CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/154638/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Corcoran, Padraig and Jones, Christopher 2023. Topological data analysis for geographical information science using persistent homology. International Journal of Geographical Information Science 37 (3) , pp. 712-745. 10.1080/13658816.2022.2155654

Publishers page: https://doi.org/10.1080/13658816.2022.2155654

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Topological data analysis for geographical information science using persistent homology

Padraig Corcoran and Christopher B. Jones

School of Computer Science & Informatics, Cardiff University

ARTICLE HISTORY
Compiled December 1, 2022

ABSTRACT
Topological Data Analysis (TDA) is an emerging field of research which considers the application of topology to data analysis. Recently, these methods have been successfully applied to research problems in the field of Geographical Information Science (GIS) and there is much potential for future applications. In this article, we provide an introduction to the fundamentals of TDA for GIS researchers and practitioners and highlight specific benefits that TDA methods provide relative to some conventional methods. We focus on the method of *persistent homology* which is the most commonly used TDA method. We describe how persistent homology can be applied to data types commonly encountered in the GIScience domain, namely sets of points, networks and sequences of images. We also describe the application of persistent homology to two specific GIS problems, which are the point pattern analysis of UK city pubs and the analysis of UK rainfall radar imagery. In each case we stress the specific benefits of TDA methods that include, for example, generating an output signature in a form that can be subject to subsequent analyses; identification of void regions in point patterns; and providing a relatively simple method to track objects in spatio-temporal images.

KEYWORDS
Geographical Information Science; Topological Data Analysis; Persistent Homology

1. Introduction

Geographical information science (GIS[1]) is a research field concerned with solving geographical or spatial problems. On the other hand, topology is a research field of mathematics that broadly speaking is concerned with modelling properties which are preserved under continuous deformations such as stretching and twisting. Such properties relate to the nature of connectivity and include the properties that one object is contained inside another or that two objects touch. Many useful geographical facts can be modelled using topology. For example, the fact that a school is located in a particular geographical region or the fact that there exists a walking route between someone's home and a shop is naturally modelled using topology. Given this, concepts and techniques from the field of topology are considered fundamental to the field of GIS (Worboys and Duckham 2004).

CONTACT P. Corcoran. Email: corcoranp@cardiff.ac.uk

[1]We use the acronym here in this sense as opposed to geographical information systems to which it commonly also refers.

The importance of topology in spatial information science led to the development of qualitative spatial reasoning (QSR) models, notably the 9-intersection model (Egenhofer and Franzosa 1991) and the region connection calculus (RCC) (Randell *et al.* 1992), that model topological spatial relations between pairs of objects. Such models have been important in supporting the development of GIS applications for identifying and querying topological relations between geometry objects of regions, lines and points. Complementary to developments in QSR, the concept of topology is invoked in GIS in the form of topologically structured data, where the term topology refers to the explicit recording of connectivity between point, line and polygon geometry objects. In contrast to this focus on well defined objects, there has in recent decades been a growing interest in the field of Topological Data Analysis (TDA) that applies principles of topology to the analysis of less precise and often noisy data such as point clouds and network structures. TDA differs from the concepts of topology exploited in QSR models in applying neighbourhood relations of topology that enable connectivity to be defined with respect to distances from points, and hence supports the definition of connected components consisting of points within some distance of each other. This might be regarded as analogous to familiar GIS methods of kernel density estimation and density based clustering, but TDA introduces concepts of the persistence of connectivity across scales (ranges of distance), as well as the explicit representation of voids or holes, which are empty regions of space surrounded by a connected component. This support for analysis of holes in space is a remarkable and potentially very useful aspect of persistent homology methods, with the need for identifying spatial voids occurring in geospatial studies relating for example to cell net coverage (Menon and Joe Prathap 2016), detection of silence (Meyer 2021), and gap analysis in ecology (Jennings 2000). The output of TDA methods, particularly persistent homology, can be treated as a form of signature of a dataset, and is particularly valuable in providing representations that can easily be input to other forms of analysis such as similarity measurement and machine learning classifiers.

Research in the field of TDA has led to the development of several novel and useful methods for data analysis, including persistent homology (Zomorodian and Carlsson 2005) and the mapper algorithm (Singh *et al.* 2007). These methods are general and have been successfully applied to many different types of data in many different research fields. De De Silva and Ghrist (2007) used persistent homology to automatically detect holes in sensor network coverage. Bendich *et al.* (2016) demonstrated that persistent homology applied to the tree structure of blood vessels in the brain can distinguish between the factors of age and sex. Jakubowski *et al.* (2020) applied persistent homology in natural language processing to detect polysemous words, i.e. words with multiple meanings. In the context of social networks analysis, Carstens and Horadam (2013) demonstrated that persistent homology can characterise social dynamics such as collaboration. In an application to shape analysis, Turner *et al.* (2014) used persistent homology to define a measure of shape similarity and used it to determine object similarity. Nicolau *et al.* (2011) used the mapper algorithm to identify a new subgroup of breast cancers with unique properties.

In the field of GIS, TDA methods have also been successfully applied to several research problems. For example, TDA methods have been used to compare the connectivity structures of different street networks (Ahmed *et al.* 2014) and to generalise digital elevation models (Corcoran 2019a). This success can be attributed to the fact that TDA methods have a number of attributes which make them useful for solving practical problems involving real data in the domain of GIS. Notably, TDA methods are robust to noise whereby small changes in the input do not result in a significant

change in the results of the analysis. These methods also offer the ability to perform statistical inferences, such as performing a hypothesis test (Bubenik *et al.* 2015), with respect to topological features exhibited by the data. Like almost all types of data, geographical data can exhibit uncertainty or noise (Longley *et al.* 2015). Therefore, the ability to process such data in a robust manner and to perform statistical inferences is very useful. Finally, as indicated above, TDA methods also facilitate the application of machine learning methods with respect to topological features. Given the ever-growing interest in and capabilities of machine learning methods, this is of particular importance.

To demonstrate the above attributes of TDA methods, let us briefly consider the GIS problem of point pattern analysis which broadly speaking concerns the problem of automatically detecting patterns in sets of points representing some spatial phenomenon. Figure 1 displays two sets of points representing the set of pub locations for the cities of Cardiff and Manchester in the UK. Persistent homology is a method for computing the existence and persistence of connected components and holes across different scales[2]. In the context of this application, connected components correspond to spatially distinct or separated clusters of pubs while holes correspond to void regions where there are no pubs. A visual inspection of Figure 1 reveals that Cardiff city has a number of significant connected components and Manchester has a number of significant holes. Persistent homology can detect the existence of these features across different scales in a robust manner whereby we have some confidence that their detection is the consequence of a significant or meaningful pattern in the data. Furthermore, persistent homology can be used as a platform to perform subsequent statistical inferences and machine learning. For example, we could compute a mean representation which describes the mean pattern of connected components and holes across a number of different cities. We could also perform clustering of cities to determine those cities with similar patterns of connected components and holes. Finally, we could perform a statistical test to determine if the difference in patterns of connected components and holes for the cities of Cardiff and Manchester is statistically significant. As will be discussed later in this article, performing the above types of analysis would be very challenging to do using traditional methods for point pattern analysis such as kernel density estimation (KDE) and density-based spatial clustering.

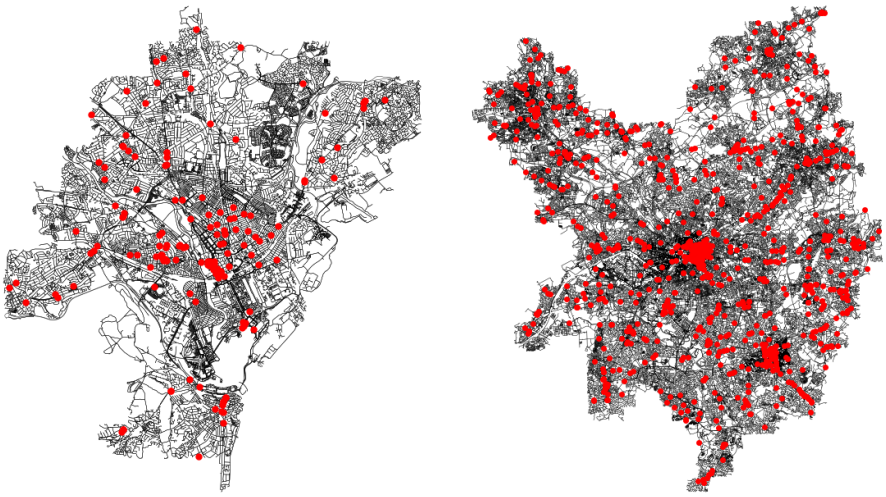
TDA is a relatively new tool in the field of GIS which, as discussed above, has many useful attributes and in turn potential applications. In this article we aim to provide an introduction to the fundamentals of TDA for the GIS community. In doing so we demonstrate how TDA methods provide solutions to a number of fundamental GIS problems, such as point pattern analysis introduced above, where the TDA methods have specific benefits that make them attractive relative to some conventional methods such as point density estimation and clustering.

We address the following three research problems:

(1) For which types of spatial and geographic analysis is TDA applicable and why?
(2) What are the benefits of persistent homology for purposes of point pattern analysis when compared with widely used point density estimation and clustering methods?
(3) What is the benefit of applying persistent homology methods to the spatio-temporal analysis of remote-sensed images?

We also aim to highlight possible future applications of TDA to the field. As an area

[2]The term homology broadly speaking refers to the study of connected components and holes.



(a) (b)

Figure 1. The set of Cardiff and Manchester city pub locations are displayed using red dots in (a) and (b) respectively. In both cases, the city street network in question is also represented in the background to provide context.

of applied mathematics, on some levels TDA can become very technical. Therefore to make this article suitable for a more general audience, we do not venture too deeply into the corresponding underlying mathematics. We instead provide a slightly higher level abstraction of TDA which includes references to more in-depth information. We hope that this approach will have the effect of generating greater interest in TDA from the GIS community, and motivate those in the community to learn more about the field and to gain benefits in applying it to their research problems. At this point, it is worthwhile mentioning that one does not necessarily need to be a mathematician to apply TDA to a given problem. Many TDA methods have existing implementations that can be used like a black box, requiring only an understanding of the inputs and outputs of the method in question. Although there exist a number of other articles which provide an introduction to TDA, none of these works specifically considers the GIS domain. They instead consider alternative domains, such as neuroscience (Sizemore *et al.* 2019), network science (Aktas *et al.* 2019, Serrano *et al.* 2020) and the very general domain of data science (Chazal and Michel 2021). Feng *et al.* (2022) discuss some applications of TDA to spatial systems but do not provide a detailed introduction to the topic.

While TDA is a broad field with many different methods, in this paper we focus almost exclusively on the method of *persistent homology* which is the most commonly used method by a large margin. However, in the conclusion of this article, we highlight some other methods and give corresponding references for the interested reader.

The remainder of this article is structured as follows. In Section 2 we describe the most commonly used workflow for applying persistent homology to a given problem. In doing so, we describe how to apply this workflow to the three types of data. Namely, sets of points, networks and temporal sequences of images. In Section 3 we demonstrate the application of this workflow to two specific GIS problems. Namely, the point pattern analysis of UK city pubs and the analysis of UK rainfall radar imagery. Finally, in Section 4 we draw conclusions and discuss possible directions for future research.

2. The persistent homology workflow

When applying persistent homology to a given problem, the most commonly used workflow contains the following three steps. In the first step of this workflow, the data in question is modelled using a filtration. A filtration is a sequence of triangulations of the data with some additional structure which models connectivity between data elements. In a standard filtration the sequence is nested and hence forms a hierarchy where each (possibly partial) triangulation is associated with a distance-based scale parameter that determines whether neighbouring elements are connected.

In the second step of the workflow, the persistent homology of the filtration is computed. Broadly speaking, this computation models information relating to the existence and persistence across parameter values (representing a measure of distance) of connected components and holes of different dimensions present in the filtration and in turn the data. Thus progression between levels of the filtration is associated with the appearance and disappearance of these features as the distance parameter changes in value. Outputs of the persistent homology computation are a set of persistence diagrams that record the ranges of persistence (existence) of connected components and holes across particular range values of the parameter. As we will see later these methods can be applied such that the persistence is across time ranges rather than scale values.

In the final step of the workflow, analysis, data mining or machine learning is applied to the result of the persistent homology computation. Note that, this step may involve a preprocessing step which converts the persistent homology output into a representation more amenable to subsequent analysis.

In the following subsections we describe each of these three workflow steps in turn. In doing so, we describe how three types of data commonly encountered in GIS can be modelled using a filtration, which in turn allows the workflow to be applied to these types of data.

2.1. *Filtration construction*

A filtration is a sequence of simplicial complexes that equate geometrically to partial or complete triangulations of the data with some additional structure which models connectivity between data elements. In the context of persistent homology, a triangulation corresponds to a simplicial complex model of the data where a simplicial complex is a higher dimensional generalisation of a network. More formally, a simplicial complex $\mathcal{K}$ is a finite set of sets such that for each $\sigma \in \mathcal{K}$ all subsets of σ are also contained in $\mathcal{K}$. An example of a simplicial complex is the set of sets $\{\{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{c, d\}\}$. Each element σ of a given simplicial complex is called a simplex or more specifically a k-simplex where $k = |\sigma| - 1$ is the dimension of the simplex. Thus in the triangulation representation, a vertex is a 0-simplex, an edge is 1-simplex and a face is 2-simplex. A simplicial complex $\mathcal{K}'$ is a subcomplex of simplicial complex $\mathcal{K}$, denoted $\mathcal{K}' \subseteq \mathcal{K}$, if and only if $\mathcal{K}'$ is a subset of $\mathcal{K}$.

A simplicial k-complex $\mathcal{K}$ is a simplicial complex where the largest dimension of any simplex in $\mathcal{K}$ equals k. A simplicial 0-complex is equivalent to a set of points. A simplicial 1-complex is equivalent to a network containing sets of nodes and arcs (or vertices and edges). A simplicial 2-complex is equivalent to a network containing sets of nodes and arcs (or vertices and edges) plus a set of two-dimensional faces.

As discussed above, a filtration is a sequence of simplicial complexes which model a

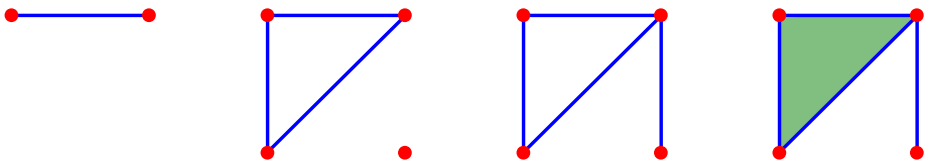


Figure 2. A sequence of four simplicial complexes which form a standard filtration are displayed. In these figures 0-simplices are represented by red circles, 1-simplices are represented by blue lines and 2-simplices are represented by green triangles.

given data with some additional structure. There are a few different structures one can consider. The most appropriate depends on the problem and data in question. Here we describe two filtrations which have distinct structures. We refer to the first filtration as the *standard filtration* since it is the one most used in practice. A standard filtration is a sequence of m simplicial complexes $\mathcal{K}_1, \mathcal{K}_2, \ldots \mathcal{K}_m$, such that each simplicial complex is a subset of the next in the sequence, and hence they satisfy the following condition:

$$\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \cdots \subseteq \mathcal{K}_m \tag{1}$$

Figure 2 illustrates an example standard filtration containing four simplicial complexes. We can see that the simplicial complex in Figure 2(a) is a subset of the simplicial complex in Figure 2(b) and so on.

A standard filtration requires that each simplicial complex is a subset of the next in the sequence. A given sequence of simplicial complexes that are derived from real data, such as a temporal sequence of raster images, may not satisfy this requirement and in such cases, a standard filtration cannot be constructed. A *zig-zag filtration* relaxes this requirement by introducing an intermediate simplicial complex between each pair of consecutive simplicial complexes in the original sequence such that each simplicial complex in the new sequence is a subset of the next **or** previous in the sequence (Carlsson and De Silva 2010). Given a sequence of m simplicial complexes $\mathcal{K}_1, \mathcal{K}_2, \ldots \mathcal{K}_m$, one can construct a zig-zag filtration by constructing the sequence $\mathcal{K}_1, \mathcal{K}_1 \cup \mathcal{K}_2, \mathcal{K}_2, \mathcal{K}_2 \cup \mathcal{K}_3, \ldots, \mathcal{K}_m$. Since a simplicial complex will always be a subset of a simplicial complex formed through the union with another simplicial complex, this sequence satisfies the following subset relations and therefore is a valid zig-zag filtration:

$$\mathcal{K}_1 \subseteq \mathcal{K}_1 \cup \mathcal{K}_2 \supseteq \mathcal{K}_2 \subseteq \mathcal{K}_2 \cup \mathcal{K}_3 \ldots \ \mathcal{K}_m \tag{2}$$

That is, each simplicial complex is a subset of the next or previous in the sequence. Note that, one can also construct an alternative zig-zag filtration by replacing each union operation with an intersection operation and reversing the directions of the subset operations. However, computing the persistent homology of both zig-zag filtrations will return the same result (Carlsson and De Silva 2010).

Figure 3 illustrates an example sequence of four simplicial complexes. This sequence does not satisfy the subset relation necessary to be a standard filtration. For example, the simplicial complex in Figure 3(a) is not a subset of that in Figure 3(b). Figure 4 displays the zig-zag filtration constructed from the above sequence. In this figure the simplicial complex in Figure 4(b) is union of those in Figures 4(a) and 4(c), and the simplicial complex in Figure 4(d) is union of those in Figures 4(c) and 4(e) and so on.

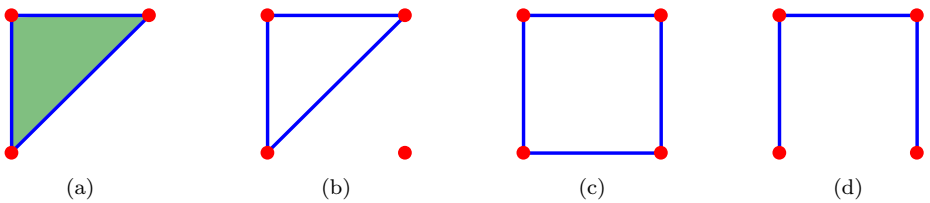


Figure 3. A sequence of four simplicial complexes which do not form a standard filtration are displayed.

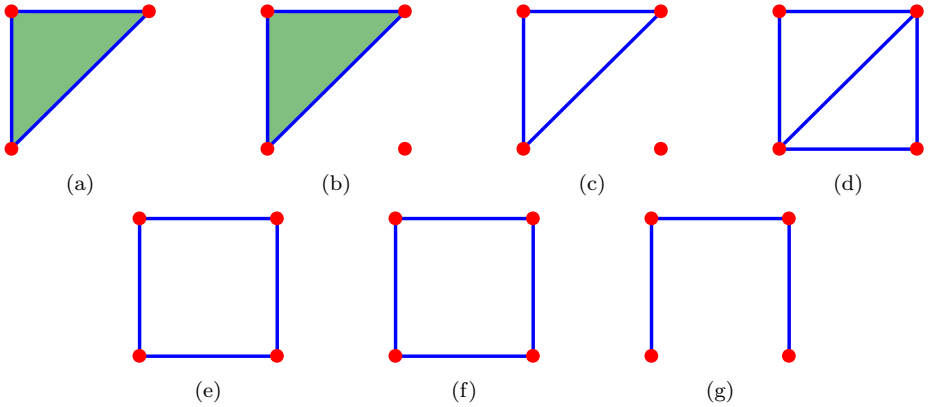


Figure 4. A sequence of seven simplicial complexes which form a zig-zag filtration are displayed.

There exist a number of methods for modelling a given dataset using a filtration. The most appropriate method depends on the data and problem being addressed. In the following subsections we describe methods for modelling a set of points using a standard filtration, modelling a network using a standard filtration and modelling a sequence of images using a zig-zag filtration. Note that, for each type of data there exist potentially many approaches to performing the modelling in question. In this work we only present one possible approach for each data type and this approach may not be the most appropriate for a given problem. To mitigate this lack of coverage, we provide references to relevant works where more information can be found.

2.1.1. Standard filtration of a set of points

In this section we consider the case where the data equals a set of n-dimensional points $S \subseteq \mathbb{R}^n$. Many spatial datasets, such as the locations of facilities or services, are commonly modelled using methods of *spatial point pattern analysis*. There are many such methods, some of which focus on detecting whether a set of points has a non-random clustered data distribution and others based for example on detecting individual clusters or dense regions. The point-based filtration methods that we present here, along with the associated outputs of persistent homology analyses, differ somewhat from other point pattern analysis methods in characterising the structure of patterns with measures of the strength and number of clusters (connected components) as well as detecting other structure in the data, particularly void regions (holes) surrounded by a connected component.

To support the procedure for computing a standard filtration of a set of points we start by introducing the Vietoris-Rips complex (VR complex), where each individual complex is characterised by the fact that its geometric elements are all within some

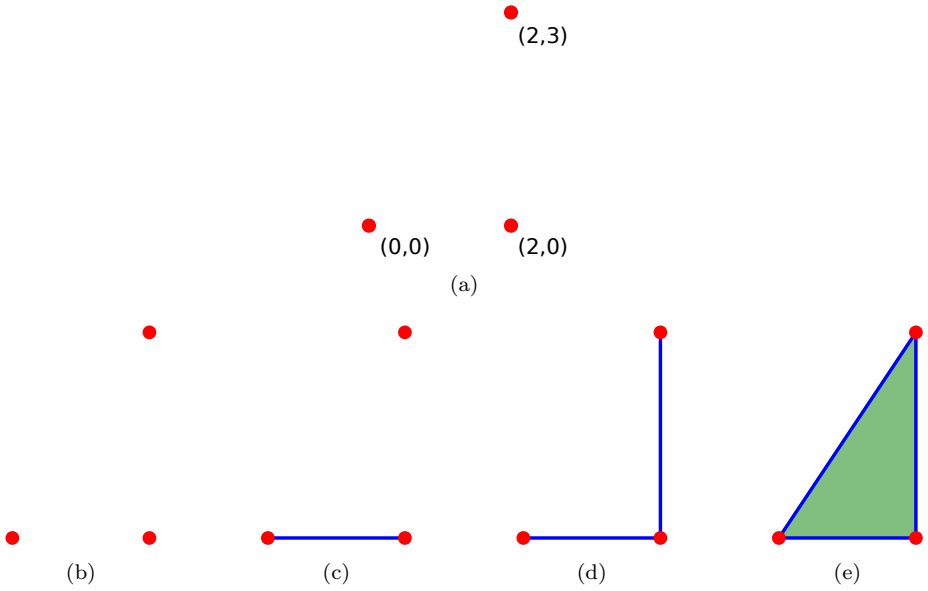


Figure 5. A set of three points $S \subseteq \mathbb{R}^2$ plus corresponding coordinate values are displayed in (a). The VR complexes $V_r(S)$ for r equal to 0.0, 2.0, 3.0 and 3.6 are displayed in (b), (c), (d) and (e) respectively. Note that in the case of (e) for example the longest edge between a pair of points has a length of 3.6 which is its corresponding r value.

specified distance of each other. The VR complex of S is parameterized by $r \in \mathbb{R}$ and is denoted $V_r(S)$ and defined as follows where d is the Euclidean distance metric:

$$V_r(S) = \{\sigma \subseteq S : d(i,j) \leq r, \forall\, i,j \in \sigma\} \tag{3}$$

A subset of k elements in S corresponds to a $k-1$-simplex in $V_r(S)$ if and only if each pair of elements in this subset is less than or equal to r distance apart. To illustrate the VR complex, consider the set $S \subseteq \mathbb{R}^2$ containing three points displayed in Figure 5(a). The corresponding VR complexes $V_r(S)$ for r equal to 0.0, 2.0, 3.0 and 3.6 are displayed in Figures 5(b), 5(c), 5(d) and 5(e) respectively. We can see from these figures that, as the value of r increases, additional simplices are added to the simplicial complex.

We construct a standard filtration by considering the sublevel sets of a VR complex with respect to the parameter r. By continuously increasing the value of this parameter, we get a parametrized family of distinct VR complexes. Each of these is a subcomplex of the VR complex $V_\infty(S)$; that is, the simplicial complex containing the finite set of all subsets of S. Therefore, there exists a finite sequence of m VR complexes $V_{r_1}(S), V_{r_2}(S), \ldots, V_{r_m}(S)$ where $V_{r_i-\varepsilon}(S) \neq V_{r_i}(S)$ for $\varepsilon > 0$. That is, r_i equals the parameter value when $V_{r_{i-1}}(S)$ changes to $V_{r_i}(S)$. This sequence satisfies the following subset relations and therefore is a valid standard filtration.

$$V_{r_1}(S) \subseteq V_{r_2}(S) \subseteq \cdots \subseteq V_{r_m}(S) \tag{4}$$

Given a set of points S, we refer to the above standard filtration as the VR filtration of S. Consider again the set of points $S \subseteq \mathbb{R}^2$ displayed in Figure 5(a). The VR filtration of this set corresponds to the sequence of four VR complexes $V_r(S)$ with r equal to 0.0, 2.0, 3.0 and 3.6. As indicated above, these VR complexes are displayed

in Figures 5(b), 5(c), 5(d) and 5(e) respectively.

Apart from the VR filtration, there exist other methods for modelling a set of n-dimensional points using a standard filtration including the Ĉech complex filtration and the Alpha Complex filtration (Edelsbrunner and Harer 2010). In the Ĉech complex, the elements of an individual simplex are defined as connected on the basis of the intersection of balls of specified radius that surround the points.

2.1.2. Standard filtration of a network

In this section we consider the case where the data takes the form of a network. Many spatial datasets, such as streets and air transportation, are commonly modelled as networks. A network or graph is a tuple (V, E) where V is a set of elements called nodes or vertices and E is a set of pairs of elements of V called edges or arcs. If these pairs are unordered, the network is called undirected. On the other hand, if the pairs are ordered, the network is called directed. If a network $G = (V, E)$ has an associated map $w : V \to \mathbb{R}$ which maps each vertex to a real value, the network is called a vertex-weighted network. If a network $G = (V, E)$ has an associated map $w : E \to \mathbb{R}$, the network is called an edge-weighted network.

There exists a large array of methods for constructing filtrations of networks. A review of these methods can be found in Aktas *et al.* (2019). The most appropriate method to use depends on the context. In this subsection we describe a method for constructing a filtration of an undirected vertex-weighted network and a method for constructing a filtration of an undirected edge-weighted network. In both cases, to construct a filtration we first model the network as a simplicial 1-complex $\mathcal{K}$. This is achieved by modelling each vertex as a corresponding 0-simplex and modelling each edge as a corresponding 1-simplex. For example, consider the network $G = (V, E)$ where $V = \{a, b, c\}$ and $E = \{(a, b), (b, c)\}$. This network is modelled by the simplicial 1-complex $\mathcal{K} = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}\}$. Given $\mathcal{K}$, we next define a map $f : \mathcal{K} \to \mathbb{R}$ which maps each simplex σ to a real value. This map models the significance of each simplex with respect to a given property. For example, consider the case where the network in question represents a street network where vertices correspond to intersections and edges correspond to street segments. In this case, the map f could model the traffic congestion at each vertex and edge.

How best to define the map f will be application dependent. In the case of an edge-weighted network with associated map $w : E \to \mathbb{R}$, the map f can be defined as follows:

$$f(\sigma) = \min\{w(\beta) : |\beta| = 2, \sigma \subset \beta\} \tag{5}$$

This map diffuses values defined on 1-simplices (graph edges) to values defined on 0-simplices (graph vertices). For a given vertex σ, it assigns a weight equal to the minimum weight assigned to an adjacent edge which is represented by β in the equation. The cardinality constraint specifies that β must be a 1-simplex. To illustrate the definition of this map, consider again the network $G = (V, E)$ where $V = \{a, b, c\}$ and $E = \{(a, b), (b, c)\}$. Also consider an associated map $w : E \to \mathbb{R}$ where $w((a, b)) = 1$ and $w((b, c)) = 2$. In this case the result of the map f is that $f(\{a\}) = 1$, $f(\{b\}) = 1$, $f(\{c\}) = 2$, $f(\{a, b\}) = 1$ and $f(\{b, c\}) = 2$. The above approach to defining the map f may be useful in the context of modelling street network traffic congestion where congestion levels are only measured at edges which correspond to street segments.

Given a simplicial 1-complex $\mathcal{K}$ and a map $f : \mathcal{K} \to \mathbb{R}$, we can define a new simplicial

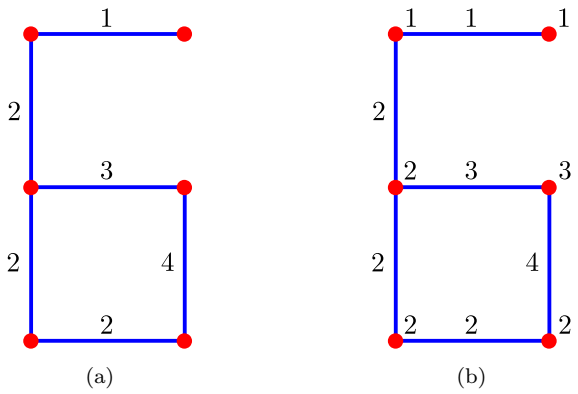


Figure 6. An edge-weighted network is displayed in (a) where the values assigned to each edge by the map w are represented. The simplicial 1-complex corresponding to this network is displayed in (b) where the values assigned to each simplex by the map f are represented.

1-complex $\mathcal{K}_r$ which is parameterized by $r \in \mathbb{R}$ and is a sub-complex of $\mathcal{K}$:

$$\mathcal{K}_r = \{\sigma \in \mathcal{K} : f(\sigma) \leq r\} \tag{6}$$

In a similar vein to the VR filtration described above, we construct a standard filtration by considering the sublevel sets of $\mathcal{K}$ with respect to the parameter r. By continuously increasing the value of this parameter, we get a parametrized family of distinct simplicial 1-complexes. Each of these is a subcomplex of $\mathcal{K}_\infty$ which equals $\mathcal{K}$. Therefore, there exists a finite sequence of m simplicial 1-complexes $\mathcal{K}_{r_1}, \mathcal{K}_{r_2}, \ldots, \mathcal{K}_{r_m}$ where $\mathcal{K}_{r_i-\varepsilon} \neq \mathcal{K}_{r_i}$ for $\varepsilon > 0$. This sequence satisfies the following subset relations and therefore is a valid standard filtration.

$$\mathcal{K}_{r_1} \subseteq \mathcal{K}_{r_2} \subseteq \cdots \subseteq \mathcal{K}_{r_m} \tag{7}$$

To illustrate the above process for constructing a standard filtration of a network, consider the edge-weighted network displayed in Figure 6(a). The simplicial 1-complex corresponding to this network is displayed in Figure 6(b) where the map f that assigns a value to each simplex is defined using Equation 5. The standard filtration corresponding to this simplicial complex equals the sequence of four simplicial complexes $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{K}_3$ and $\mathcal{K}_4$. This sequence is displayed in Figures 7(a), 7(b), 7(c) and 7(d) respectively.

2.1.3. Zig-zag filtration of an image sequence

In this section we consider the case where the data equals a sequence or time series of m images $I_1, I_2, \ldots, I_m$. Spatio-temporal datasets, such as land-use and land-cover classifications, are commonly modelled as a sequence of images. We assume each image I_i in the sequence is an element of the space $\{0,1\}^{p\times q}$. That is, a matrix or grid of size $p \times q$ where each element of the matrix takes a value in the set $\{0, 1\}$.

We model the sequence of images $I_1, I_2, \ldots, I_m$ using a corresponding sequence of simplicial 2-complexes $\mathcal{K}_1, \mathcal{K}_2, \ldots \mathcal{K}_m$. Each image I_i is modelled as a simplicial 2-complex using the following approach which is known as a Freudenthal triangulation. For each cell in I_i we define a corresponding 0-complex in $\mathcal{K}_i$ if the cell has a value equal to 1. For each pair of cells in the I_i which are vertically, horizontally or main

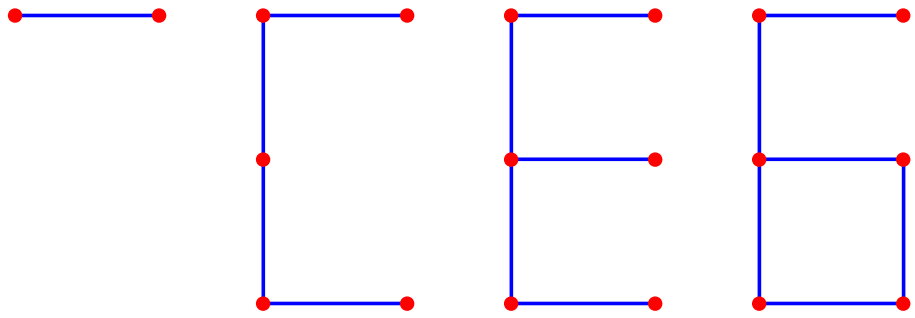


Figure 7. A standard filtration corresponding to the simplicial complex in Figure 6(b) is displayed.

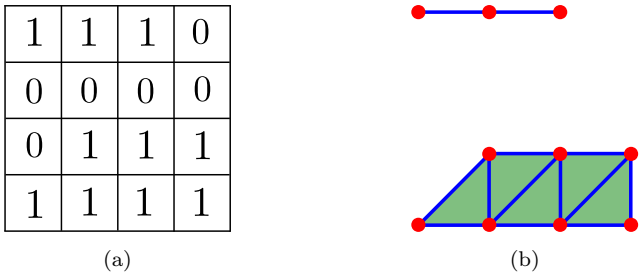


Figure 8. For an image $I_i \in \{0,1\}^{4\times 4}$ displayed in (a), the corresponding simplicial 2-complex $\mathcal{K}_i$ is displayed in (b).

diagonally adjacent, we define a corresponding 1-complex in $\mathcal{K}_i$ if both cells have a value equal to 1. For each triple of cells where all pairs are vertically, horizontally or main diagonally adjacent, we define a corresponding 2-complex in $\mathcal{K}_i$ if all three cells have a value equal to 1. To illustrate this modelling, consider the example image $I_i \in \{0,1\}^{4\times 4}$ displayed in Figure 8(a). The simplicial 2-complex $\mathcal{K}_i$ corresponding to this image is displayed in Figure 8(b).

Given the above sequence of m simplicial complexes $\mathcal{K}_1, \mathcal{K}_2, \ldots \mathcal{K}_m$, we construct a corresponding zig-zag filtration by constructing the sequence $\mathcal{K}_1, \mathcal{K}_1 \cup \mathcal{K}_2, \mathcal{K}_2, \mathcal{K}_2 \cup \mathcal{K}_3, \ldots, \mathcal{K}_m$.

2.2. *Persistent homology computation*

Persistent homology is a method which takes as input a filtration and returns information relating to the existence and persistence of connected components and holes of different dimensions in the filtration. The actual computation varies depending on whether the filtration is a standard (Zomorodian and Carlsson 2005) or a zig-zag filtration (Carlsson *et al.* 2009). In this work we refer to these computations as standard and zig-zag persistent homology respectively. Note that, both computations are quite mathematically technical requiring a working knowledge of algebraic topology to understand. We do not present this material here and instead describe persistent homology in terms of the corresponding method inputs and outputs. An interested reader seeking a more in-depth description should consult the articles referenced above or the textbook by Edelsbrunner and Harer (2010).

Given an input filtration $\mathcal{K}_1, \mathcal{K}_2, \ldots \mathcal{K}_m$, the output from persistent homology is

a set of mathematical objects called *persistence diagrams* that, as indicated earlier, record the ranges of existence (with respect to the distance parameter) of either connected components or of holes depending on the dimension of the diagram. A persistence diagram is a multiset of elements in the space $\{(p, q) \in \mathbb{R}^2 |\ p < q\}$ where p and q are values of the distance parameter (Edelsbrunner and Harer 2010). Each persistence diagram has an associated dimension indicating the specific information it models. An element (p, q) in the zero-dimensional persistence diagram models that a connected component appeared and subsequently disappeared in simplicial complexes $\mathcal{K}_p$ and $\mathcal{K}_q$ respectively. An element (p, q) in the one-dimensional persistence diagram models that a one-dimensional hole appeared in simplicial complex $\mathcal{K}_p$ and subsequently disappeared in simplicial complex $\mathcal{K}_q$. Finally, an element (p, q) in the two-dimensional persistence diagram models that a two-dimensional hole or void appeared and subsequently disappeared in simplicial complexes $\mathcal{K}_p$ and $\mathcal{K}_q$ respectively. To compute persistence diagrams, persistent homology implicitly matches connected components and holes between consecutive simplicial complexes in the filtration. This matching can be reduced to determining if the connected components and holes in question intersect between consecutive simplicial complexes in the filtration.

If a connected component or hole appears in a given filtration at simplicial complex $\mathcal{K}_p$ but does not subsequently disappear, it is modelled by an element (p, ∞) in the corresponding persistence diagram. The value $q - p$ corresponding to an element (p, q) in a given persistence diagram is known as the *persistence* of the element in question[3]. Note that, to compute a persistence diagram of a given dimension, one only needs to consider a filtration containing simplicial complexes of one dimension higher. For example, to compute a zero-dimensional persistence diagram one only needs to consider a filtration containing simplicial 1-complexes.

Consider the sequence of four simplicial complexes displayed in Figures 2(a), 2(b), 2(c) and 2(d) that we denote $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{K}_3$ and $\mathcal{K}_4$ respectively. This sequence forms a standard filtration and therefore we can compute the corresponding persistence diagrams by applying standard persistent homology. The zero-dimensional persistence diagram corresponding to this sequence equals the set $\{(2, 3), (1, \infty)\}$. The element $(1, \infty)$ corresponds to the connected component which appears in $\mathcal{K}_1$ and never disappears. Note that, the set of simplices corresponding to this connected component increases in size during the sequence as additional simplices are added to it. The element $(2, 3)$ corresponds to the connected component consisting of a single point which appears in $\mathcal{K}_2$ and disappears in $\mathcal{K}_3$ when it becomes connected to the previous connected component. The one-dimensional persistence diagram corresponding to this sequence equals the set $\{(2, 4)\}$. The element $(2, 4)$ corresponds to the one-dimensional hole which appears in $\mathcal{K}_2$ and disappears in $\mathcal{K}_4$ when it becomes filled in by a 2-simplex. Illustrations of these persistence diagrams (corresponding to Figure 2) are shown in Figures 9(a) and 9(b) and further explanation of such diagrams is provided in Section 2.3.1.

Next consider the sequence of four simplicial complexes displayed in Figures 3(a), 3(b), 3(c) and 3(d) that we denote $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{K}_3$ and $\mathcal{K}_4$ respectively. This sequence does not form a standard filtration because, for example, a 2-simplex is removed from $\mathcal{K}_1$ to form $\mathcal{K}_2$. Therefore we can compute the corresponding persistence diagrams by constructing a zig-zag filtration and applying zig-zag persistent homology. The zero-dimensional persistence diagram corresponding to this sequence equals the set

[3]There is an alternative form of output to a persistence diagram, known as a bar code consisting of a stack of horizontal lines the length of each or which represents the value of *persistence*.

$\{(2,3),(1,\infty)\}$. The element $(1,\infty)$ corresponds to the connected component which appears in $\mathcal{K}_1$ and never disappears. Note that, the set of simplices corresponding to this connected component changes during the sequence as additional simplices are added to it while others are removed. The element $(2,3)$ corresponds to the connected component which appears in $\mathcal{K}_2$ and disappears in $\mathcal{K}_3$ when it becomes connected to the previous connected component. The one-dimensional persistence diagram corresponding to this sequence equals the set $\{(2,4)\}$. The element $(2,4)$ corresponds to the one-dimensional hole which appears in $\mathcal{K}_2$ and disappears in $\mathcal{K}_4$ when a 1-simplex on its boundary is removed.

When multiple connected components or holes merge, a single connected component or hole respectively will persist and all others will disappear. For example, consider again the sequence of simplicial complexes in Figure 2, where one connected component disappears when it becomes connected to another connected component. When such a merger happens, it must be decided which connected component or hole persists. Many implementations of standard and zig-zag persistent homology use a solution known as the *elder rule* whereby the connected component or hole which appeared first in the filtration is the one which persists (Otter *et al.* 2017). Alternatively, one can use a solution where the largest connected component or hole is the one which persists (Corcoran 2019b).

Finally, it is worthwhile to briefly consider the relationship between persistent homology and Betti numbers which is an alternative model of topological features. The Betti numbers equal the number of connected components and holes of different dimensions in a single given simplicial complex. They do not consider the persistence of connected components and holes across a sequence of simplicial complexes. Therefore, persistent homology can be considered a richer model of topological features.

2.3. *Persistent homology analysis*

The output from both standard and zig-zag persistent homology is a set of persistence diagrams. As described at the beginning of this section, the next step in the persistent homology workflow is to perform an analysis of this output.

In recent years there has been a lot of research in the development of new methods to assist in performing this analysis. Consequently, there exist a large number of such methods and we, unfortunately, cannot review them all here. Hence in this section we only describe those methods which we consider to be fundamental or particularly useful. A reader seeking a more in-depth review can consult the following review articles (Pun *et al.* 2018, Hensel *et al.* 2021). We have structured our description of methods into the following two parts. In Section 2.3.1 we describe methods which use visualisation and manual interpretation. Subsequently, in Section 2.3.2 we describe more automated methods which use data science and machine learning methods. Ultimately, the suitability of a given method for performing analysis will depend on the data of interest, how the persistent homology of this data is computed and the specific research question one is attempting to answer.

2.3.1. Visualisation and manual interpretation

It is common practice to visualise persistence diagrams using two-dimensional figures. In these figures, each element in a given persistence diagram is represented as a point in the corresponding figure where the appearance and disappearance values are represented by the x- and y-axis respectively. For example, Figures 9(a) and 9(b) display

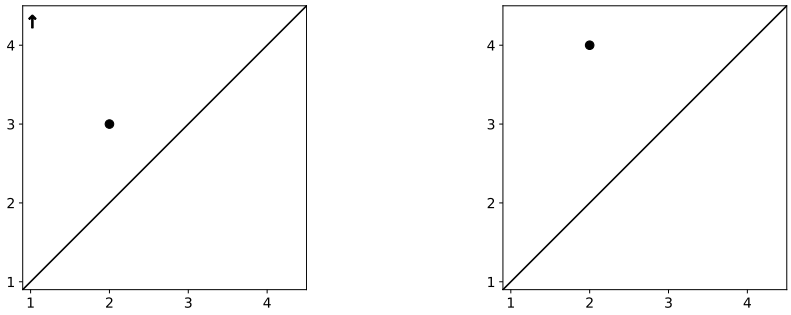


Figure 9. The persistence diagrams $\{(2,3),(1,\infty)\}$ and $\{(2,4)\}$ are displayed in (a) and (b) respectively. Points (p,∞) which have infinite persistence are represented using arrows at the corresponding locations $(p,4.3)$.

two such figures corresponding to the persistence diagrams $\{(2,3),(1,\infty)\}$ and $\{(2,4)\}$ respectively. Note that, in this type of visualisation all points will lie above the diagonal and points with smaller corresponding persistence will be located closer to the diagonal.

How to interpret a given persistence diagram depends on the data being modelled and how the corresponding filtration was constructed. We now demonstrate this by considering the VR filtration of a set of points and the zig-zag filtration of an image time series. First, consider the set of two-dimensional points displayed in Figure 10(a) which contains three compact clusters. The zero and one-dimensional persistence diagrams corresponding to the VR filtration of this set of points are displayed in Figures 10(b) and 10(c) respectively. The zero-dimensional persistence diagram contains many elements with small persistence and three elements with significant persistence where two of these elements have finite persistence and one has infinite persistence. The elements with small persistence correspond to the connected components formed by each individual data point. The three elements with significant persistence indicate that the data contains three significant clusters. Of these three elements, the persistence of the two elements with finite persistence equals the distance between the two clusters closer together and the distance between these two clusters and the third cluster (these distances being the 'resolution' parameter values at which the respective pairs of clusters would merge). The one-dimensional persistence diagram contains many elements with small persistence and no elements with significant persistence. This indicates that the data does not contain any significant one-dimensional holes. A real-world example of the application of persistent homology to sets of points is presented in Section 3.2.

Next consider the set of two-dimensional points displayed in Figure 10(d) which contains a single compact cluster in the form of a figure of eight which in turn contains two holes. The zero and one-dimensional persistence diagrams corresponding to the VR filtration of this set of points are displayed in Figures 10(e) and 10(f) respectively. The zero-dimensional persistence diagram contains a single element of significant persistence reflecting the fact that the data contains a single cluster as indicated above. The one-dimensional persistence diagram contains two elements of significant persistence indicating that the data contains two significant one-dimensional holes. The persistence of these elements, which are both finite values, equals the diameter of the holes in question. It is important to note that although we have considered sets of two-dimensional points above, the analysis generalises to higher dimensional data. In

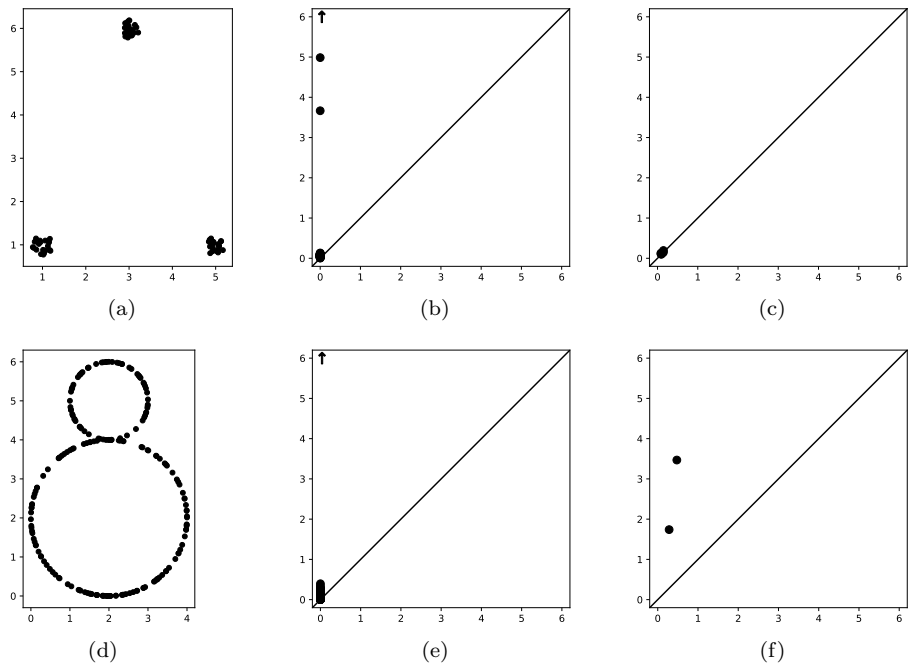


Figure 10. The zero- and one-dimensional persistence diagrams corresponding to the VR filtration of the set of points in (a) are displayed in (b) and (c) respectively. The zero- and one-dimensional persistence diagrams corresponding to the VR filtration of the set of points in (d) are displayed in (e) and (f) respectively.

such cases, persistent homology provides a means of inferring topological information which cannot easily be inferred by data visualisation.

Finally, consider the zig-zag filtration of the sequence of images described in Section 2.1.3. Let us assume that the image sequence is a time series where the time interval between images is β. In this case an element (p, q) in the zero-dimensional persistence diagram represents the fact that a connected component appeared at time $p\beta$, persisted for $(q - p)\beta$ time and disappeared at time $q\beta$. A similar interpretation can be applied to a one-dimensional persistence diagram. The application of zig-zag filtration and persistence diagrams to a real dataset is described in Section 3.3.

2.3.2. Data science and machine learning methods

In this section we describe data science and machine learning methods for the analysis of persistence diagrams. Given a single persistence diagram, one can perform an analysis by computing and interpreting summary statistics describing the elements in the diagram. Such statistics include the number of elements and the mean and variance of these elements' persistence. Similarly, given a set of persistence diagrams, one can perform an analysis by computing and interpreting summary statistics describing this set. Such statistics include the mean number of elements in the set of diagrams. Which statistics to compute will depend on the specific research question one is attempting to answer. In the following sections we explain some of these statistics and provide specific examples of their application to particular research challenges.

Many data science and machine learning methods require a distance measure or metric to be defined on the input space. For example, the k-nearest neighbours and k-means clustering algorithms perform classification and clustering using a distance measure. Two popular metrics defined on the space of persistence diagrams are the

bottleneck distance and the *p-th Wasserstein distance*. Recall that a persistence diagram is a multiset of elements in the space $\{(p, q) \in \mathbb{R}^2 |\ p < q\}$. Before defining the above distances, we first define a distance between two given elements $x = (x_1, x_2)$ and $y = (y_1, y_2)$ as $\|x - y\|_\infty = \max\{|x_1 - y_1|, |x_2 - y_2|\}$. We also assume that each persistence diagram also contains the additional elements $\{(p, q) \in \mathbb{R}^2 |\ p = q\}$ of which there is an infinite number. If we consider the persistence diagram visualisations in Figure 9, these additional elements lie on the diagonal of the figures. Given two persistence diagrams X and Y, let $\eta : X \to Y$ be a bijection from X to Y. The bottleneck distance between X and Y is then defined as follows (Edelsbrunner and Harer 2010):

$$W_\infty(X, Y) = \inf_{\eta: X \to Y} \sup_{x \in X} \|x - \eta(x)\|_\infty \tag{8}$$

This metric determines the bijection which minimizes the maximum distance between corresponding elements in X and Y where the function $\eta(x)$ finds those corresponding elements. Note that, if X and Y contain a different number of elements, all bijections will map some elements in X to elements in the set $\{(p, q) \in \mathbb{R}^2 |\ p = q\}$ or some elements in the set $\{(p, q) \in \mathbb{R}^2 |\ p = q\}$ to elements in Y. Broadly speaking, the bottleneck distance will assign smaller distances to pairs of persistence diagrams where a bijection exists mapping elements in X to elements in Y in similar locations. If an element in X is not mapped to an element in Y or vice versa, it will instead be mapped to an element in the set $\{(p, q) \in \mathbb{R}^2 |\ p = q\}$. Therefore, the further these unmapped elements are from this set the greater the bottleneck distance. This models the fact that the further elements are from the set $\{(p, q) \in \mathbb{R}^2 |\ p = q\}$, the greater their corresponding persistence and in turn their significance.

The bottleneck distance considers the maximum distance between corresponding elements in X and Y. This makes it sensitive to outliers in the set of distance values. To overcome this, the p-th Wasserstein distance considers the sum, instead of the maximum, of all p-th powers of distances and is defined as follows:

$$W_p(X, Y) = \left[\inf_{\eta: X \to Y} \sum_{x \in X} \|x - \eta(x)\|_\infty^p \right]^{1/p} \tag{9}$$

The standard and zig-zag persistent homology are both stable with respect to the bottleneck and p-th Wasserstein distance measures (Botnan and Lesnick 2018, Skraba and Turner 2020). Thus if you change the input to persistent homology slightly, the change in the output persistence diagrams as measured by the bottleneck and p-th Wasserstein distance measures will be small. This is an important property because if persistent homology was not stable, one could not determine if the persistence diagrams obtained were a function of the actual structure in the input data or noise. Finally, it is important to note that computing the bottleneck and p-th Wasserstein distance measures is computationally expensive because both measures require an optimisation procedure to compute the bijections between persistence diagrams (Kerber *et al.* 2017).

As discussed above persistent homology returns a set of persistence diagrams which are multisets of elements. Many popular machine learning methods assume their inputs are elements in a vector space. Such methods include the support vector machine, the random forest and the multilayer perceptron (MLP). There do exist some machine learning methods which can be applied to sets but these methods are difficult to ap-

ply in practice (Zaheer *et al.* 2017). The desire to combine machine learning with persistent homology has led to the development of an array of methods for transforming persistence diagrams into a representation which is an element in a vector space (Chevyrev *et al.* 2018). Two popular such methods are *persistence landscapes* (Bubenik *et al.* 2015) and *persistence images* (Adams *et al.* 2017). For an example of the application of persistence landscapes to analysis of swarm behaviour see Corcoran and Jones (2017).

3. Applications of persistent homology in GIS

Persistent homology has many potential applications to problems in the GIS domain. In Section 3.1 we present an overview of existing applications previously published in the literature. In Sections 3.2 and 3.3 we consider in more detail the two applications of point pattern analysis of UK city pub (public house/bar) locations and the analysis of UK rainfall radar data. In doing so we describe existing solutions to each problem and highlight the relative benefits of the proposed persistent homology solutions.

3.1. *Overview of applications*

Feng and Porter (2020) propose a method for analysing the topological properties of street networks using persistent homology. The authors construct a standard filtration based on a parameter that measures the distance to the nearest street before computing the standard persistent homology of this filtration. This method models the number, size and shape of regions enclosed or surrounded by streets. In subsequent work, Feng *et al.* (2022) propose a method for analysing the topological properties of COVID-19 infections. The authors construct a standard filtration with a function that measures the density of infections. Using the corresponding persistent homology they demonstrate that this method can detect infection hotspots. Feng and Porter (2021) propose a method for analysing the topological properties of election voting patterns. A standard filtration is constructed with a function defined on geographical regions measuring voting preference, before computing the persistent homology. The method is demonstrated to identify regions with voting patterns different from surrounding regions. In related work, Duchin *et al.* (2021) propose a method for analysing the topological properties of election gerrymandering.

Corcoran and Jones (2021) propose a method for analysing the connectivity of street networks. The authors construct a standard filtration with a function that measures the degree of connectivity provided by different street types, before computing the persistent homology of this filtration. The method is demonstrated to identify regions of poor and good connectivity. They also demonstrated that clustering based on 2-th Wasserstein distance can identify cities with similar connectivity properties. Wu *et al.* (2017) analyse traffic congestion in street networks using persistent homology. They employ a standard filtration with a function that measures the speed of traffic which is proportional to the level of congestion. By computing the standard persistent homology of this filtration the authors demonstrate that this method can identify regions experiencing traffic congestion. Carmody and Sowers (2021) extended the method of Wu *et al.* (2017) to consider street networks where the speed of travel along a street can vary depending on the direction of travel.

Ahmed *et al.* (2014) propose a model for determining local differences between street networks. They construct a standard filtration with a function that measures

the distance to the nearest street, before computing the standard persistent homology of this filtration. Specifically, they use a version of standard persistent homology which considers local instead of global topological features. The method is shown to recognize changes in street networks over time and assess the quality of street networks inferred from GPS data.

Corcoran (2019a) proposes a method for performing generalisation or simplification of images representing digital elevation models. The authors construct a standard filtration with a function based on height. They compute the standard persistent homology of this filtration and optimise a function defined with respect to the persistent homology.

Corcoran and Jones (2016, 2017) propose a method for analysing the topological properties of swarm behaviour exhibited by a set of agents. The author constructs a zig-zag filtration that uses kernel density estimation of agent locations. They compute the zig-zag persistent homology of this filtration and convert the persistence diagrams to persistent landscape representations. The authors demonstrate that clustering based on 2-th Wasserstein distance can identify the common swarm behaviours of flock, torus, and disorder in a school of fish.

3.2. *Point pattern analysis of UK city pubs*

Point pattern analysis concerns the problem of analysing the spatial patterns of sets of points. There exist a large array of methods for performing this task. Many of these methods involve computing descriptive statistics such as the minimum bounding box, the mean or centre location of the points, the mean distance between pairs of points, measures of point density and measures of point distribution randomness based on quadrat counts (Baddeley *et al.* 2015). More advanced methods include kernel density estimation (KDE) and spatial clustering using methods such as DBSCAN. Persistent homology is most similar to these latter methods in the sense that all methods are commonly used to analyse clustering structures. In this section we consider the problem of performing point pattern analysis of sets of points corresponding to UK city pub locations. In doing so we highlight the benefits that persistent homology offers relative to KDE or spatial clustering. Robins and Turner (2016) previously considered the application of persistent homology to point pattern analysis but did not consider applications in the GIS domain.

The UK contains 70 cities. To define the boundaries of these cities, we used the Urban Centre Database (UCD), which is a project supported by the European Commission's Joint Research Centre and Directorate-General for Regional and Urban Policy (Florczyk *et al.* 2019). This database contains the geographical boundaries for most major urban areas in the world which are derived from a fusion of census population and remotely sensed image data. Some of the UK's cities are very small urban areas and therefore are not represented in the UCD. For example, the city of Armagh has a population of less than 15 thousand and is not represented. Furthermore, in some cases, a set of spatially close UK cities are represented as a single urban area in the UCD. For example, the cities of Wolverhampton and Birmingham are represented as a single urban area entitled Birmingham. In this work, we only considered those UK cities represented in the UCD which corresponded to 48 distinct urban areas. For each city we extracted the locations of all pubs located inside its corresponding boundary using OpenStreetMap (OSM) (Corcoran and Mooney 2013). We defined a pub to be any OSM feature with the tag *amenity=bar* or *amenity=pub*. These tags were deter-

City	Area	No. Pubs	City	Area	No. Pubs
Aberdeen	57	100	Lancaster	37	77
Bangor	20	13	Leeds	472	842
Bath	22	73	Leicester	122	162
Belfast	163	78	Lincoln	38	65
Birmingham	668	1048	Liverpool	418	565
Brighton	108	280	London	1865	3669
Bristol	169	382	Manchester	674	1021
Cambridge	37	78	Newcastle	241	510
Canterbury	15	42	Newport	57	60
Cardiff	109	152	Norwich	55	109
Carlisle	17	31	Nottingham	170	295
Chelmsford	31	49	Oxford	46	104
Chester	27	80	Peterborough	58	44
Coventry	126	186	Plymouth	88	131
Derby	83	112	Portsmouth	166	237
Derry	28	31	Preston	79	102
Dundee	62	87	Sheffield	247	400
Durham	99	142	Southampton	144	145
Edinburgh	123	325	Southend-on-Sea	75	43
Exeter	31	63	St Albans	27	48
Glasgow	306	354	Stoke-on-Trent	118	159
Gloucester	45	69	Sunderland	59	85
Hereford	17	31	Swansea	60	75
Hull	102	204	Worcester	30	59

Table 1. This table displays the names of the 48 UK cities considered plus the corresponding geographical area (measured in km^2) and the number of pubs.

mined to be appropriate after studying the OSM wiki which defines the meaning of different tags[4]. Table 1 displays for each city the corresponding geographical area and number of pubs.

For each city, we constructed a VR filtration of the corresponding set of pub locations as described in Section 2.1.1. We subsequently computed the standard persistent homology of this filtration as described in Section 2.2. Figures 1(a) and 1(b) display the set of pub locations corresponding to Cardiff and Manchester city respectively. The zero-dimensional persistence diagrams corresponding to these sets are displayed in Figures 11(a) and 11(c) respectively. Recall that zero-dimensional persistence diagrams model the existence of connected components or clusters. The persistence diagram corresponding to Cardiff city contains more elements of greater persistence than that corresponding to Manchester city. This is a consequence of the fact that Cardiff city contains clusters of pubs spatially separated from other pubs. For example, if we examine the Cardiff city pub locations in Figure 1(a) we can see such clusters in the centre, the south, the west and the north east. On the other hand, the spatial distribution of pubs in Manchester city is more uniform and any clusters of pubs are not spatially separated from other pubs.

The one-dimensional persistence diagrams corresponding to the sets of Cardiff and Manchester pub locations are displayed in Figures 11(b) and 11(d) respectively. Recall that, one-dimensional persistence diagrams model the existence of holes. The persis-

[4] `https://wiki.openstreetmap.org/wiki/Map_features#Amenity`

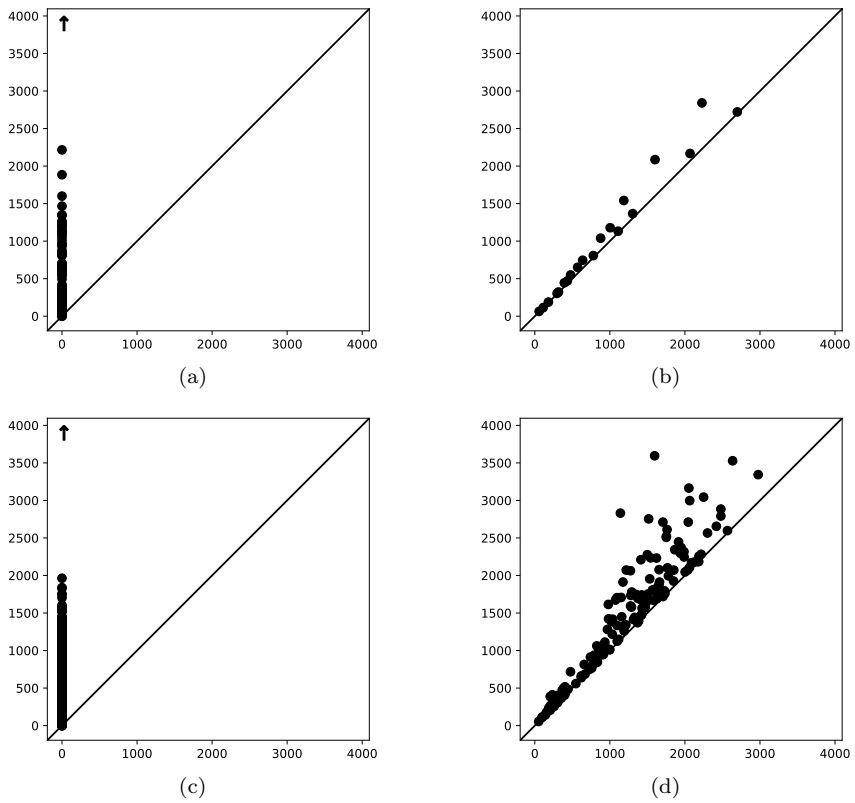
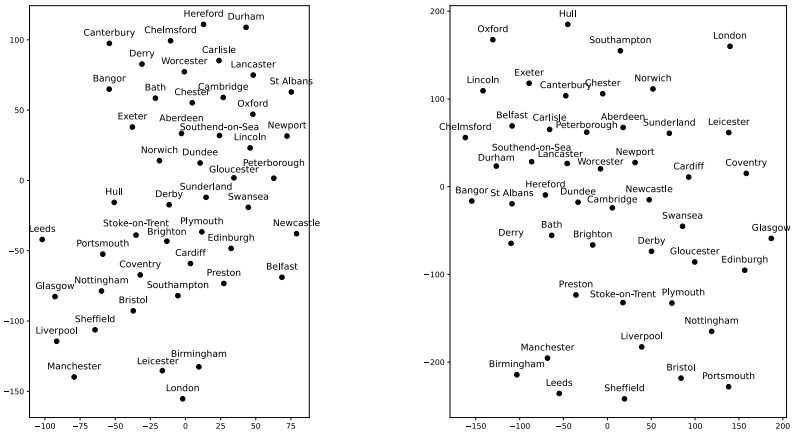


Figure 11. The zero- and one-dimensional persistence diagrams corresponding to the set of Cardiff city pub locations are displayed in (a) and (b) respectively. The zero- and one-dimensional persistence diagrams corresponding to the set of Manchester city pub locations are displayed in (c) and (d) respectively.

tence diagram corresponding to Manchester city contains more elements of greater persistence than that corresponding to Cardiff city. This is a consequence of the fact that Manchester city contains more larger regions containing no pubs which are surrounded by pubs. For example, if we examine the Manchester city pub locations in Figure 1(b) we can see such regions in the north west and north east.

To analyse the entire set of UK cities, for each pair of cities we computed the 2-nd Wasserstein distance between the corresponding pair of zero-dimensional persistence diagrams and the 2-nd Wasserstein distance between the corresponding pair of one-dimensional persistence diagrams. This computation gives a zero-dimensional persistence diagram distance matrix and a one-dimensional persistence diagram distance matrix. For each distance matrix we computed a corresponding representation of each city as a point in $\mathbb{R}^2$ using the t-SNE manifold learning technique (Maaten and Hinton 2008). These representations are displayed in Figures 12(a) and 12(b) respectively. Examining these figures we can see the formation of clusters in both representations. For example, in both representations London city is a member of a distinct cluster. In the case of the zero-dimensional persistence diagrams, the cluster in question also contains the cities of Leicester and Birmingham. On the other hand, in the case of the one-dimensional persistence diagrams, the cluster in question only contains the city of London. Other patterns evident in these diagrams include the fact that both the zero- and one-dimensional persistence diagrams corresponding to Cardiff and Manchester are relatively dissimilar. The reasons for this difference were discussed previously.

To further examine this clustering behaviour, we performed hierarchical single-



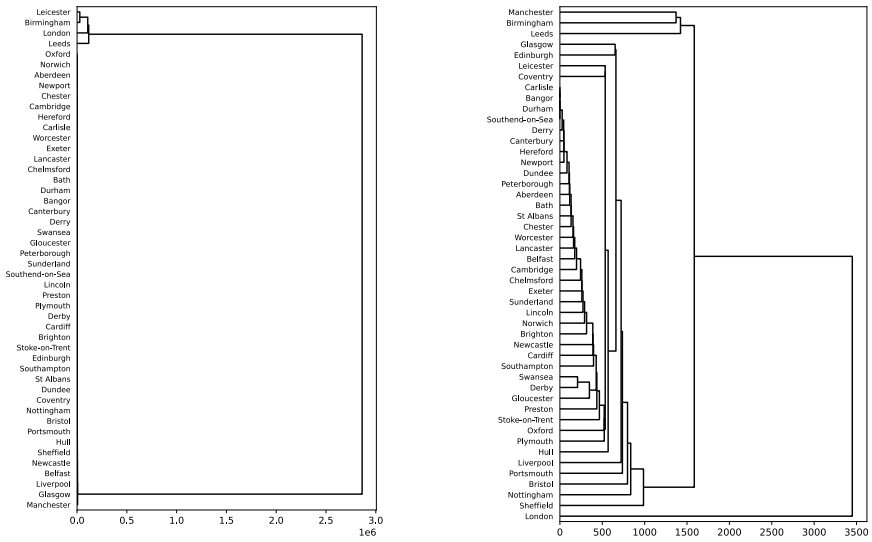
(a) (b)

Figure 12. A representation of each UK city as a point in $\mathbb{R}^2$ computed by applying t-SNE to the pairwise 2-th Wasserstein distances between the zero- and one-dimensional persistence diagrams are displayed in (a) and (b) respectively.

linkage clustering using distance matrices to obtain dendrogram representations (Everitt *et al.* 2011). The dendrogram representations corresponding to the zero- and one-dimensional persistence diagrams are displayed in Figures 13(a) and 13(b) respectively. In both dendrograms we can again see that London city is a member of a distinct cluster. To understand the reasons for this, note from Table 1 that London city contains a larger number of pubs relative to other cities. Furthermore, consider the zero- and one-dimensional persistence diagrams corresponding to London city which are displayed in Figures 14(a) and 14(b) respectively. Relative to the persistence diagrams corresponding to Cardiff and Manchester displayed in Figure 11, both diagrams contain a larger number of elements of smaller persistence. This can be attributed to the higher density of pubs in London that results in a larger number of spatially close clusters and a larger number of smaller regions surrounded by pubs.

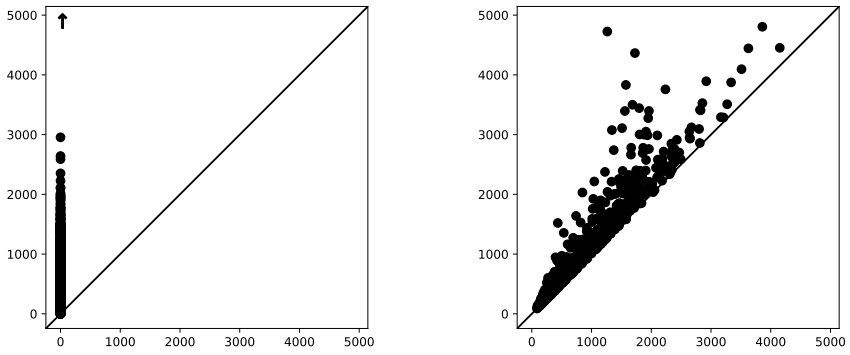
As discussed above, persistent homology has similarities to the methods of KDE and spatial clustering for point pattern analysis. We now discuss the benefits that persistent homology offers relative to these methods. KDE and spatial clustering methods such as DBSCAN (Schubert *et al.* 2017) are the most commonly used methods for detecting clusters in sets of points. In KDE clusters are detected by visual inspection of the density function or by thresholding the density function followed by spatial clustering. In DBSCAN clusters are detected by grouping points that are sufficiently spatially close to each other. Both KDE and DBSCAN have a scale parameter which must be selected and both methods are not stable with respect to the choice of these parameters. Specifically, the choice of the KDE kernel bandwidth parameter and the DBSCAN maximum distance parameter can significantly affect the number and shape of the clusters detected. Persistent homology does not suffer from this undesirable sensitivity to scale parameter selection. Persistent homology instead considers the existence of connected components and holes across all scales simultaneously. In fact, as indicated previously, persistent homology has a stability property that ensures a small change in the input data results in at most a small change in the output (Cohen-Steiner *et al.* 2007).

Although KDE and density-based clustering methods can infer the existence of clusters, these methods do not compute any information relating to the scale of these



(a) (b)

Figure 13. Dendrograms for UK cities computed by applying single linkage clustering to the pairwise 2-th Wasserstein distances between the zero- and one-dimensional persistence diagrams are displayed in (a) and (b) respectively.



(a) (b)

Figure 14. The zero- and one-dimensional persistence diagrams corresponding to the set of London city pub locations are displayed in (a) and (b) respectively.

clusters. On the other hand, as well as inferring the existence of clusters, persistent homology also measures the persistence of these clusters across particular scale (or another parameter) values. Recall from the illustrative example in Figure 10, that in this context persistence models how separated and hence how distinct the clusters are. Persistent homology can also infer the existence and scale of void regions modelled as holes. Being able to detect holes has many potential applications in point pattern analysis. For example, if the points correspond to instances of a service or facility, the existence of a hole indicates a lack of availability in that region. To date, there has been little research on the topic of detecting holes in point pattern analysis. Finally, the outputs from persistent homology facilitate the application of many downstream data mining and machine learning tasks. This is because metrics have been defined with respect to these outputs. Furthermore, these outputs can be transformed into a vector space representation. On the other hand, defining a metric on the space of KDE and density-based clusterings, or transforming these clusterings into a vector space representation is not straightforward.

3.3. *Spatio-temporal analysis of UK rainfall radar imagery*

The tracking of objects in geographical data is a fundamental problem in the field of GIS with many applications (Worboys and Duckham 2006). For example, in order to make a weather forecast is it necessary to track weather features or objects such as a storm. Successful object tracking requires that correspondences between the same object existing at different discrete times can be determined. In many cases, object properties will change over time making it challenging to correctly infer these correspondences. The topological properties of many objects in geographical data will change over time. For example, in the case of a weather storm, changes in topological properties include the formation of holes plus the splitting into and merging of multiple connected components. For this reason, many researchers in the domain of GIS have considered the development of methods for object tracking in geographical data (Jiang and Worboys 2009).

In this section we demonstrate how persistent homology can be used to analyse the topological patterns of UK rainfall radar images. A significant contribution of persistent homology here is the use of zig-zag homology methods for tracking individually identifiable weather features between successive time frames. This analysis applies the methods presented in Corcoran and Jones (2018), Corcoran (2019b). Rainfall radar images were obtained from the UK Meteorological (Met) Office which provides an image time series where the interval between consecutive images is 15 minutes. Each image represents the level of rainfall at each location in a 500x500 regular grid over Ireland and the UK. Specifically, the level of rainfall is represented as lying in one of 8 intervals of rainfall levels measured in terms of the number of millimetres (mm) of rainfall per hour.

We obtained a time series of images between 11:45 on 6 December 2021 and 7:45 on 12 December 2021. This time series contains a total of 1,041 images. The start of this time series corresponds to the time period during which a significant storm called Storm Barra passed over Ireland and the UK. The end of this time series corresponds to a period of less rainfall after the storm had passed. For this work, we converted the original images into binary images where values of 0 and 1 represent less than and greater than or equal to 0.01mm of rainfall per hour respectively. The threshold value of 0.01mm was chosen because the Met Office considers levels of rainfall less than this

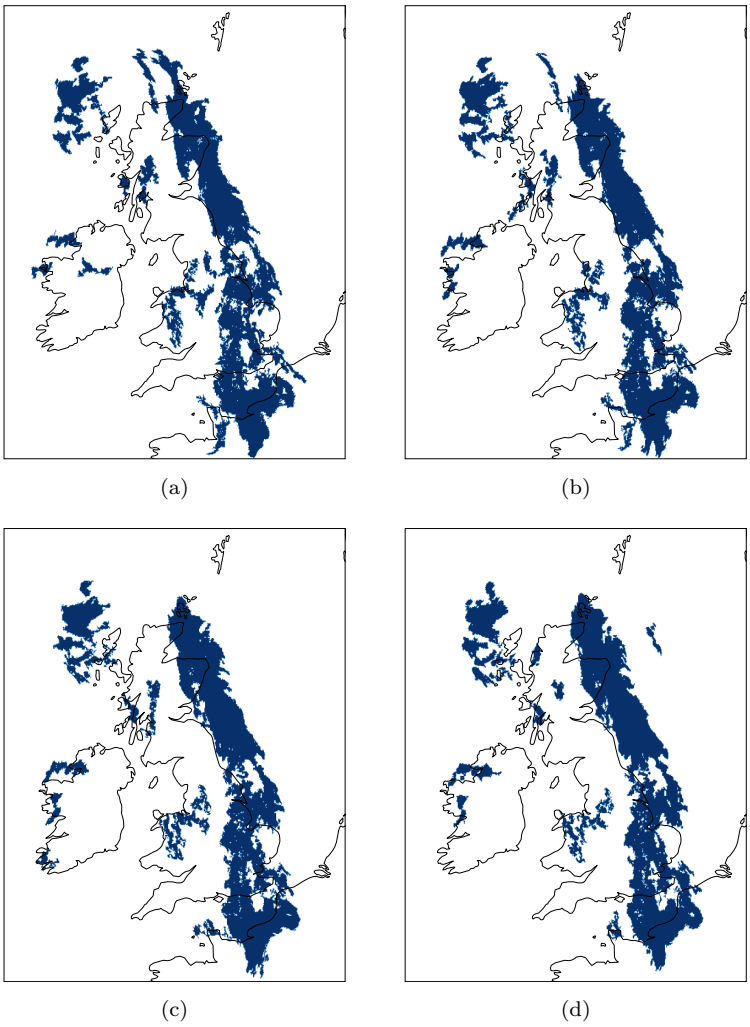


Figure 15. Four consecutive images in the rainfall radar image time series are displayed (a), (b), (c) and (d). In each image the existence of rainfall is represented by the colour blue.

value as no rainfall. Figure 15 displays a subset of the above time series containing four consecutive images.

In our analysis we study the topological evolution of connected components in the radar images which correspond to rain clouds. This evolution may involve the events of rain clouds appearing, disappearing, merging and splitting. For this analysis we only need to consider the zero-dimensional persistence diagram. If we were to study the evolution of holes with connected components we would need to consider the one-dimensional persistence diagram. The radar images can sometimes contain many small connected components. We consider these to be topology noise and therefore preprocessed the images to remove all connected components less than 50 pixels in size.

For the image time series we constructed a zig-zag filtration as described in Section 2.1.3. We subsequently computed the zig-zag persistent homology of this filtration as described in Section 2.2. Note that we used a more advanced implementation of zig-zag persistent homology than that described above adopting the method presented in Corcoran (2019b) in which they consider spatially close connected components to

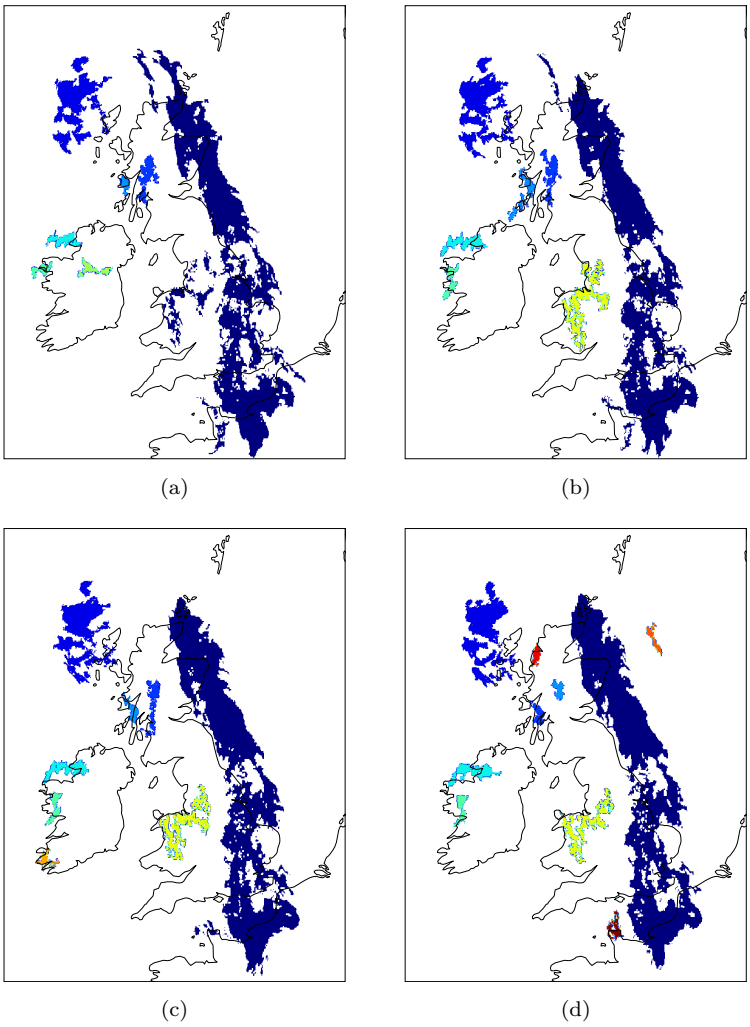


Figure 16. A cartographic (as opposed to diagrammatic) representation of the zero-dimensional persistence diagram $\{(1,\infty),(1,\infty),(1,\infty),(1,\infty),(1,\infty),(1,\infty),(1,2),(2,\infty),(3,4),(4,\infty),(4,\infty),(4,\infty)\}$ corresponding to the time series in Figure 15 is illustrated. Each coloured weather feature is a connected component and corresponds to an element in the persistence diagram. Note that the prominent blue and cyan features correspond to persistence diagram elements with coordinates $(1,\infty)$.

be the same connected component.

Consider again the time series containing four consecutive images displayed in Figure 15. Applying the above methodology to this time series gives the zero-dimensional persistence diagram $\{(1,\infty),(1,\infty),(1,\infty),(1,\infty),(1,\infty),(1,\infty),(1,2),(2,\infty),(3,4),(4,\infty),(4,\infty),(4,\infty)\}$. A cartographic representation of this persistence diagram is displayed in Figure 16 where each rainfall feature is a connected component corresponding to an individual element and is represented using a unique colour. For example, one of the elements with value $(1,\infty)$ corresponds to the large dark blue coloured connected component in the east. This connected component persists across all time four steps 1 to 4 inclusive, hence does not disappear and has infinite persistence. Similarly, the element $(2,\infty)$ corresponds to the smaller connected component in the centre which persists across the three steps 2 to 4 inclusive and is represented by a yellow colour.

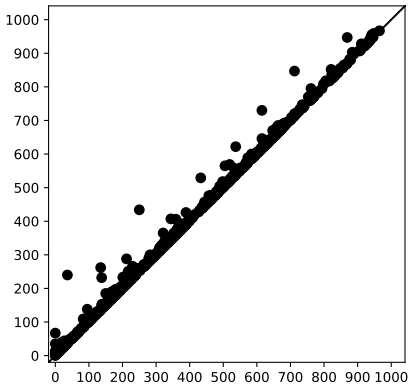


Figure 17. The zero-dimensional persistence diagram corresponding to the time series of 1,041 rainfall radar images is displayed.

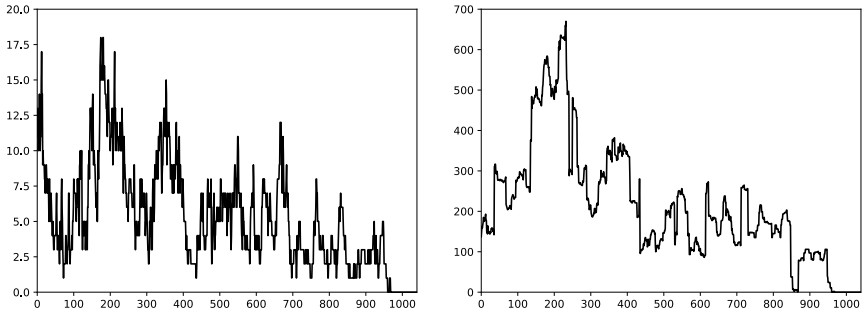
Figure 17 displays the zero-dimensional persistence diagram corresponding to the full time series containing 1,041 images instead of the smaller time series containing only 4 images considered above. Let us denote this persistence diagram as D. The total number of elements in D is 773. To interpret the temporal behaviour of D, for each time i we computed the following two statistics which measure the number of elements that persist during time i and the sum of their persistence values respectively.

$$|\{(p, q) \in D : p \leq i \leq q\}| \tag{10}$$

$$\sum_{\{(p,q)\in D : p\leq i\leq q\}} q - p \tag{11}$$

Plots of the above statistics versus time are displayed in Figures 18(a) and 18(b) respectively. From these plots we see that both statistics have larger values at the start and smaller values at the end of the time series. In fact, both statistics have a value of 0 at the very end of the time series. This indicates that the number of connected components and the persistence of these connected components decreased over the course of the time series. This reflects the fact that the start of the time series corresponds to the time period during which a storm passed over the region while the end of the time series corresponds to a period of less rainfall after the storm had passed.

Several models for tracking objects with changing topological properties have been previously proposed. Liu and Schneider (2010, 2011) proposed a model which considers objects corresponding to multiple connected components. However, this model is conceptual in nature and no corresponding computational model is proposed. Worboys and Duckham (2006) and Jiang and Worboys (2009) proposed models for using a sensor network to track topological changes in objects. However, again these models are mostly conceptual in nature and cannot directly be applied to real data without further development. For example, both models assume a continuous representation of change in topology which does not exist in real data where changes occur at discrete time steps. Consequently, the authors limited their evaluation to simulated continuous data. The inability of the above models to be applied to real data can be attributed in part to the fact they are formulated in terms of logic which is not suitable for modelling



(a) (b)

Figure 18. The number of elements that persist at each time and the sum of their persistence are plotted in (a) and (b) respectively. In both plots time is represented by the x axis while the statistic in question is represented by the y axis.

noisy data with uncertainty. On the other hand, persistent homology is fundamentally designed to model data with these properties.

Another advantage of the proposed model for object tracking is that the outputs from persistent homology facilitate the application of many downstream data mining and machine learning tasks. Although we do not explore this research direction in this paper, it has previously been considered. For example, Corcoran and Jones (2016, 2017) demonstrated that clustering the persistence diagrams of swarm behaviour can discover the swarm behaviours of flock, torus, and disordered.

4. Conclusions

This article aims to provide an introduction to topological data analysis for GIS researchers and practitioners. In doing so we have provided multiple examples of how TDA methods can be of benefit in working with geospatial data, addressing our first research question. With regard to the second and third research questions that focus on the benefits of persistent homology methods for the analysis of point patterns and remotely sensed spatio-temporal data, we have enumerated specific benefits in combination with a detailed description of two case studies of the application of these methods. Given that some of the methods described for point pattern analysis might be regarded as analogous to well-known GIS methods such as kernel density estimation and density-based clustering, we have stressed the distinctive advantages of the presented TDA methods with regard, for example, to avoiding the need for parameter selection; the generation of outputs that characterise the scale and the numbers of features; the output of signatures of the analysed patterns of data that facilitate analysis such as similarity measurement, machine learning and trajectory computation; and the fact that the methods can detect and analyse void regions, in addition to reporting on clusters of objects.

TDA has grown into a very large research field in recent years and it would not be possible to consider all aspects and methods in a single article. Therefore, we instead considered those aspects and methods we believe to be most necessary and useful. For example, although zig-zag persistent homology might be considered a relatively advanced topic for an introductory article, we have chosen to include it here as it has proven to be particularly useful in spatio-temporal applications that work with sequences of time-stamped data sets.

Due to its prominence in the field of TDA, in this article, we have almost exclusively focused on the method of persistent homology. However, the field of TDA contains many other methods which are also potentially very applicable to problems in the GIS domain. Such methods include the mapper algorithm (Singh *et al.* 2007), contour trees (Carr *et al.* 2003), Reeb graphs (Biasotti *et al.* 2008) and discrete Morse theory (De Floriani *et al.* 2015). To date, there have only been a handful of works which have considered the application of these methods to GIS problems. For example, Dey *et al.* (2017) proposes a method for inferring a street network from GPS data using discrete Morse theory. We expect the application of TDA methods to GIS problems to continue to grow in the future and in turn the application of these other methods to do so also.

In our coverage of persistent homology we described how persistent homology can be applied to three different types of data commonly encountered in GIS. Namely, sets of points, networks and sequences of images. However, there are other types of data also commonly encountered in GIS which we did not consider. This includes sets of lines and polygons which may share boundaries. This type of data was only recently considered by Feng and Porter (2021) and it appears to remain an open question as to how persistent homology can best be applied to such data. In our case study of the application of persistent homology to analyse the spatial distribution of pubs, we highlighted that these methods help us to understand the characteristics of not just the regions of space that are occupied (the connected components) but also the empty or void regions that are represented as holes in the analysis. Numerous fruitful geospatial applications of the analysis of holes or voids can be envisaged, including the detection of regions of isolation or voidness in building density, cell net coverage, noise pollution, different types of vegetation and fauna, and demographic and socio-economic characteristics. Detection of these voids with persistent homology methods could be accompanied by various forms of analysis to understand or explain their presence with regard to the occurrence of other geographic phenomena, whether physical or social.

It is clear that there are many interesting challenges in the application of TDA to geospatial data, but we hope that this summary and review of some of the main methods will help to realise what we believe to be the considerable potential of TDA to advance the state of the geospatial data and information sciences.

5. Data and codes availability statement

The Python code used to perform the analyses presented in Sections 3.2 and 3.3 is openly available in the figshare repository at `https://doi.org/10.6084/m9.figshare.19521760`.

The city pub location data used in the analysis of Section 3.2 was obtained from OpenStreetMap. The code used to obtain this data is available in the above figshare repository.

The sequence of rainfall radar images used in the analysis of Section 3.3 was obtained from the UK Meteorological (Met) Office using their DataPoint service (`https://www.metoffice.gov.uk/datapoint`). The sequence in question is available in the above figshare repository.

Further information regarding the code and data is contained in the description of the above figshare repository.

6. Contributions

Padraig Corcoran contributed to all aspects of computer programming, data analysis and writing.

Christopher B. Jones contributed to all aspects of data analysis and writing.

7. Biographical sketch

Padraig Corcoran is a Senior Lecturer and Director of Research in the School of Computer Science and Informatics at Cardiff University. Dr Corcoran completed a PhD in computer science at Maynooth University in Ireland. He subsequently obtained a two year European Marie Curie International Outgoing Fellowship (IOF) which he completed at the Computer Science and Artificial Intelligence Laboratory (CSAIL) at Massachusetts Institute of Technology (MIT).

Christopher B. Jones is Professor of Geographical Information Systems at Cardiff University, having worked previously at the University of South Wales, the University of Cambridge, BP Exploration and the British Geological Survey. He studied geology at Imperial College and Bristol University and has a PhD on periodicities in fossil growth rhythms from the University of Newcastle upon Tyne. Current research focuses on geographical information retrieval and earlier research on cartography led to the development of the Maplex cartographic name placement software.

References

Adams, H., *et al.*, 2017. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18.

Ahmed, M., Fasy, B.T., and Wenk, C., 2014. Local persistent homology based distance between maps. *In*: *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 43–52.

Aktas, M.E., Akbas, E., and El Fatmaoui, A., 2019. Persistence homology of networks: methods and applications. *Applied Network Science*, 4 (1), 1–28.

Baddeley, A., Rubak, E., and Turner, R., 2015. *Spatial point patterns: methodology and applications with R*. CRC press.

Bendich, P., *et al.*, 2016. Persistent homology analysis of brain artery trees. *The annals of applied statistics*, 10 (1), 198.

Biasotti, S., *et al.*, 2008. Reeb graphs for shape analysis and applications. *Theoretical computer science*, 392 (1-3), 5–22.

Botnan, M. and Lesnick, M., 2018. Algebraic stability of zigzag persistence modules. *Algebraic & geometric topology*, 18 (6), 3133–3204.

Bubenik, P., *et al.*, 2015. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16 (1), 77–102.

Carlsson, G. and De Silva, V., 2010. Zigzag persistence. *Foundations of computational mathematics*, 10 (4), 367–405.

Carlsson, G., De Silva, V., and Morozov, D., 2009. Zigzag persistent homology and real-valued functions. *In*: *Proceedings of the twenty-fifth annual symposium on Computational geometry*. 247–256.

Carmody, D.R. and Sowers, R.B., 2021. Topological analysis of traffic pace via persistent homology. *Journal of Physics: Complexity*, 2 (2), 025007.

Carr, H., Snoeyink, J., and Axen, U., 2003. Computing contour trees in all dimensions. *Computational Geometry*, 24 (2), 75–94.

Carstens, C.J. and Horadam, K.J., 2013. Persistent Homology of Collaboration Networks. *Mathematical Problems in Engineering*, 2013, 1–7. Available from: `https://ideas.repec.org/a/hin/jnlmpe/815035.html`.

Chazal, F. and Michel, B., 2021. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, 4.

Chevyrev, I., Nanda, V., and Oberhauser, H., 2018. Persistence paths and signature features in topological data analysis. *IEEE transactions on pattern analysis and machine intelligence*, 42 (1), 192–202.

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J., 2007. Stability of persistence diagrams. *Discrete & computational geometry*, 37 (1), 103–120.

Corcoran, P., 2019a. Topological generalization of continuous valued raster data. *In*: *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 428–431.

Corcoran, P., 2019b. Topology based object tracking. *Mathematical and Computational Applications*, 24 (3), 84.

Corcoran, P. and Jones, C.B., 2016. Spatio-temporal modeling of the topology of swarm behavior with persistence landscapes. *In*: *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–4.

Corcoran, P. and Jones, C.B., 2017. Modelling topological features of swarm behaviour in space and time with persistence landscapes. *IEEE Access*, 5, 18534–18544.

Corcoran, P. and Jones, C.B., 2018. Robust tracking of objects with dynamic topology. *In*: *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 428–431.

Corcoran, P. and Jones, C.B., 2021. A persistent homology model of street network connectivity. *Transactions in GIS*, 26 (1), 155–181.

Corcoran, P. and Mooney, P., 2013. Characterising the metric and topological evolution of openstreetmap network representations. *The European Physical Journal Special Topics*, 215 (1), 109–122.

De Floriani, L., *et al.*, 2015. Morse complexes for shape segmentation and homological analysis: discrete models and algorithms. *In*: *Computer Graphics Forum*. Wiley Online Library, vol. 34, 761–785.

De Silva, V. and Ghrist, R., 2007. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7 (1), 339–358.

Dey, T.K., Wang, J., and Wang, Y., 2017. Improved road network reconstruction using discrete morse theory. *In*: *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–4.

Duchin, M., Needham, T., and Weighill, T., 2021. The (homological) persistence of gerrymandering. *Foundations of Data Science*, 0, –.

Edelsbrunner, H. and Harer, J., 2010. *Computational topology: an introduction*. American Mathematical Soc.

Egenhofer, M.J. and Franzosa, R.D., 1991. Point-set topological spatial relations. *International Journal of Geographical Information System*, 5 (2), 161–174.

Everitt, B.S., *et al.*, 2011. *Cluster analysis 5th ed.* West Sussex, UK: John Wiley.

Feng, M., Hickok, A., and Porter, M.A., 2022. Topological data analysis of spatial systems. *In*: *Higher-order systems*. Springer, 389–399.

Feng, M. and Porter, M.A., 2020. Spatial applications of topological data analysis: Cities, snowflakes, random structures, and spiders spinning under the influence. *Physical Review Research*, 2 (3), 033426.

Feng, M. and Porter, M.A., 2021. Persistent homology of geospatial data: A case study with voting. *SIAM Review*, 63 (1), 67–99.

Florczyk, A., *et al.*, 2019. Description of the GHS urban centre database 2015. (KJ-02-19-103-EN-N (online)). Available from: `https://ghsl.jrc.ec.europa.eu/ghs_stat_ucdb2015mt_r2019a.php`.

Hensel, F., Moor, M., and Rieck, B., 2021. A survey of topological machine learning methods.

Frontiers in Artificial Intelligence, 4, 52.

Jakubowski, A., Gasic, M., and Zibrowius, M., 2020. Topology of word embeddings: Singularities reflect polysemy. *In*: *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*. 103–113.

Jennings, M.D., 2000. Gap analysis: concepts, methods, and recent results. *Landscape ecology*, 15 (1), 5–20.

Jiang, J. and Worboys, M., 2009. Event-based topology for dynamic planar areal objects. *International Journal of Geographical Information Science*, 23 (1), 33–60.

Kerber, M., Morozov, D., and Nigmetov, A., 2017. Geometry helps to compare persistence diagrams. *ACM Journal of Experimental Algorithmics*, 22, 1–20.

Liu, H. and Schneider, M., 2010. Detecting the topological development in a complex moving region. *Journal of Multimedia Processing and Technologies*, 1 (3), 160–180.

Liu, H. and Schneider, M., 2011. Tracking continuous topological changes of complex moving regions. *In*: *Proceedings of the 2011 ACM Symposium on Applied Computing*. ACM, 833–838.

Longley, P.A., *et al.*, 2015. *Geographic information science and systems*. John Wiley & Sons.

Maaten, L.v.d. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9 (Nov), 2579–2605.

Menon, V.G. and Joe Prathap, P., 2016. Opportunistic routing with virtual coordinates to handle communication voids in mobile ad hoc networks. *In*: *Advances in signal processing and intelligent recognition systems*. Springer, 323–334.

Meyer, M., 2021. Tracking silence: place, embodiment, and politics. *In*: Kingsbury, P., Secor and A., eds. *Into the Void*. University of Nebraska Press, 103–117. Available from: `https://hal.archives-ouvertes.fr/hal-03457692`.

Nicolau, M., Levine, A.J., and Carlsson, G., 2011. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108 (17), 7265–7270.

Otter, N., *et al.*, 2017. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6, 1–38.

Pun, C.S., Xia, K., and Lee, S.X., 2018. Persistent-homology-based machine learning and its applications–a survey. *arXiv preprint arXiv:1811.00252*.

Randell, D.A., Cui, Z., and Cohn, A.G., 1992. A spatial logic based on regions and connection. *In*: *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, KR'92, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc., 165–176.

Robins, V. and Turner, K., 2016. Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Physica D: Nonlinear Phenomena*, 334, 99–117.

Schubert, E., *et al.*, 2017. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42 (3), 1–21.

Serrano, D.H., Hernández-Serrano, J., and Gómez, D.S., 2020. Simplicial degree in complex networks. applications of topological data analysis to network science. *Chaos, Solitons & Fractals*, 137, 109839.

Singh, G., *et al.*, 2007. Topological methods for the analysis of high dimensional data sets and 3d object recognition. *PBG@ Eurographics*, 2.

Sizemore, A.E., *et al.*, 2019. The importance of the whole: topological data analysis for the network neuroscientist. *Network Neuroscience*, 3 (3), 656–673.

Skraba, P. and Turner, K., 2020. Wasserstein stability for persistence diagrams. *arXiv preprint arXiv:2006.16824*.

Turner, K., Mukherjee, S., and Boyer, D.M., 2014. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3 (4), 310–344. Available from: `https://doi.org/10.1093/imaiai/iau011`.

Worboys, M.F. and Duckham, M., 2004. *Gis: a computing perspective*. CRC press.

Worboys, M. and Duckham, M., 2006. Monitoring qualitative spatiotemporal change for

geosensor networks. *International Journal of Geographical Information Science*, 20 (10), 1087–1108.

Wu, Y., *et al.*, 2017. Congestion barcodes: Exploring the topology of urban congestion using persistent homology. *In*: *IEEE International Conference on Intelligent Transportation Systems*. 1–6.

Zaheer, M., *et al.*, 2017. Deep sets. *In*: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds. *Advances in Neural Information Processing Systems*. Curran Associates, Inc., vol. 30. Available from: `https://proceedings.neurips.cc/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf`.

Zomorodian, A. and Carlsson, G., 2005. Computing persistent homology. *Discrete & Computational Geometry*, 33 (2), 249–274.