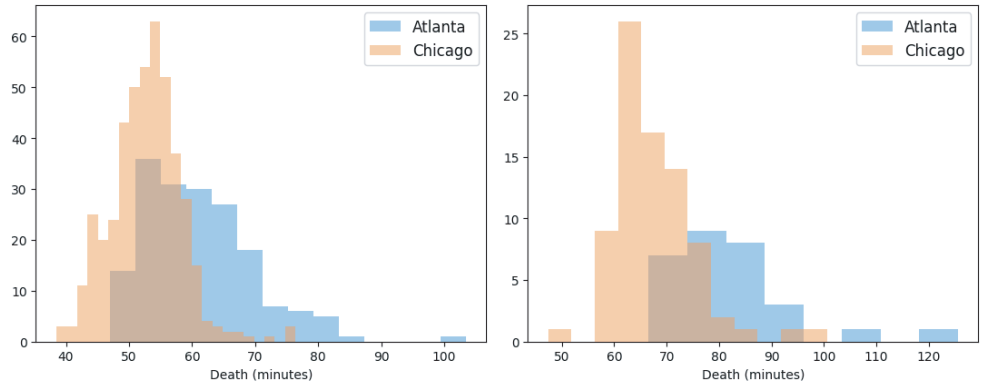
Fig. 6 *Box plots of the death values of the* 0*D and* 1*D homology classes for each city. We only consider homology classes whose death/birth ratio is at least* 1.05*. Salt Lake City has no such* 1*D homology classes.*

Table 1 *The medians and variances of the homology-class death values for each city. (As we discuss in the main text, we consider Los Angeles County rather than only the city of Los Angeles.) We consider homology classes whose death/birth ratio is at least* 1.05*. Salt Lake City has no such* 1*D homology classes.*

City	Homology dimension	Median (minutes)	Variance (minutes)
Atlanta	0	59.9	75.4
	1	77.1	150.8
Chicago	0	53.1	30.2
	1	66.3	59.7
Jacksonville (Florida)	0	42.8	75.7
	1	57.5	394.4
Los Angeles County	0	59.6	53.3
	1	76.1	84.6
New York City	0	65.1	49.2
	1	82.9	207.1
Salt Lake City	0	82.8	37.3
	1	N/A	N/A

values for Atlanta and Chicago. Our PDs and visualizations of summary statistics suggest that Chicago has better polling-site coverage than Atlanta.

We use the death simplices to locate and visualize holes in polling-site coverage. We interpret the death simplex of a homology class as the "epicenter" of an associated coverage hole because the death simplex represents the last part of the hole to be covered. The death simplex of a 0D homology class is an edge between two polling sites; there is a hole in coverage between those two sites. Similarly, the death simplex of a 1D homology class is a triangle that is the convex hull of three polling sites; there is a hole in coverage between those three sites. In Figures 8 and 9, we show the



(a) 0D homology classes (b) 1D homology classes

Fig. 7 *Histograms of the death values of the 0D and 1D homology classes for Atlanta and Chicago. We only consider homology classes whose death/birth ratio is at least 1.05.*

death simplices with the largest death values[9] for the 0D and 1D homology classes,[10] respectively. For example, consider Figure 8(a,b) and Figure 9(a,b), in which we show the 0D and 1D homology-class death simplices for Atlanta and Chicago. The areas of lowest coverage (i.e., the areas that have the death simplices with the largest death values) in Atlanta tend to be in the southwest, whereas the areas of lowest coverage in Chicago tend to be in the northwest and southeast. There is one 1D homology class in Atlanta that has a significantly larger death filtration value than the other classes in Atlanta and any of the classes in Chicago. This homology class represents a hole in coverage in southwest Atlanta (see Figure 9(a)).

5. Conclusions and Discussion.

5.1. Summary. We showed that persistent homology (PH), which is a type of topological data analysis (TDA), is a helpful approach to studying accessibility and equitability. It allows one to examine holes in resource coverage with respect to an appropriate choice of "distance," which one constructs to incorporate important features of a problem of interest. The distance can be based on geography, time, or something else. In the present paper, we used PH to study and quantify holes in polling-site coverage in six U.S. cities (technically, five cities and Los Angeles County). For each city, we constructed a filtration in which a homology class that dies at time t represents a geographical region in which it takes t minutes to cast a vote (including

[9]More precisely, for each city and each homology dimension (0 and 1), we show the death simplices whose death values have a z-score of at least 1. We calculate the z-scores as follows. Let d be the death value of a p-dimensional homology class (where $p = 0$ or $p = 1$) for city C. The z-score of d is $z = (d - \mu_{C,p})/\sigma_{C,p}$, where $\mu_{C,p}$ and $\sigma_{C,p}$ are the mean and standard deviation of the distribution of death values of the p-dimensional homology classes for city C.

[10]In Figure 9, in which we show the death simplices of the 1D homology classes, some of the polling sites appear to be covered by death simplices whose vertices are other polling sites. At least two factors may contribute to this. One factor is that our measure of distance is not a Euclidean metric, even though we plot the death simplices in Figure 9 as Euclidean triangles. The Euclidean triangles can sometimes cover polling sites that are not among its vertices, but geodesic triangles may not cover those polling sites. Another possibility is that a polling site x has such a long waiting time that it does not show up in the filtration until after the homology class whose death simplex includes x has already died.

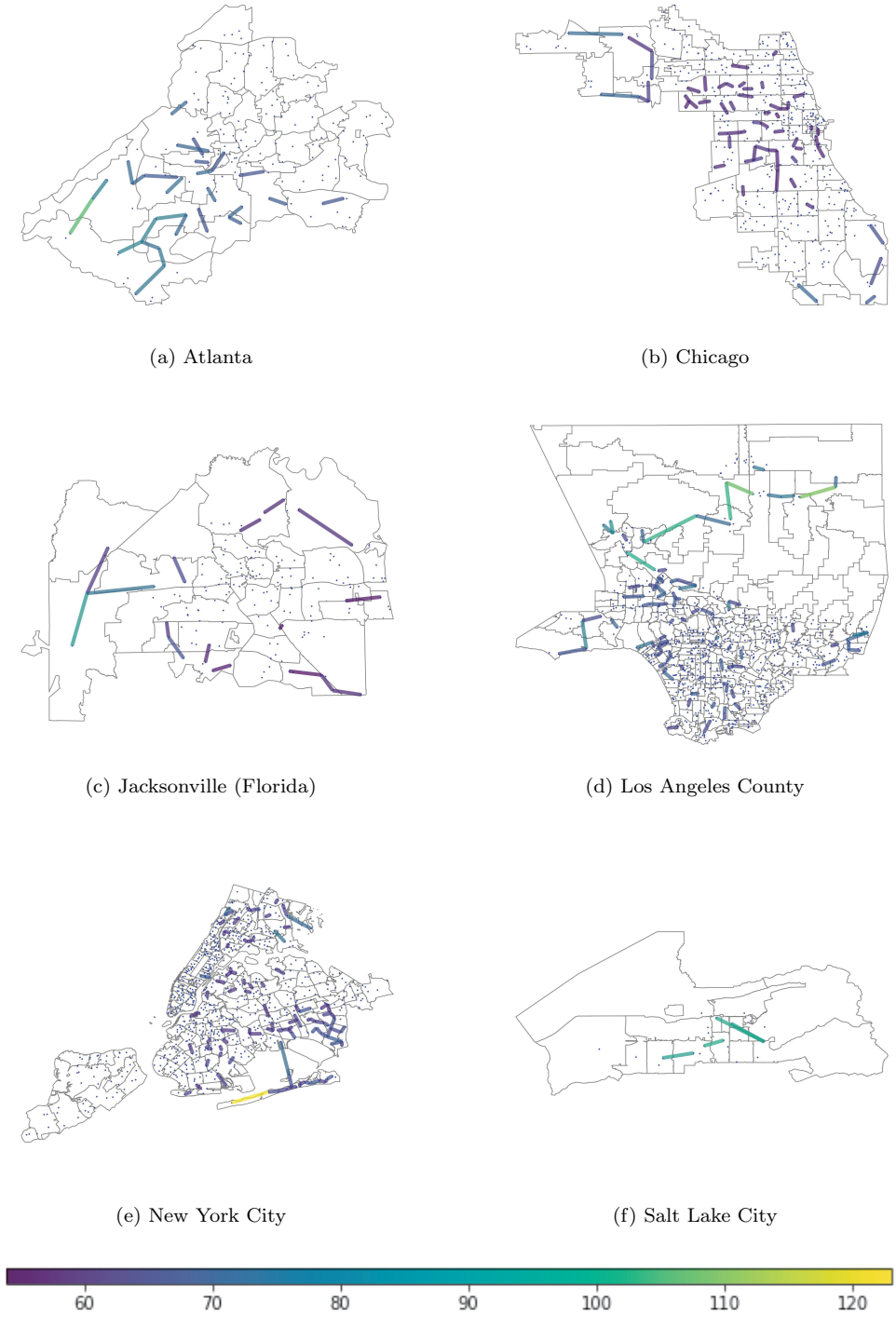


Fig. 8 *Death simplices with the largest death values for the 0D homology classes. The colors correspond to the death values (in minutes). We only consider homology classes whose death/birth ratio is at least* 1.05.

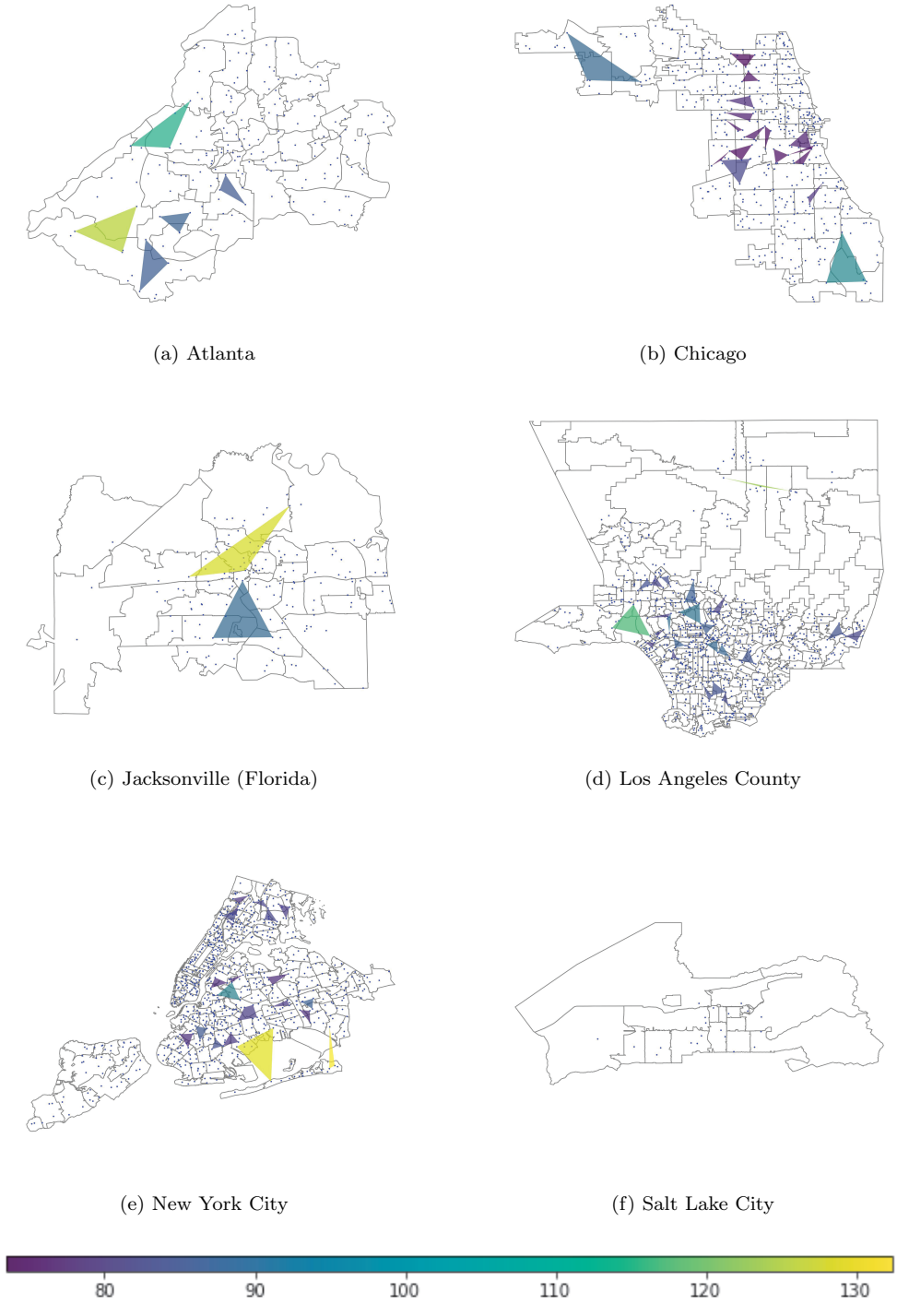


Fig. 9 *Death simplices with the largest death values for the* 1*D homology classes. The colors correspond to the death values (in minutes). We only consider homology classes whose death/birth ratio is at least* 1.05.

both travel time and waiting time). We interpreted the death simplex of a homology class as the location of the corresponding hole in resource coverage. The information in the PH allowed us both to compare the accessibility of voting across our chosen cities and to determine the locations of the coverage holes within each city.

A key benefit of PH is that it enabled us to identify holes in polling-site coverage at all time scales. It also allowed us to use a distance that we designed for the problem at hand, rather than merely using geographical distance, which does not capture important factors in resource accessibility [7]. We based our distance function on estimates of travel time, which is more reasonable and accurate than geographical distance for capturing resource accessibility [34].

5.2. Limitations. To conduct our study, we needed to estimate a variety of quantities (see section 3), including travel times, waiting times, and demographic information. We also made several simplifications due to computational and monetary constraints. We now discuss some issues that are important to address before attempting to incorporate our approach into policy-making.

One limitation of our study is our estimation of travel times. As we discussed in section 3.1, we computed travel times using the Google Maps API. Because of monetary constraints, we only computed a subset of the relevant travel times and used a graph-based estimate to determine the others. Additionally, we computed each travel time between polling sites only once. Computing more precise estimates of travel times is important to better capture the accessibility of polling sites. One way to do this is to compute travel times between the same two polling sites multiple times across different days and times of day and take an average. Such additional computations can also help yield estimates of best-case and worst-case scenarios.

Another limitation of our study is the granularity of our data. As we discussed in section 3.2, our waiting-time data is at the scale of congressional districts. Because there is heterogeneity in the waiting times at different polling sites in the same congressional district, it is important to obtain finer-grained data for the waiting times at polling sites. Having finer-grained waiting times (e.g., if possible, procuring an estimated waiting time for each polling site) would improve our ability to capture voting accessibility.

We also made several topological approximations. We worked with a weighted VR filtration, which approximates a weighted Čech filtration, which in turn approximates the nested set $\{\bigcup_i B(x_i, r_{x_i}(t))\}_{t\in\mathbb{R}}$ of spaces, where $\{x_i\}$ is a set of polling sites and $r_{x_i}(t)$ is the radius function that we defined in section 3. The nested set of spaces is directly relevant to our application, as the holes in $\bigcup_i B(x_i, r_{x_i}(t))$ are the true holes in polling-site coverage. We made our approximations, which are standard in TDA and are well-justified (see our discussion in section 2) [33], to reduce computational cost. However, the convexity condition of the Nerve Theorem, which justifies the approximation of $\bigcup_i B(x_i, r_{x_i}(t))$ by a weighted Čech complex, is not guaranteed to be satisfied for all times t. The Nerve Theorem implies that the weighted Čech complex is homotopy-equivalent to $\bigcup_i B(x_i, r_{x_i}(t))$ whenever the balls $B(x_i, r_{x_i}(t))$ are convex. This condition always holds in Euclidean space, but it is not guaranteed to hold in the space that we defined in section 3.[11] Homotopy-equivalence is important

[11] Although our space is not Euclidean, it is still reasonable to assume that it is approximately *locally* Euclidean. That is, for each polling site x, there is a constant $a > 0$ such that if y is sufficiently close to x, then $d(x, y) \approx a \cdot d_E(x, y)$, where $d(x, y)$ is defined by (3.1) and $d_E(x, y)$ is the Euclidean distance. This approximation holds because car-ownership rates and traffic conditions do not vary much within a sufficiently small neighborhood. We verified empirically that our distance function

because homotopy-equivalent spaces have the same homology and thus have the same set of holes.

Finally, our approach only detects holes in the convex hull of a set of resource sites. Although this may be inconsequential if resource sites are sufficiently spread out geographically, it can be problematic if the resource sites are overly concentrated near a few locations. One way to address this issue is to incorporate city boundaries into the construction of the filtrations. This would help capture holes in coverage in regions that lie outside the convex hull of the resource sites, and it would also help identify the filtration-parameter value t at which an entire city is covered by the balls $B(x_i, r_{x_i}(t))$.

5.3. Future Work. As we discussed in section 5.2, we made several topological approximations of our mathematical object of interest, which is the nested set $\{\bigcup_i B(x_i, r_{x_i}(t))\}_{t\in\mathbb{R}}$ of spaces. Instead of using a weighted VR filtration, one can construct a more direct approximation of $\{\bigcup_i B(x_i, r_{x_i}(t))\}_{t\in\mathbb{R}}$. One can first discretize a city by imposing a grid on it. For each point on such a grid, one can then construct the filtered cubical complex that is induced by the travel time to the nearest polling site. However, this is much more computationally expensive than our approach, and it would also entail many more travel-time queries (which cost money) than in the present paper.[12]

It is also important to incorporate city boundaries into the construction of filtrations. One way to do this is as follows. Let $x_1, \ldots, x_n$ denote the resource sites, and let $y_1, \ldots, y_m$ denote the points that one obtains by discretizing a city boundary. One can extend our distance function (3.1) by defining[13]

$$d(x_i, y_j) := \frac{2}{P}[P_{Z(x_i)}\tilde{d}(x_i, y_j) + P_{Z(y_j)}\tilde{d}(y_j, x_i)]\,, \tag{5.1}$$

where P, P_Z, and $\tilde{d}$ are the same as in the distance function (3.1) and

$$d(y_i, y_j) = \begin{cases} 0\,, & y_i \text{ and } y_j \text{ are adjacent points of the discretized city boundary} \\ \infty\,, & \text{otherwise}\,. \end{cases} \tag{5.2}$$

At each filtration-parameter value, the simplicial complex that one constructs using the distance function (3.1) with the extensions (5.1) and (5.2) includes both the points that one obtains by discretizing the boundary and the edges that connect adjacent boundary points. The largest death value is then the filtration-parameter value t that corresponds to the time at which an entire city is covered by the balls $\{B(x_i, r_{x_i}(t))\}$ (i.e., when there are no longer any holes in coverage).

In our paper, we used death simplices to locate holes in coverage, but other approaches are also possible. For example, by calculating minimal generators [28], one

is approximately locally Euclidean by showing that, for each polling site x, there is a strong linear correlation between the pairwise distances $d(x, y)$ and the pairwise Euclidean distances $d_E(x, y)$ when y is sufficiently close to x. Because our distance function is approximately locally Euclidean, sufficiently small balls (with respect to our distance function) behave like Euclidean balls, so the Nerve Theorem is applicable for sufficiently small filtration values.

[12] Our distance function (3.1) is symmetric, but recall that it is not a metric because it does not satisfy the triangle inequality. Therefore, we cannot use techniques such as distance transforms and level-set propagation to reduce the computational complexity of calculating the filtration $\{\bigcup_i B(x_i, r_{x_i}(t))\}_{t\in\mathbb{R}}$.

[13] The factor of 2 arises from the fact that x_i is a resource site but y_j is not.

can identify representative cycles that encircle holes. The topological pipeline "hyperTDA" was introduced recently [3] to analyze the structure of minimal generators by constructing a hypergraph, calculating hypergraph centrality measures, and employing community detection. This approach may provide insights into the spatial structure of minimal generators. Another potentially viable approach is to use decorated merge trees (DMTs) [11] to locate holes in coverage. DMTs allow one to match holes with associated clusters of points.

Although we have explored a specific case study (namely, the accessibility of polling sites), it is also relevant to conduct similar investigations for other resources, such as public parks, hospitals, vaccine distribution centers, grocery stores, Planned Parenthood clinics, and Department of Motor Vehicles (DMV) locations. One can use similar data to construct a filtration, although it may be necessary to modify the choices of distance and weighting. One can also use ideas from mobility theory [4] to help construct suitable distances and weightings. For example, all DMV offices offer largely the same services, so it seems reasonable to assume that people will go to their nearest office. Therefore, in a study of DMV accessibility, it seems appropriate to use travel time as a distance function, just as we did in our analysis of polling sites. However, in other applications, it is not reasonable to use travel time alone as a distance function. For example, different grocery stores[14] may offer different products at different prices, so travel time alone may not be appropriate as a choice of distance function. Additionally, although waiting time is a significant factor for investigating the coverage of polling sites, there are many applications for which it does not make sense to incorporate waiting time. For example, the time that is spent in a public park or recreation center is typically not a barrier to access. In applications in which waiting times are not an accessibility factor, it seems more appropriate to use a standard VR filtration than a weighted VR filtration. With salient modifications (such as those that we described in this subsection and in section 5.2), we can apply our approach to many other types of resource sites.

Acknowledgments. We thank Chris Anderson, Juan Manuel Balcazar, Lyndie Chiou, Alex Sherman, and Renata Turkes for helpful comments and discussions. We also thank two anonymous referees for helpful comments on an earlier version of this paper.

REFERENCES

[1] H. Adams and J. Segert, *Simplicial Complex Filtration Demonstrations in Mathematica*, 2011; available at https://www.math.colostate.edu/~adams/research/ (accessed 2 May 2022). (Cited on p. 485)

[2] H. Anai, F. Chazal, M. Glisse, Y. Ike, H. Inakoshi, R. Tinarrage, and Y. Umeda, *DTM-based filtrations*, in 35th International Symposium on Computational Geometry (SoCG 2019), G. Barequet and Y. Wang, eds., Leibniz Int. Proc. Inform. 129, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2019, art. 58. (Cited on pp. 482, 485)

[3] A. Barbensi, H. R. Yoon, C. D. Madsen, D. O. Ajayi, M. P. H. Stumpf, and H. A. Harrington, *Hypergraphs for Multiscale Cycles in Structured Data*, https://arxiv.org/abs/2210.07545, 2022. (Cited on p. 498)

[4] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, *Human mobility: Models and applications*, Phys. Rep., 734 (2018), pp. 1–74. (Cited on p. 498)

[14] See [25] for a recent study of grocery-store accessibility.

[5] K. Borsuk, *On the imbedding of systems of compacta in simplicial complexes*, Fund. Math., 35 (1948), pp. 217–234. (Cited on p. 484)
[6] L. Brabyn and R. Barnett, *Population need and geographical access to general practitioners in rural New Zealand*, New Zealand Med. J., 117 (2004), art. U996. (Cited on p. 483)
[7] L. Brabyn and P. Beere, *Population access to hospital emergency departments and the impacts of health reform in New Zealand*, Health Inform. J., 12 (2006), pp. 227–237. (Cited on p. 496)
[8] R. C. Browning, E. A. Baker, J. A. Herron, and R. Kram, *Effects of obesity and sex on the energetic cost and preferred speed of walking*, J. Appl. Physiol., 100 (2006), pp. 390–398. (Cited on p. 489)
[9] M. K. Chen, K. Haggag, D. G. Pope, and R. Rohla, *Racial Disparities in Voting Wait Times: Evidence from Smartphone Data*, Working Paper 26487, National Bureau of Economic Research, 2019; available at https://ssrn.com/abstract=3492890. (Cited on p. 489)
[10] P. Corcoran and C. B. Jones, *Topological data analysis for geographical information science using persistent homology*, Int. J. Geogr. Infor. Sci., 37 (2023), pp. 712–745. (Cited on pp. 483, 484)
[11] J. Curry, H. Hang, W. Mio, T. Needham, and O. B. Okutan, *Decorated merge trees for persistent topology*, J. Appl. Comput. Topol., 6 (2022), pp. 371–428. (Cited on p. 498)
[12] V. de Silva and R. Ghrist, *Coordinate-free coverage in sensor networks with controlled boundaries via homology*, Int. J. Robot. Res., 25 (2006), pp. 1205–1222. (Cited on p. 483)
[13] V. de Silva and R. Ghrist, *Coverage in sensor networks via persistent homology*, Alg. Geom. Topology, 7 (2007), pp. 339–358. (Cited on p. 483)
[14] T. K. Dey and Y. Wang, *Computational Topology for Data Analysis*, Cambridge University Press, Cambridge, UK, 2022. (Cited on p. 484)
[15] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*, American Mathematical Society, Providence, RI, USA, 2010. (Cited on p. 484)
[16] M. Feng, A. Hickok, and M. A. Porter, *Topological data analysis of spatial systems*, in Higher-Order Systems, F. Battiston and G. Petri, eds., Understanding Complex Systems, Springer International Publishing, Cham, Switzerland, 2022, pp. 389–399. (Cited on p. 484)
[17] M. Feng and M. A. Porter, *Spatial applications of topological data analysis: Cities, snowflakes, random structures, and spiders spinning under the influence*, Phys. Rev. Res., 2 (2020), art. 033426. (Cited on p. 484)
[18] M. Feng and M. A. Porter, *Persistent homology of geospatial data: A case study with voting*, SIAM Rev., 63 (2021), pp. 67–99, https://doi.org/10.1137/19M1241519. (Cited on p. 484)
[19] S. Fowler, *Why do nonwhite Georgia voters have to wait in line for hours? Too few polling places*, NPR, 2020; available at https://www.npr.org/2020/10/17/924527679/why-do-nonwhite-georgia-voters-have-to-wait-in-line-for-hours-too-few-polling-pl. (Cited on p. 483)
[20] J. Gimpel and J. Schuknecht, *Political participation and accessibility of the ballot box*, Political Geogr., 22 (2003), pp. 471–488. (Cited on p. 482)
[21] Google Developers, *Distance matrix API*; available at https://developers.google.com/maps/documentation/distance-matrix (accessed 4–7 November 2021). (Cited on pp. 482, 489)
[22] M. Haspel and H. G. Knotts, *Location, location, location: Precinct placement and the costs of voting*, J. Politics, 67 (2005), pp. 560–573. (Cited on p. 482)
[23] T. L. Hawthorne and M.-P. Kwan, *Using GIS and perceived distance to understand the unequal geographies of healthcare in lower-income urban neighbourhoods*, Geogr. J., 178 (2012), pp. 18–30. (Cited on p. 483)
[24] A. Hickok, D. Needell, and M. A. Porter, *Analysis of spatial and spatiotemporal anomalies using persistent homology: Case studies with COVID-19 data*, SIAM J. Math. Data Sci., 4 (2022), pp. 1116–1144, https://doi.org/10.1137/21M1435033. (Cited on p. 484)
[25] D. Horton, T. Logan, D. Skipper, and E. Speakman, *Hundreds of Grocery Outlets Needed across the United States to Achieve Walkable Cities*, https://arxiv.org/abs/2404.01209, 2024. (Cited on p. 498)
[26] H. Kanengiser, *In New York, where you live can determine how hard it is to vote*, City Limits, 2019; available at https://citylimits.org/2019/04/25/nyc-polling-place-shortage-inequality/. (Cited on p. 483)
[27] J. A. Kauba and T. Weighill, *Topological Analysis of U.S. City Demographics*, https://arxiv.org/abs/2310.08334, 2023. (Cited on p. 484)
[28] L. Li, C. Thompson, G. Henselman-Petrusek, C. Giusti, and L. Ziegelmeier, *Minimal cycle representatives in persistent homology using linear programming: An empirical study with user's guide*, Front. Artif. Intel., 4 (2021), art. 681117. (Cited on p. 497)
[29] X.-Y. Li, P.-J. Wan, and O. Frieder, *Coverage in wireless ad hoc sensor networks*, IEEE

Trans. Computers, 52 (2003), pp. 753–763. (Cited on p. 483)
[30] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, *Coverage problems in wireless ad-hoc sensor networks*, in Proceedings IEEE INFOCOM 2001, Conference on Computer Communications, Twentieth Annual Joint Conference of the IEEE Computer and Communications Society, Vol. 3, 2001, pp. 1380–1387. (Cited on p. 483)
[31] OpenStreetMap contributors, *OpenStreetMap*, Version 1.1.1, 2021, https://www.openstreetmap.org (accessed 19–22 January 2022). (Cited on p. 488)
[32] S. J. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, Heidelberg, Germany, 2003. (Cited on p. 484)
[33] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, *A roadmap for the computation of persistent homology*, EPJ Data Sci., 6 (2017), art. 17. (Cited on pp. 484, 496)
[34] J. Pearce, K. Witten, and P. Bartie, *Neighbourhoods and health: A GIS approach to measuring community resource accessibility*, J. Epidemiol. Community Health, 60 (2006), pp. 389–395. (Cited on pp. 483, 496)
[35] S. Scarr, M. Sharma, and M. Hernandez, *Roads, boats and elephants: How India mobilised a million polling stations*, Reuters, 2019; available at https://graphics.reuters.com/INDIA-ELECTION-STATIONS/010092FY33Z/index.html. (Cited on p. 482)
[36] R. Schnalzer, *Traffic is terrible again. Here's how to get it closer to spring* 2020 *levels*, Los Angeles Times, 2021; available at https://www.latimes.com/business/story/2021-07-22/los-angeles-traffic-congestion-commute-pandemic. (Cited on p. 483)
[37] B. J. Stolz, H. A. Harrington, and M. A. Porter, *The Topological "Shape" of Brexit*, https://arxiv.org/abs/1610.00752, 2016. (Cited on p. 484)
[38] The White House, *FACT SHEET: President Biden announces* 90% *of the adult U.S. population will be eligible for vaccination and* 90% *will have a vaccination site within* 5 *miles of home by April* 19; available at https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/29/fact-sheet-president-biden-announces-90-of-the-adult-u-s-population-will-be-eligible-for-vaccination-and-90-will-have-a-vaccination-site-within-5-miles-of-home-by-april-19/, 2021. (Cited on p. 482)
[39] U.S. Census Bureau, *American Community Survey,* 2015–2019 *Estimates*; available at https://data.census.gov/ (accessed 7 November 2021). (Cited on p. 489)
[40] H. Zhang and J. C. Hou, *Maintaining sensing coverage and connectivity in large sensor networks*, Ad Hoc & Sensor Wireless Netw., 1 (2005), pp. 89–124. (Cited on p. 483)