

# Stability and Statistical Inferences in the Space of Topological Spatial Relationships

PADRAIG CORCORAN<sup>1</sup> AND CHRISTOPHER B. JONES

School of Computer Science and Informatics, Cardiff University, Cardiff CF14 3UX, U.K.

Corresponding author: Padraig Corcoran (corcoranp@cardiff.ac.uk)

**ABSTRACT** Modeling topological properties of the spatial relationship between objects, known as the *topological relationship*, represent a fundamental research problem in many domains including artificial intelligence and geographical information science. Real-world data are generally finite and exhibit uncertainty. Therefore, when attempting to model topological relationships from such data, it is useful to do so in a manner which is both *stable* and facilitates *statistical inferences*. Current models of the topological relationships do not exhibit either of these properties. We propose a novel model of topological relationships between objects in the Euclidean plane, which encodes topological information regarding connected components and holes. Specifically, a representation of the persistent homology, known as a persistence scale space, is used. This representation forms a Banach space that is *stable* and, as a consequence of the fact that it obeys the *strong law of large numbers* and the *central limit theorem*, facilitates statistical inferences. The utility of this model is demonstrated through a number of experiments.

**INDEX TERMS** Spatial relationships, topology, stable, statistical inference.

## I. INTRODUCTION

There are many real world scenarios where one is required to model the spatial relationship between objects [1]. A meteorologist may wish to model the spatial relationship between different environmental variables toward understanding climate change [2]. A city planner may wish to model the spatial relationship between noise pollution produced by heavy traffic and residential areas. Models of spatial relationships have also been employed when designing intelligent robotic systems [3]. There exist many models of spatial relationships and these models typically focus exclusively on modelling metric, order or topological properties of the relationship in question [4]. Metric properties of a spatial relationship relate to things such as distance, size and orientation. Order properties of a spatial relationship relate to things such as the partial and total order of objects as described by prepositions such as *in front of*, *behind*, *above*, and *below* [5]. Finally, topological properties of a spatial relationship relate to things which are invariant under continuous transformations of the ambient space such as a homotopy. An indicator of whether or not two objects intersect is an example of a topological property. In this article we use the term *topological relationship* when referring to the topological properties of a given spatial relationship.

Real world data are generally finite and exhibit uncertainty. Therefore, when attempting to model topological relationships from such data it is useful to do so in a manner which is both stable and facilitates statistical inferences. Informally, a model is stable if a small change in the input data produces at most a small change in the resulting model [6]; a formal definition of Lipschitz stability, which is a type of stability, is provided in Appendix VI-A. Given the presence of data uncertainty, model stability is necessary for robustness. Statistics is an effective paradigm for making inferences given data regarding phenomena which cannot be directly observed. A model which facilitates statistical inferences with respect to topological relationships is useful in many contexts. For example, consider a sensor network with a small number of sensors where each is constantly moving and capable of detecting the presence or absence of different objects whose locations remain constant over time. Given the small size of such a network, at a given time, object locations and in turn topological relationships cannot be precisely modelled from sensor measurements. A solution to this problem would be to perform a statistical inference whereby the topological relationships in question are modelled at  $n$  distinct time steps, where each of these  $n$  models is considered to be an independent *sample* from the

*sampling distribution* of the model, and the *expected value* of these models is approximated. Such an approximation could be made using the sample mean if the model in question exhibited the statistical property of the *strong law of large numbers* where by the sample mean approaches the expected value as the number of samples increases. Current models of the topological relationships are not stable and do not exhibit those statistical properties necessary for performing many useful statistical inferences. Proposing a model of topological relationships which overcomes these limitations represents the theme of this article.

In this article we consider the problem of modelling topological relationships between objects in the ambient space  $\mathbb{R}^2$  given a finite set of points  $S$  in that space and the ability to detect the presence or absence of different objects at each of these points. The finite size of  $S$  results in uncertainty with respect to object locations and the degree of this uncertainty is a function of the size of  $S$ . This is equivalent to the sensor network problem discussed above where  $S$  corresponds to the set of sensors. We propose a novel model of topological relationships between objects which encodes topological information regarding connected components and holes. Specifically, a representation of the *persistent homology*, known as a *persistence scale space*, is used. This model forms a *Banach space* which is stable whereby a small change in the set  $S$ , as measured by the *Hausdorff distance*, produces at most a small change in the resulting model. This model also exhibits the following statistical properties which facilitate statistical inferences. It exhibits the *strong law of large numbers* described above. It also exhibits the *central limit theorem* whereby the distribution of the sample mean converges to a normal distribution centred at the expected value as the number of samples increases.

The layout of this article is as follows. In section II we review important works on modelling topological relationships. In section III the proposed model of topological relationships is presented. In section IV the accuracy and utility of this model is demonstrated through a number of experiments. Finally in section V conclusions from this work and possible future research directions are discussed.

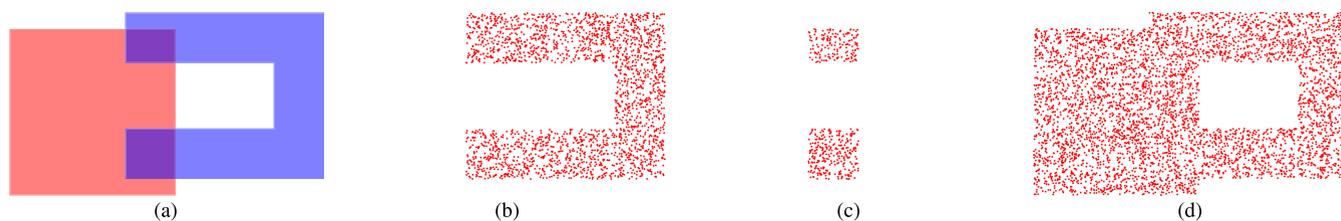
## II. RELATED WORKS

There exist many models of topological relationships with two of the most cited by a significant margin being the *Intersection Model* (IM) [7] and the *Region Connection Calculus* (RCC) [8]. These models assume object locations are known precisely and these are modelled as subsets of  $\mathbb{R}^2$ . In the IM, a number of subsets of the ambient space are considered where each is a binary set relation of object interiors, boundaries and exteriors. For example, one subset considered is the intersection of object interiors. Having determined these subsets the spatial relationship in question is modelled by evaluating whether or not these subsets equal the null set or their dimension. The success of the IM can be attributed in part to the fact that it models important topological features in a simple and interpretable manner.

In fact, many instances of the model correspond to spatial relationships which can be described using the natural language terms such as contains or disjoint; here a natural language term is an English language description which does not refer to binary set relations and mathematical topology concepts. For example, if the intersection of the interiors and boundaries of two objects is the null set the spatial relationship in question can be described using the natural language term disjoint. The IM is described in greater detail in Appendix VI-A where we also prove this model to be unstable. The RCC models topological relationships between objects as one of five or eight topological relationships such as disconnected (DC) and externally connected (EC). The IM and RCC models are in some sense equivalent; after accounting for physical constraints, there are exactly eight feasible instances of the IM and these correspond to the five or eight RCC topological relationships [9]. In their original form, both the RCC and IM assume the objects in question equal single connected components. A number of generalisations of the IM and RCC have been proposed which consider objects equalling one or more connected components [10].

As stated previously, when attempting to model topological relationships from data which exhibits uncertainty it is important to do so in a manner which is both stable and facilitates statistical inferences. Worboys [11] defined the following five factors which cause uncertainty in spatial data. Incompleteness due to lack of information; Inconsistency arising from conflicts in information; Vagueness resulting from objects not having crisp or sharp boundaries; Imprecision resulting from limits in the resolution at which measurements are made or stored; Errors which are a consequence of deviation from true values. If a spatial relationship is represented using natural language terms this may also result in uncertainty; for example, if a spatial relationship is represented using the terms *near* or *far* there is uncertainty with respect to the distance between the objects in question [12]. These latter terms correspond to vague qualifications of the topological relationship of disjoint.

Many models of topological relationships have been proposed which model uncertainty. Generally this is achieved by generalising the IM or RCC models in some way. For example [13]–[15] generalise these models using fuzzy set theory to offer robustness to vagueness. A number of generalisations of the IM and RCC have been proposed which model spatial uncertainty with respect to object locations. Tøssebro and Nygård [16] proposed a probabilistic model where a probability distribution is defined over the true location of each object. Clementini and Felice [17], [18], and Clementini [19] proposed to model objects using *broad boundaries* where an object boundary is represented by a region corresponding to all its possible locations. Bejaoui *et al.* [20] proposed to model each object using two components; one corresponding to the minimal and one corresponding to the maximal possible extent of the object. In each of these works the authors generalised the IM to model topological relationships between objects modelled in the manner in question.



**FIGURE 1.** Objects  $A$  and  $B$  are illustrated in (a) using the colours red and blue respectively. For a set  $S$  containing 6,000 points, the subset of the set of points contained in object  $B$  are illustrated in (b). The subsets of the points  $S$  contained in the intersection and union of  $A$  and  $B$  are illustrated in (c) and (d) respectively.

Cohn and Gotts [21] proposed a similar ‘egg-yolk’ model for objects and generalised the RCC model [8] to model topological relationships between objects modelled in this manner.

Although the models described above model spatial uncertainty with respect to object locations, they are not necessarily stable and do not consider the problem of performing statistical inferences. This does not mean that these models could not potentially be generalised to perform such inferences. For example, the metric refinements proposed by [22] could be used to measure the degree to which a property of a topological relationship exists.

A number of generalisations of the IM and RCC have been proposed which model topological relationships between objects whose locations, and in turn topological relationships, change as a function of time [23]–[25]. We do not consider this aspect of modelling topological relationships.

### III. MODEL OF TOPOLOGICAL RELATIONSHIPS

In this article we consider the following instance of the problem of modelling topological relationships. We assume the existence of two objects  $A$  and  $B$  in the ambient space  $\mathbb{R}^2$  for which we wish to model the corresponding topological relationship. Furthermore, we assume the maximum of the length and width of the axis aligned minimum bounding box containing both objects is equal to 10. The spatial locations of the objects is unknown and cannot be directly observed. Instead we assume a finite set  $S$  of points contained in the bounding box is known. Furthermore, we assume the ability to detect the presence or absence of the objects  $A$  and  $B$  at each of those points and, in turn, determine those subsets of  $S$  contained in  $A$  and  $B$ . This corresponds to performing *rejection sampling* of finite sets of points contained in  $A$  and  $B$ . Rejection sampling is a commonly used method for drawing samples from one space given the ability to draw samples from another [26]. To illustrate, consider the objects  $A$  and  $B$  illustrated in Figure 1(a) which form a running example in this article. For a given set  $S$  of size 6,000, the subset of  $S$  contained in the object  $B$  of Figure 1(a) is illustrated in Figure 1(b). The finite size of the set  $S$  results in uncertainty with respect to the locations of  $A$  and  $B$ . The degree of such uncertainty varies as a function of the size of  $S$ . If the size of  $S$  is small the size of the subsets contained in  $A$  and  $B$  is also small and there is in turn a greater degree of uncertainty

with respect to the locations of these objects. On the other hand, if the size of  $S$  is larger these object locations can be more precisely determined. Given this uncertainty, when attempting to model topological relationships from such data it is important to do so in a manner which is both stable and facilitates statistical inferences.

In this article we propose a model of topological relationships which goes some way toward achieving the above goal. The construction of the proposed model consists of the following three steps. In the first step a number of subsets of the set  $S$  are considered where each is a binary set relation of the objects  $A$  and  $B$ . This step is similar to the IM model although the subsets considered equal a finite number of points as opposed to subsets of  $\mathbb{R}^2$  which contain an infinite number of points. Unlike the IM model, the proposed model does not model the spatial relationship by evaluating whether or not these subsets equal the null set or their dimension. Instead, in the second step of the proposed model, the *persistent homology* of each subset is computed to give a corresponding *persistence diagram*. This is a representation which describes the topology of the subset in terms of the number of  $k$  dimensional holes it contains plus the range of scales across which these holes persist (note that a zero dimensional hole corresponds to a path connected component). In the third step, this information is in turn mapped to a function space representation known as a *persistence scale space* which forms a *Banach space* that is *stable*, obeys the *strong law of large numbers* and the *central limit theorem*. These properties are a consequence of the fact that this representation exploits the insight that those  $k$  dimensional holes which do not persist over a large range of scales are not statistically significant and can be considered *topological noise*. Fasy *et al.* [27] presents a formal definition of statistical significance in the context of persistence diagrams. Given these properties, the proposed model represents a more suitable platform for performing statistical inferences than existing models of topological relationships. Each of these steps is described in turn in the following three subsections.

#### A. SUBSETS

When modelling the topological relationship between objects  $A$  and  $B$  we do not wish to model global topological properties such as the number of  $k$  dimensional holes both objects contain. Instead, we wish to model topological properties

of the spatial relationship in question. Toward this goal, we propose to model the topological properties of subsets of the set  $S$  where each subset is a binary set relation of the objects in question. This corresponds to sampling from the subsets in question by performing *rejection sampling*. For example the intersection and union of those points contained in the objects of Figure 1(a) are illustrated in Figure 1(c) and Figure 1(d) respectively. The consideration of such subsets is motivated by the insight that the topological properties of these subsets model distinct topological properties of the spatial relationship in question. For example, if that subset corresponding to the binary set relation of intersection contains zero connected components, the spatial relationship in question may be described using the natural language term *disjoint*.

The specific subsets to consider depend on the topological properties one wishes to model and in turn the problem one wishes to solve. For example, if one wishes to perform a general clustering of topological relationships, a reasonable solution would be to consider a large number of subsets which capture a variety of topological properties and subsequently perform a clustering based on these properties. On the other hand, if one wishes to determine if a given topological relationship equals that corresponding to the natural language term *intersect*, a reasonable solution would be to consider the single subset corresponding to the binary set relation of intersection and determine whether or not this subset contains zero 0 dimensional holes. Similarly, if one wishes to determine if a given topological relationship equals that corresponding to the natural language term *contains*, a reasonable solution would be to consider the single subset corresponding to the binary set relation of exclusive or and determine whether or not this subset contains one 0 dimensional hole and one 1 dimensional hole. Determining necessary and sufficient conditions for instances of the proposed model to correspond to various natural language terms is beyond the scope of this work.

## B. PERSISTENCE DIAGRAM

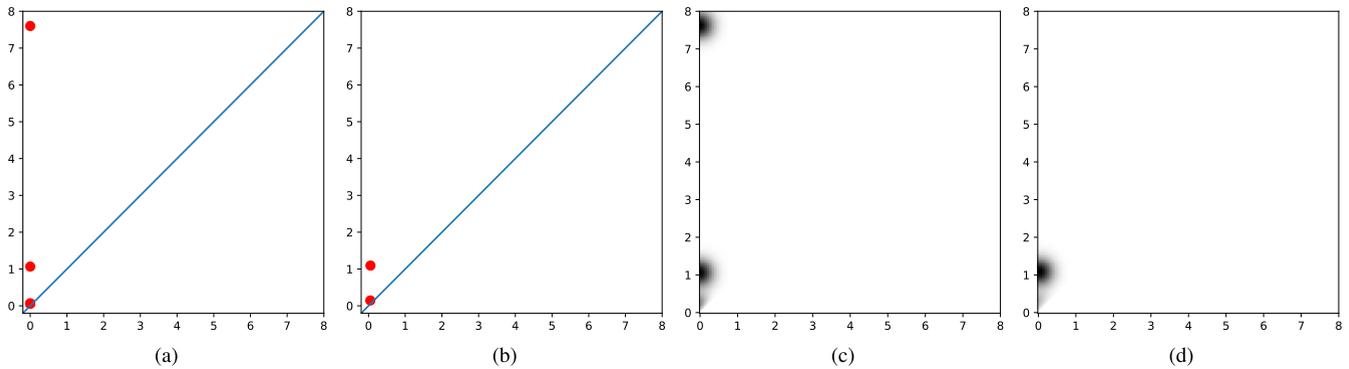
In this step the *persistent homology* of each subset is computed to give a corresponding set of *persistence diagrams*. This section only briefly introduces the concept of persistent homology and its computation. More technical details are contained in Appendix VI-B.

Each persistence diagram is a representation which describes the topology of the subset in terms of the number of  $k$  dimensional holes it contains plus the range of scales across which these holes persist. In this context scale corresponds to the radius of a set of balls centred at each point in the subset. As one increases the value of this radius holes may both appear and disappear. More formally, a persistence diagram for  $k$  dimensional holes is a multiset of points  $(i, j)$  in the space  $\{(i, j) \in \mathbb{R}^2, i \leq j\}$  where a point  $(i, j)$  indicates that a  $k$  dimensional hole appeared at scale  $i$  and disappeared at scale  $j$ . The disappearance of a  $k$  dimensional hole may be the consequence of its size becoming zero or it

merging with another  $k$  dimensional hole. The *persistence* of the  $k$  dimensional hole in question is the value  $j - i$ . If a  $k$  dimensional hole appears at scale  $i$  but does not disappear, it is represented in the persistence diagram by a point  $(i, u)$  where  $u$  is an upper bound on the scale at which  $k$  dimensional holes may disappear. Since we assume the maximum of the length and width of the axis aligned minimum bounding box containing both objects is equal to 10, we set the value of  $u$  equal to 7.6 for all  $k$  dimensional holes. This is a valid upper bound given the fact that the objects in question are scaled to be contained in a specified bounding box and a union of balls centered at the points is considered (see Appendix VI-B for details). It is important to note that some authors omit from persistence diagrams those points corresponding to  $k$  dimensional holes which do not disappear [28]. In the context of the current problem, it is important not to omit such points because doing so would remove the ability to differentiate between a number of important cases. This includes differentiating between a subset containing zero 0 dimensional holes and a subset containing one 0 dimensional holes.

The persistent homology computation is performed in two steps. In the first step the subset in question is represented using a combinatorial representation known as a *filtration*. The persistence diagrams are subsequently computed as a function of this filtration. The technical details of this computation are provided in Appendix VI-B. Since we assume the ambient space is  $\mathbb{R}^2$  it is only necessary to consider 0 and 1 dimensional holes where a 0 dimensional hole corresponds to a path connected component. That is, for each subset specified in section III-A two corresponding *persistence diagrams* are computed; one for 0 dimensional holes and one for 1 dimensional holes.

The persistence diagram corresponding to the 0 dimensional holes of that subset in Figure 1(c) is illustrated in Figure 2(a). In this diagram the  $x$ -axis represents the scale at which 0 dimensional holes appear while the  $y$ -axis represents the scale at which 0 dimensional holes disappear. Each point in the persistence diagram is represented by a red point above the diagonal which is in turn represented by a blue line. Note that points lying closer to the diagonal have lower persistence. This persistence diagram contains two points where the corresponding persistence values are relatively large. All other points in this persistence diagram lie close to the diagonal and therefore their corresponding persistence values are relatively small (note that, there are multiple points in this category but they all have similar coordinates and therefore appear to be a single point in the figure). These two significant points correspond to the two clusters in Figure 1(c). Note that, the point at location (0,1) disappears when the cluster it corresponds to disappears when it merges with the other cluster following increase in the scale value. Similarly, the persistence diagram corresponding to the 1 dimensional holes of that subset in Figure 1(d) is illustrated in Figure 2(b). This persistence diagram contains one point where the corresponding persistence value is relatively large. This point corresponds to the single large hole in Figure 1(d).



**FIGURE 2.** The persistence diagram corresponding to the 0 dimensional holes (connected components) of that subset in Figure 1(c) is illustrated in (a). The persistence diagram corresponding to the 1 dimensional holes of that subset in Figure 1(d) is illustrated in (b). The function space representations of the persistence diagrams in (a) and (b) are illustrated in (c) and (d) respectively. In these figures the x-axis represents the scale at which  $k$  dimensional holes appear while the y-axis represents the scale at which  $k$  dimensional holes disappear.

It is important to consider the persistence of  $k$  dimensional holes for two reasons. Firstly, the number of  $k$  dimensional holes and their persistence encodes important topological information regarding the subset in question and in turn the topological relationship in question. Consider the persistence diagram in Figure 2(a) corresponding to the 0 dimensional holes of that subset in Figure 1(c). This persistence diagram contains two points where their persistence values are relatively large. This is clearly important information regarding the topological relationship in question. It is important here to recall that the IM model only encodes if subsets equal the null set or their dimension. Secondly, the persistence diagram representation is stable whereby a small change in a given subset, as measured by the *Hausdorff distance*, produces at most a small change in the resulting persistence diagram, as measured by the *bottleneck distance* [6], [29].

**C. FUNCTION SPACE REPRESENTATION**

As stated in the introduction of this article, the goal of this research is the development of a model of topological relationships which is both stable and facilitates statistical inferences. The bottleneck and Wasserstein distance functions may be used to compute the distance between two persistence diagrams [30]. Both these distance functions are stable with respect to the locations of objects. However they do not provide a way of computing a mean persistence diagram. We therefore convert each persistence diagram into an alternative function space representation which facilitates this. Note that a function space is a space where the objects in that space are functions [31].

Let  $\mathcal{D}$  be the space of persistence diagrams. Let  $p = (b, d)$  denote a point in a persistence diagram  $F \in \mathcal{D}$  and  $\bar{p} = (d, b)$  denote its mirror image across the diagonal [32], [33]. Furthermore, let  $\Omega$  be the space  $\{x = (x_1, x_2) \in \mathbb{R}^2, x_2 \geq x_1\}$ . In this work we employ the map  $\Phi_\sigma : \mathcal{D} \rightarrow L^2(\Omega)$  defined in Equation 1. Here  $L^2(\Omega)$  is a Banach Space consisting of real-valued functions on  $\Omega$  [31]; that is, a vector space of real-valued functions on  $\Omega$  which is equipped with a norm. The norm in question is the  $L^2$ -norm which is denoted  $\|\cdot\|_2$ .

Conceptually this map places a Gaussian function at each point in the corresponding persistence diagram while suppressing those Gaussian functions which lie closer to the diagonal. The scale of the Gaussian functions in question is equal to the parameter  $\sigma$  which we set equal to the value 0.3. Figures 2(c) and 2(d) illustrate the result of applying the map in Equation 1 to those persistence diagrams illustrated in Figures 2(a) and 2(b) respectively.

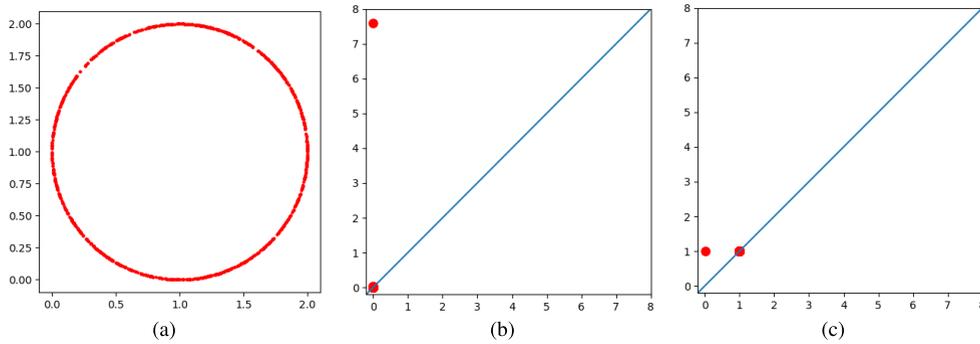
Although the persistence diagrams in Figures 2(a) and 2(b) encode the same information as the function space representations of Figures 2(c) and 2(d) respectively, they are different types of mathematical objects. A persistence diagram is a multiset of points while a function space representation is a function.

$$\Phi_\sigma(F) : \Omega \rightarrow \mathbb{R}, \quad x \mapsto \frac{1}{4\pi\sigma} \sum_{p \in F} e^{-\frac{\|x-p\|^2}{4\sigma}} - e^{-\frac{\|x-\bar{p}\|^2}{4\sigma}} \quad (1)$$

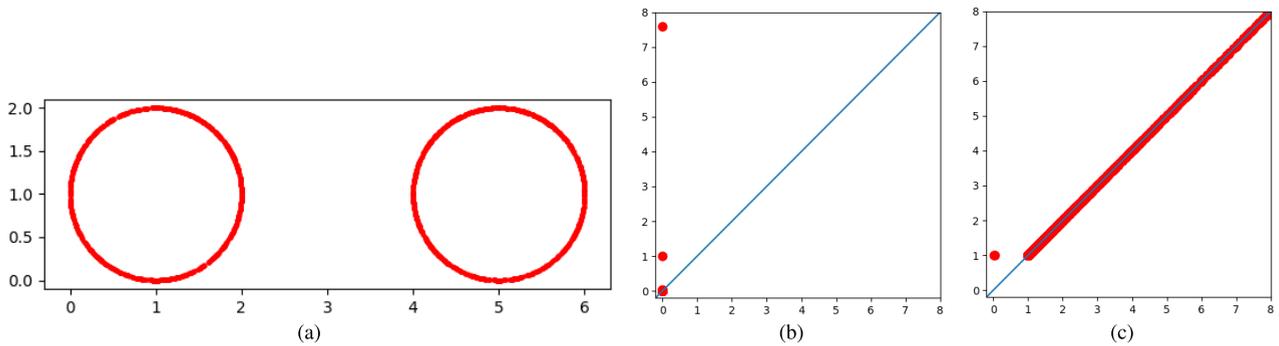
The map in Equation 1 was originally proposed by [32], [33]. The authors proved that the function space produced by this map facilitates statistical inferences. Specifically they proved the following three facts. This space is *stable* whereby a small change in the set  $S$  as measured by the *Hausdorff distance* produces at most a small change in the resulting model. It obeys the *strong law of large numbers*; that is, a sample mean converges almost surely to the expected value. Furthermore, this convergence obeys the *central limit theorem*; a reader unfamiliar with concepts relating to probability in a Banach space may consult [34] for details.

**IV. EXPERIMENTS**

In this section we present four experiments which demonstrate the accuracy of the proposed model of topological relationships and its ability to facilitate a number of important statistical inferences. In a first experiment we demonstrate the accuracy of the proposed model. In a second experiment we demonstrate the model can correctly infer a topological relationship given a set of samples from the sampling distribution of the model. In a third experiment we demonstrate the model



**FIGURE 3.** For the set of points illustrated in (a), the corresponding persistence diagrams for 0 and 1 dimensional holes are illustrated in (b) and (c) respectively.



**FIGURE 4.** For the set of points illustrated in (a), the corresponding persistence diagrams for 0 and 1 dimensional holes are illustrated in (b) and (c) respectively.

can perform a statistical test with the null hypothesis that two topological relationships are equal against the alternative hypothesis that they are different. In a final experiment we demonstrate the model can perform the data mining tasks of clustering and retrieval of similar topological relationships. These four experiments are described in turn in the following four subsections.

**A. MODEL ACCURACY**

This section demonstrates the accuracy of the proposed model with respect to representing the topology of a given set of points. This is achieved by considering sets of points for which accurate ground truth persistence diagrams can be inferred and comparing the persistence diagram computed by the proposed model to this ground truth.

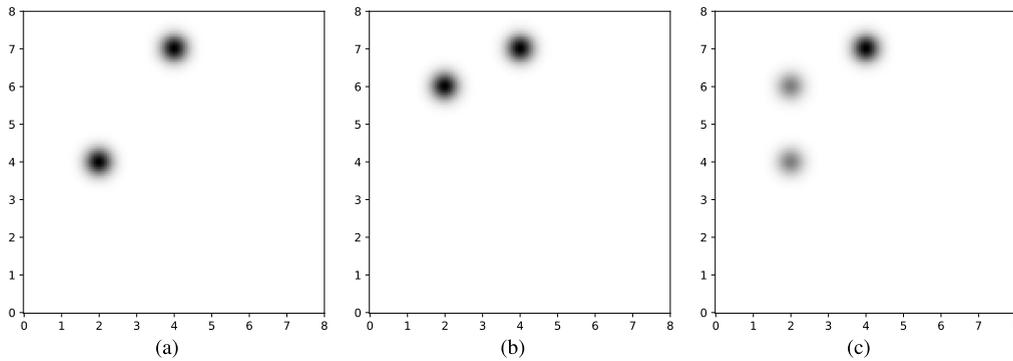
Consider the set of points illustrated in Figure 3(a) which equals 1,000 points uniformly sampled on a circle centred at the point (1, 1) with radius equal to 1. The corresponding persistence diagrams for 0 and 1 dimensional holes are illustrated in Figures 3(b) and 3(c) respectively. The persistence diagram for 0 dimensional holes accurately contains a single significant point at the location (0, 7.6) corresponding to the single significant connected component which never disappears. Here a point is determined significant if it does not lie close to the diagonal and, in turn, its persistence value is not close to zero. Note that, 7.6 is the value of an upper bound described in Section III-B. The persistence diagram for 1 dimensional holes accurately contains a single significant

point at the location (0, 1). The value of 1 in the second coordinate of this point accurately indicates that the hole formed by the circle disappears when the radius of the union of balls centred at the points is greater than or equal to the value 1.

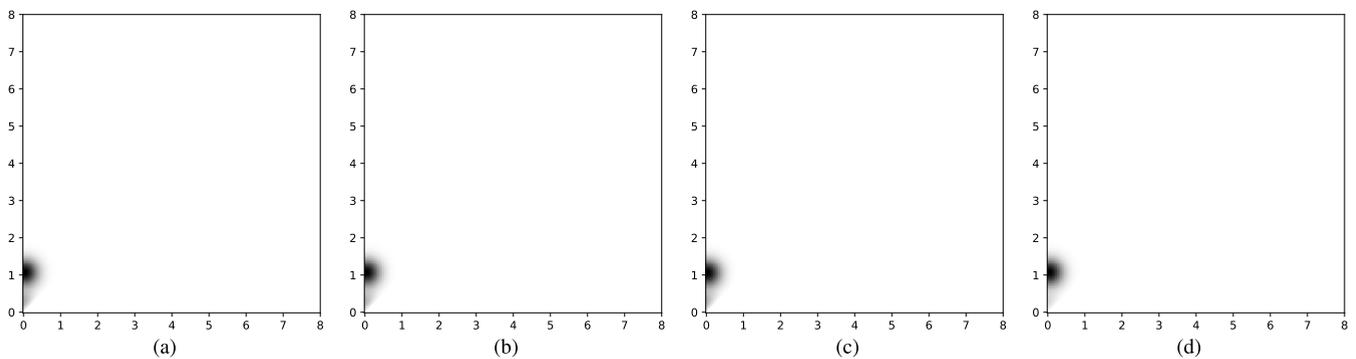
Next, consider the set of points illustrated in Figure 4(a) which equals 2,000 points uniformly sampled on two circles centred at the points (1, 1) and (5, 1) with radius equal to 1. The corresponding persistence diagrams for 0 and 1 dimensional holes are illustrated in Figures 4(b) and 4(c) respectively. The persistence diagram for 0 dimensional holes accurately contains two significant points at the locations (0, 1) and (0, 7.6) corresponding to the two significant connected components. The value of 1 in the second coordinate of the point (0, 1) accurately indicates that the two connected components merge when the radius of balls centred at the points is greater than or equal to the value 1. The persistence diagram for 1 dimensional holes accurately contains two significant points at the location (0, 1) corresponding to the two significant 1 dimensional holes of radius 1.

**B. INFERRING EXPECTED VALUE**

Consider the situation where one is attempting to model a topological relationship given a set of samples from the sampling distribution of the model. In this situation a reasonable solution is to approximate the expected value of the model from the samples. In the proposed model samples from the



**FIGURE 5.** Two function space representations are displayed in (a) and (b). The mean of these is displayed in (c). In these figures the x-axis represents the scale at which  $k$  dimensional holes appear while the y-axis represents the scale at which  $k$  dimensional holes disappear.



**FIGURE 6.** The function space representations of three samples drawn from a topological relationship are illustrated in (a)-(c). The function space representation of the mean of ten such samples is illustrated in (d). In these figures the x-axis represents the scale at which 1 dimensional holes appear while the y-axis represents the scale at which 1 dimensional holes disappear.

sampling distribution correspond to function space representations. Owing to the fact that the proposed model obeys the strong law of large numbers, the mean of such samples converges to its expected value. To illustrate this concept of a mean consider the two function space representations displayed in Figures 5(a) and 5(b). Note that both share a single *peak* in the same location. The mean of these function space representations is displayed in Figures 5(c). In this the height of the shared peak is maintained while the heights of the other peaks are suppressed.

To demonstrate that the proposed model facilitates the above inference of estimating the expected value consider again the topological relationship illustrated in Figure 1(a). For this relationship 10 independent sets  $\{S_1, \dots, S_{10}\}$  of points were sampled from the ambient space where each set contains 5,000 points.

For each  $S_i$  the persistence diagram  $F_i$  corresponding to the 1-homology group of the union of points contained in each object was computed. The mapping of Equation 1 was in turn applied to each  $F_i$  to give the corresponding function space representations  $(\Phi_\sigma(F_1), \dots, \Phi_\sigma(F_{10}))$ . Figures 6(a), 6(b) and 6(c) illustrate three of these representations.

The function space representations  $(\Phi_\sigma(F_1), \dots, \Phi_\sigma(F_{10}))$  correspond to samples from the sampling distribution of the model. Computing the mean of these samples reduces

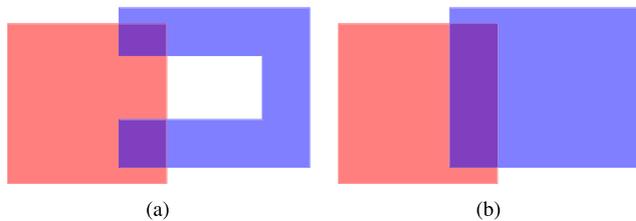
to applying the map of Equation 1 to the union of points in the persistence diagrams  $\{F_1, \dots, F_{10}\}$  and normalizing; that is,  $\frac{1}{10} \Phi_\sigma(F_1 \cup \dots \cup F_{10})$ . The result of this mapping is illustrated in Figure 6(d). Owing to the fact that the proposed model obeys the strong law of large numbers, this mapping converges to its expected value. Figure 6(d) implies that there exists a single one dimensional hole where the corresponding persistence value is relatively large. Examining the original topological relationship in Figure 1(a) demonstrates this to be correct; that is, the union of both objects contains a single large hole.

### C. TWO-SAMPLE HYPOTHESIS TEST

Given the fact that our model forms a Banach space, the norm in this space induces a metric. This metric may be used to measure the distance between samples from the sampling distributions of different topological relationships. Reference [35] proposed a method for computing the distance between two instances of the IM model. However this distance function returns integers in the interval  $[0, 8]$  and therefore only provides a very coarse measure of distance.

Given the distance between two samples from the sampling distributions of two distinct topological relationships, one can perform a statistical test with the null hypothesis that the topological relationships are equal against the alternative

hypothesis that they are different. This is known as the *two-sample problem* [34], [36]. There exist many contexts for which it is necessary to perform such a statistical test. For example, given two distinct meteorological phenomena such as two hurricanes, a meteorologist may want to perform a hypothesis test to determine if the topological relationship between the spatial extent of the hurricane and that of some other potentially related environmental phenomenon was the same in both cases.



**FIGURE 7.** A statistical test may be used to determine if the topological relationships illustrated are equal given a sample from each topological relationship.

In order to demonstrate such a statistical test we considered the topological relationships in Figures 7(a) and 7(b) and employed the *bootstrap hypothesis test* [37]. These topological relationships are different in the sense that the union of the objects in Figure 7(a) forms a single connected component with a single 1 dimensional hole while the union of the objects in Figure 7(b) forms a single connected component with zero 1 dimensional holes. Note that the IM model does not distinguish between these topological relationships.

Let  $S_a$  and  $S_b$  be single samples from the sampling distributions of the topological relationships in Figures 7(a) and 7(b) respectively where the cardinality of both sets is  $n$ . For  $S_a$  and  $S_b$  we computed the function space representations of their corresponding 1-homology groups. Recall that, the 1-homology group encodes the number of 1 dimensional holes present plus their persistence values. This particular homology group is appropriate for distinguishing between the topological relationships in question but may not be appropriate in other contexts. Using the metric induced by the norm of the function space representation, the distance between the function space representations in question was computed. Toward determining if this distance could be used to accept the null hypothesis (i.e. that the relationships are equal) we computed the *bootstrap distribution* under the null hypothesis [37]. That is, we sampled with replacement two sets  $S_a^1$  and  $S_b^2$  from  $S_a$  and computed the distance between the corresponding function space representations. This step was repeated 1000 times to form the bootstrap distribution. The null hypothesis was then rejected with  $p$ -value equal to the number of distances in the bootstrap distribution greater than the distance between  $S_a$  and  $S_b$ .

This hypothesis test was repeated for varying values of  $n$ ; that is, the cardinality of  $S_a$  and  $S_b$ . For  $n$  equal to 500, 1000, 2000 and 4000, the null hypothesis was rejected with a  $p$ -value of 0.21, 0.03, 0.01 and 0.00 respectively. This result

demonstrates that, given sufficient points in the ambient space, the proposed topological relation may be used to test the hypothesis that two topological relationships are equal.

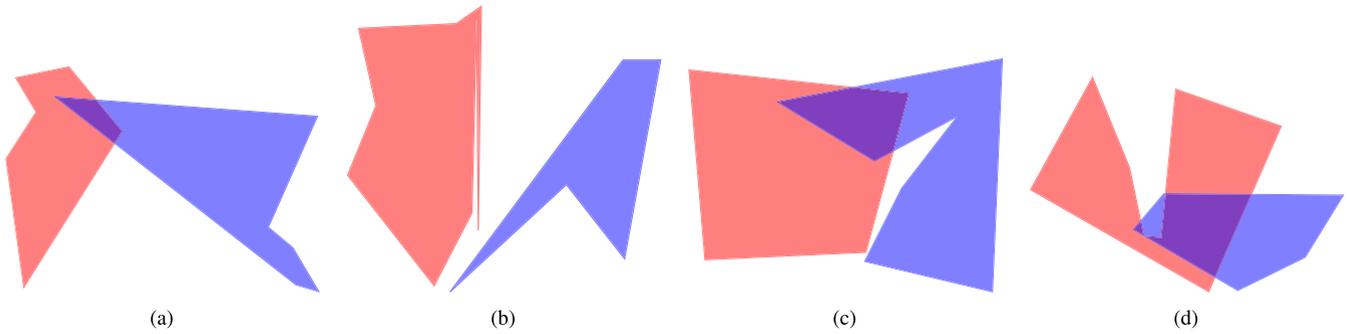
#### D. DATA MINING

Given that the proposed model allows distances between different topological relationships to be computed, one can use this model to perform a number of data mining tasks. In this section we describe how the proposed model may be used to perform clustering and retrieval of similar topological relationships. There are many contexts where performing such tasks would be useful. For example, consider a pair of environmental variables measured daily over an entire year. Toward summarisation, a meteorologist may be interested in detecting clusters of topological relationships which occurred daily between these variables. These clusters could in turn be used to explain broad weather conditions which occurred during the year in question.

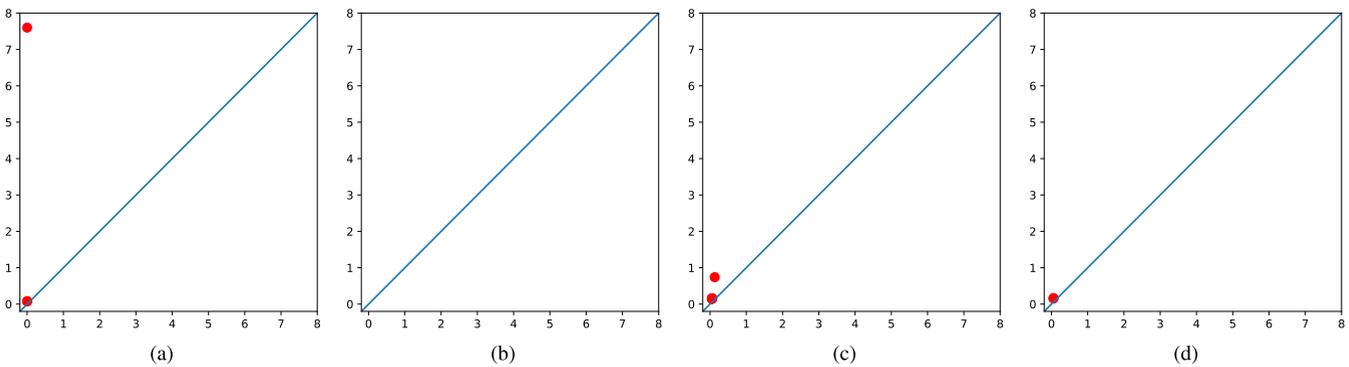
Toward performing clustering and retrieval we generated 200 sets  $S$  corresponding to distinct topological relationships using the following approach. First a pair of simple polygons corresponding to  $A$  and  $B$  as defined in section III were randomly generated using the *2-opt Moves* method of [38]. These polygons each contained between four and ten points and were initially centred at the coordinates (0, 0). The location of the polygon  $B$  was translated by adding a constant to the second coordinate of each of its points. Figure 8 displays four pairs of polygons generated using this approach. For each of these pairs of polygons, a corresponding set  $S$  was generated by uniformly sampling 6000 points from the bounding box containing  $A$  and  $B$ .

For each of the sets  $S$  generated the topological relationship in question was modelled using the proposed model. Specifically, we considered the binary set relations of union ( $A \cup B$ ), intersection ( $A \cap B$ ), symmetric difference or exclusive or ( $(A \setminus B) \cup (B \setminus A)$ ) and both relative complements ( $(A \setminus B)$  and  $(B \setminus A)$ ). Next, for each of these five subsets we computed the function space representations of their corresponding 0- and 1-homology groups. This gave ten individual function space representations for each topological relationship where each describes a different aspect of the topological relationship in question. To combine these spaces into a single space we formed the direct sum of the spaces with the direct sum norm [39]. Using this norm we computed the pair-wise distances between the 200 topological relationships and performed clustering using the  $k$ -medoids algorithm [40]. This algorithm is an iterative method which takes as input a single parameter  $k$  corresponding to the number of clusters and returns the cluster centres found. Here a cluster centre corresponds to a single representative, i.e. a pair of objects, of an entire cluster.

Figure 8 displays the pairs of polygons corresponding to the cluster centres found when the  $k$ -medoids algorithm was run with parameter  $k$  equal to 4. We can understand why our model determined those topological



**FIGURE 8.** Four pairs of simple polygons  $A$  and  $B$  generated using the method described in section IV-D are displayed. These pairs correspond to the 4 cluster centres found using the  $k$ -medoids algorithm.



**FIGURE 9.** The persistence diagrams corresponding to the connected components (0 dimensional holes) of the intersection subsets of the pairs of polygons in Figure 8(a) and 8(b) are illustrated in (a) and (b) respectively. The persistence diagrams corresponding to the 1 dimensional holes of the union subsets of the pairs of polygons in Figure 8(c) and 8(d) are illustrated in Figure (c) and (d) respectively. In these figures the x-axis represents the scale at which  $k$  dimensional holes appear while the y-axis represents the scale at which  $k$  dimensional holes disappear.

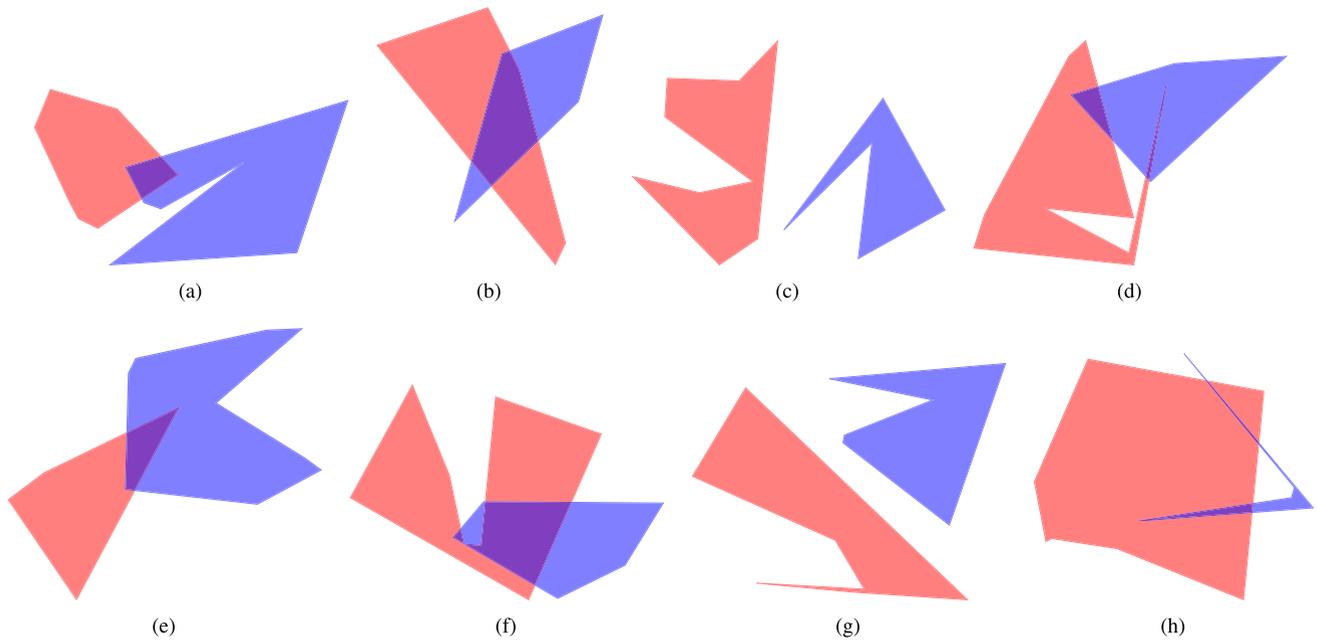
relationships to be significantly different and in turn belonging to different clusters by examining the corresponding persistence diagrams. The persistence diagrams corresponding to the connected components (i.e. the 0 dimensional holes) of the intersection subsets of the pairs of polygons in Figure 8(a) and 8(b) are illustrated in Figure 9(a) and 9(b) respectively. These persistence diagrams respectively contain one and zero points where the corresponding persistence values are relatively large. This is because the intersection of the pairs of polygons in Figure 8(a) and Figure 8(b) contain one and zero connected components respectively.

The persistence diagrams corresponding to the 1 dimensional holes of the union subsets of the pairs of polygons in Figure 8(c) and 8(d) are illustrated in Figure 9(c) and 9(d) respectively. These persistence diagrams respectively contain one and zero points where the corresponding persistence values are relatively large. This is because the union of the pairs of polygons in Figure 8(c) and Figure 8(d) contain one and zero holes respectively. Note that the hole in the union of the pair of polygons in Figure 8(c) appears once the scale of the radius of balls centred at each point is increased slightly.

The topological relationship in Figure 8(a) could be described as partial overlap. That in Figure 8(b) could be

described as disjoint. That in Figure 8(c) could be described as partial overlap with the formation of a one dimensional hole. Finally the topological relationship in Figure 8(d) could be described as partial overlap with each object splitting the other into two connected components.

Using the pair-wise distances between the 200 topological relationships described above, we performed retrieval of similar topological relationships as follows. For a given query topological relationship, that topological relationship with the smallest distance to the query was retrieved. The top row of Figure 10 displays four query topological relationships while the bottom row displays the corresponding retrieved topological relationships. It is evident that in each case the query and retrieved topological relationships are similar. The query and retrieved topological relationships in Figures 10(a) and 10(e) respectively could both be described as partial overlap. The query and retrieved topological relationships in Figures 10(b) and 10(f) respectively could both be described as partial overlap with each object splitting the other into two connected components. The query and retrieved topological relationships in Figures 10(c) and 10(g) respectively could both be described as disjoint. The query and retrieved topological relationships in Figures 10(d) and 10(h) respectively could



**FIGURE 10.** For each of the query topological relationships displayed in (a), (b), (c) and (d), the corresponding retrieved topological relationship is displayed in (e), (f), (g) and (h) respectively.

both be described as partial overlap with the formation of a one dimensional hole.

## V. CONCLUSIONS

This article proposes a novel model of topological relationships which is stable and exhibits a number of properties which facilitate statistical inferences. Existing models of topological relationships are not stable and do not consider the problem of performing statistical inferences. However, it must be noted that these models could potentially be generalised.

The proposed model formulates the problem of modelling topological relationships in terms of algebraic topology. This represents a novel formulation of the problem and consequently it presents many opportunities for further research and development. In the experiments section of this article we only consider objects corresponding to simple polygons. Without any adjustments to the model it can be applied to more general polygons such as polygons with multiple components and polygons with holes. With a slight generalisation, the model could also be applied to objects corresponding to lines and points. In this work we assume the ambient space to be a subset of  $\mathbb{R}^2$ . However the model proposed generalises to higher dimensional real coordinate spaces and more abstract spaces which can be embedded in  $\mathbb{R}^n$ .

The proposed model currently does not have any means of generating natural language descriptions of topological relationships in an automated manner. The current model outputs the topological relationships between pairs of objects in the form of persistence diagrams and function space representations that record the degree of significance of connected

components and holes. A challenge for future work is to develop automated approaches to natural language summarisation of the characteristics of the relationships such as the type (e.g. containment or overlap) and degree of applicability of a relationship [41].

Although in this article we have only applied the proposed model to synthetic data it could also be applied to real world data such as that from sensor networks. In future work the authors hope to pursue research on such applications.

## VI. APPENDIX

### A. INTERSECTION MODEL

This section describes the Intersection Model (IM) by [7] and presents some analysis of this model. The IM assumes the existence of two objects  $A$  and  $B$  in the ambient space  $\mathbb{R}^2$  for which we wish to model the corresponding topological relationship. Furthermore, this model assumes object locations are known precisely and these are modelled as subsets of  $\mathbb{R}^2$ . A number of subsets of the ambient space are considered where each is a binary set relation of object interiors, boundaries and exteriors. The subsets in question are defined in Equation 2 where  $A^\circ$ ,  $A^c$  and  $\partial A$  equal the interior, exterior and boundary of the set  $A$  respectively. Given the above subsets, the spatial relationship in question is modelled by evaluating whether or not these subsets equal the null set or their dimension. Egenhofer [7] presented an in-depth analysis of which instances of this model correspond to natural language terms. In this section we consider the version of the IM which evaluates whether or not these subsets equal the null set.

We refer to these Boolean valued functions as the features of the IM.

$$A^o \cap B^o \tag{2a}$$

$$A^o \cap \partial B \tag{2b}$$

$$A^o \cap B^e \tag{2c}$$

$$\partial A \cap B^o \tag{2d}$$

$$\partial A \cap \partial B \tag{2e}$$

$$\partial A \cap B^e \tag{2f}$$

$$A^e \cap B^o \tag{2g}$$

$$A^e \cap \partial B \tag{2h}$$

$$A^e \cap B^e \tag{2i}$$

Let  $(X, d_X)$  and  $(Y, d_Y)$  be metric spaces where  $X$  and  $Y$  are sets while  $d_X$  and  $d_Y$  are metrics on these sets respectively. A function  $f : X \rightarrow Y$  is Lipschitz stable with constant  $K$  if for all  $x_1$  and  $x_2$  in  $X$  the inequality of Equation 3 is satisfied [6]. Broadly speaking, a function is Lipschitz stable if a small change in the function input produces a small change in the function output.

$$d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2) \tag{3}$$

We now set about proving the IM is not Lipschitz stable. Let  $X$  be the space of 2-tuples of subsets of  $\mathbb{R}^2$ . Let  $d_X$  be the metric on  $X$  defined in Equation 4. Each of the terms being subtracted on the right side of this equality is a non-negative real valued function and, given this, it is straight forward to prove that  $d_X$  is a metric.

$$d_X((A_1, B_1), (A_2, B_2)) = \left| \inf_{a \in A_1^o, b \in B_1^o} \|a - b\|_2 - \inf_{a \in A_2^o, b \in B_2^o} \|a - b\|_2 \right| \tag{4}$$

Let  $Y$  be the space of Boolean values and let  $f$  be the mapping from  $X$  to  $Y$  defined in Equation 5. Note that,  $\inf_{a \in A^o, b \in B^o} \|a - b\|_2 = 0$  if and only if  $A^o \cap B^o \neq \emptyset$ . Furthermore,  $\inf_{a \in A^o, b \in B^o} \|a - b\|_2 > 0$  if and only if  $A^o \cap B^o = \emptyset$ . Let  $d_Y$  be the discrete metric on  $Y$  defined in Equation 6. The mapping  $f$  in Equation 5 is Lipschitz stable if there exists a real constant  $K$  which satisfies the inequality defined in Equation 7 for all  $(A_1, B_1), (A_2, B_2)$  in  $X$ .

$$f : (A, B) \rightarrow A^o \cap B^o \neq \emptyset \tag{5}$$

$$d_Y(f((A_1, B_1)), f((A_2, B_2))) = \begin{cases} 0, & \text{if } f((A_1, B_1)) = f((A_2, B_2)) \\ 1, & \text{otherwise} \end{cases} \tag{6}$$

$$d_Y(f((A_1, B_1)), f((A_2, B_2))) \leq K d_X((A_1, B_1), (A_2, B_2)) \tag{7}$$

*Theorem 1: The mapping  $f$  in Equation 5 is not Lipschitz stable.*

*Proof:* We prove this theorem using proof by contradiction. Assume there exists a real constant  $K$  which satisfies the inequality defined in Equation 7 for all  $(A_1, B_1), (A_2, B_2)$  in  $X$ . Let  $(A_1, B_1)$  be two sets such that  $(A_1^o \cap B_1^o \neq \emptyset)$  and in turn that  $\inf_{a \in A_1^o, b \in B_1^o} \|a - b\|_2 = 0$ . Let  $(A_2, B_2)$  be two sets such that  $(A_2^o \cap B_2^o = \emptyset)$  and in turn that

$\inf_{a \in A_2^o, b \in B_2^o} \|a - b\|_2 > 0$ . In this case Equation 7 can be written as Equation 8a where the expressions on the left and right sides of this inequality follow from evaluating Equations 6 and 4 respectively. Equation 8a is simplified in Equations 8b and 8c where the first simplification follows from the fact that the term  $\inf_{a \in A_2^o, b \in B_2^o} \|a - b\|_2$  is bounded below by 0.

$$1 \leq K |0 - \inf_{a \in A_2^o, b \in B_2^o} \|a - b\|_2| \tag{8a}$$

$$1 \leq K \inf_{a \in A_2^o, b \in B_2^o} \|a - b\|_2 \tag{8b}$$

$$\frac{1}{K} \leq \inf_{a \in A_2^o, b \in B_2^o} \|a - b\|_2 \tag{8c}$$

Equation 8c gives a lower bound for the term on the right side of this inequality. One can construct an example where this term lies in the open interval  $(0, 1/K)$  which in turn implies that  $A_2^o \cap B_2^o = \emptyset$ . This contradicts the assumption that there exists a real constant  $K$  which satisfies the inequality defined in Equation 7 for all  $(A_1, B_1), (A_2, B_2)$  in  $X$ . ■

*Corollary 1: The Intersection Model (IM) is not Lipschitz stable.*

*Proof:* The mapping  $f$  in Equation 5 is one of the features of the IM. This mapping was proven to not be Lipschitz stable in Theorem 1. Therefore, the IM is not Lipschitz stable. ■

## B. PERSISTENT HOMOLOGY

For a given subset considered in section III-A the corresponding persistent homology computation is performed in two steps. In the first step the subset in question is represented using a combinatorial representation known as a *filtration*. The persistence diagrams are subsequently computed as a function of this filtration. We now describe each of these steps in turn in the following subsections.

### 1) FILTRATION

Let  $X$  be a given subset considered in section III-A for which we wish to compute the persistent homology. Let  $\|\cdot\|$  denote the Euclidean norm and for each  $x \in X$  let  $B_r(x) = \{y \in \mathbb{R}^n \mid \|x - y\| < r\}$  for  $r \geq 0$ ; that is, a closed ball of radius  $r$  centred at  $x$ . The  $r$ -neighbourhood  $X_r$ , as defined by Equation 9, represents an intuitive means of representing the subset  $X$ . However this corresponds to an abstract mathematical representation of a continuous object upon which computation is difficult. Therefore, one generally instead uses a combinatorial representation known as a simplicial complex upon which computations may be performed.

$$X_r = \bigcup_{i=1}^m B_r(x_i) \tag{9}$$

An (abstract) simplicial complex  $\mathcal{K}$  is a finite collection of sets such that for each  $\sigma \in \mathcal{K}$  all subsets of  $\sigma$  are also contained in  $\mathcal{K}$ . Each element  $\sigma \in \mathcal{K}$  is called a simplex or  $k$ -simplex where  $|\sigma| = k + 1$  and is referred to as the dimension of the simplex. The faces of a simplex  $\sigma$  correspond to all

simplices  $\tau$  where  $\tau \subset \sigma$ . There exists a number of different simplicial complexes which may be used to represent a set of points [30]. For the purposes of this work we use a specific simplicial complex which is known as an *alpha complex* and is now described.

For  $x \in X$  let  $V_x$  be the Voronoi cell of  $x$ ; that is,  $V_x = \{y \in \mathbb{R}^n \mid \|y - x\| \leq \|y - z\|, z \in X\}$ . Furthermore, let  $R_x(r)$  be the intersection of each Voronoi cell with a ball centred at the point in question; that is,  $R_x(r) = B_r(x) \cap V_x$ . The alpha complex is isomorphic to the nerve of this cover and is defined by Equation 10. It is homotopy equivalent to the  $r$ -neighbourhood  $X_r$  where homotopy equivalence is an equivalence relation on the class of topological spaces (see [42, Ch. 7]).

$$\text{Alpha}(r) = \{\sigma \subseteq X \mid \bigcap_{y \in \sigma} R_y(r) \neq \emptyset\} \quad (10)$$

The parameter  $r$  in Equation 10 may be varied to give alpha complexes of different scales and in turn a 1-parameter family of nested alpha complexes. However only finitely many of these are distinct and are described by the sequence in Equation 11 which is called a *filtration* [30].

$$\emptyset = \mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \dots \subseteq \mathcal{K}_l \quad (11)$$

## 2) PERSISTENT HOMOLOGY

Let  $\mathcal{K}$  be a simplicial complex. The formal sum  $c$  defined by Equation 12 is called a  $k$ -chain where each  $\sigma_i \in \mathcal{K}$  is a  $k$ -simplex and each  $\lambda_i$  is an element of a given field. For the purposes of this work we consider the field  $\mathbb{Z}_2$  [43].

$$c = \sum \lambda_i \sigma_i \quad (12)$$

The vector space of all  $k$ -chains is denoted  $C_k(\mathcal{K})$ . The boundary of a  $k$ -simplex  $\sigma = [v_1, \dots, v_{k+1}]$  is a sum of its  $(k - 1)$ -dimensional faces and is defined by Equation 13 where  $\hat{v}_i$  indicates the deletion of  $v_i$  from the sequence. The boundary of a  $k$ -chain is obtained by extending this map linearly.

$$\partial_k \sigma = \sum_{i=1}^{k+1} [v_1, \dots, \hat{v}_i, \dots, v_{k+1}] \quad (13)$$

A  $k$ -chain  $c$  is a  $k$ -boundary if there exists some  $k + 1$ -chain  $d$  such that  $c = \partial d$  and a  $k$ -cycle if  $\partial c = 0$ . The set of all  $k$ -boundaries and  $k$ -cycles are denoted by  $B_k(\mathcal{K})$  and  $Z_k(\mathcal{K})$  respectively. The fact that  $\partial_{k+1} \partial_k = 0$  implies that  $B_k(\mathcal{K}) \subseteq Z_k(\mathcal{K})$ . The quotient group  $H_k(\mathcal{K}) = Z_k(\mathcal{K})/B_k(\mathcal{K})$  is the  $k$ -homology group of  $\mathcal{K}$ . Intuitively an element of the  $k$ -homology group corresponds to a  $k$ -dimensional hole in  $\mathcal{K}$ . That is, an element of the 0-homology group corresponds to a path connected component in  $\mathcal{K}$  while an element of the 1-homology group corresponds to a one dimensional hole in  $\mathcal{K}$ . The rank of  $H_k(\mathcal{K})$  is called the  $k$ -th Betti number and is denoted  $\beta_k(\mathcal{K})$ .

For a given filtration, for every  $i \leq j$  there exists an inclusion map from  $\mathcal{K}_i$  to  $\mathcal{K}_j$  and in turn an induced homomorphism from  $H_k(\mathcal{K}_i)$  to  $H_k(\mathcal{K}_j)$  for each dimension  $k$ . When

moving from  $\mathcal{K}_i$  to  $\mathcal{K}_{i+1}$  the corresponding homology groups may change in the following two ways. A homology group element *appears* at  $\mathcal{K}_{i+1}$  if it exists in  $H_k(\mathcal{K}_{i+1})$  but does not exist in  $H_k(\mathcal{K}_i)$ ; that is  $\beta_k(\mathcal{K}_{i+1}) = \beta_k(\mathcal{K}_i) + 1$ . A homology group element *disappears* at  $\mathcal{K}_{i+1}$  if it exists in  $H_k(\mathcal{K}_i)$  but does not exist in  $H_k(\mathcal{K}_{i+1})$ ; that is  $\beta_k(\mathcal{K}_{i+1}) = \beta_k(\mathcal{K}_i) - 1$ . If an element of a homology group never disappears, we assign its disappearance to be at  $\mathcal{K}_u$  where  $u$  is an upper bound. In most abstract mathematical publications a value of  $\infty$  is used instead of an upper bound. However in terms of algorithm implementation it is not feasibly to consider such a value.

If a homology group element appears at  $\mathcal{K}_i$  and disappears at  $\mathcal{K}_j$  we represent it as a point  $(i, j)$  in the space  $\{(i, j) \in \mathbb{R}^2, i \leq j\}$  with corresponding *persistence* value of  $j - i$ . The magnitude of a persistence value is important because homology group elements which persist over a larger range in a given filtration are considered to be of greater significance. In the context of this work the range in question is over the scale parameter  $r$  in the Alpha complex. The multiset of points corresponding to a  $k$ -homology group is called a *persistence diagram*. In this work the method described in [44] was used to compute all persistence diagrams.

## REFERENCES

- [1] P. Corcoran, P. Mooney, and M. Bertolotto, "Spatial relations using high level concepts," *ISPRS Int. J. Geo-Inf.*, vol. 1, no. 3, pp. 333–350, 2012.
- [2] R. W. Portmann, S. Solomon, and G. C. Hegerl, "Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 18, pp. 7324–7329, 2009.
- [3] A. Boularias, F. Duvallet, J. Oh, and A. Stentz, "Grounding spatial relations for outdoor robot navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2015, pp. 1976–1982.
- [4] M. J. Egenhofer and R. D. Franzosa, "Point-set topological spatial relations," *Int. J. Geograph. Inf. Syst.*, vol. 5, no. 2, pp. 161–174, 1991.
- [5] D. Hernandez, *Qualitative Representation of Spatial Knowledge*, vol. 804. New York, NY, USA: Springer, 1994.
- [6] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discrete Comput. Geometry*, vol. 37, no. 1, pp. 103–120, 2007.
- [7] M. J. Egenhofer, "Reasoning about binary topological relations," in *Proc. Symp. Spatial Databases*, 1991, pp. 141–160.
- [8] D. A. Randell, Z. Cui, and A. G. Cohn, "A spatial logic based on regions and connection," in *Proc. 3rd Int. Conf. Knowl. Represent. Reasoning*, 1992, pp. 165–176.
- [9] J. Chen, A. G. Cohn, D. Liu, S. Wang, J. Ouyang, and Q. Yu, "A survey of qualitative spatial representations," *Knowl. Eng. Rev.*, vol. 30, no. 1, pp. 106–136, 2015.
- [10] M. Schneider and T. Behr, "Topological relationships between complex spatial objects," *ACM Trans. Database Syst.*, vol. 31, no. 1, pp. 39–81, 2006.
- [11] M. Worboys, "Imprecision in finite resolution spatial data," *Geoinformatica*, vol. 2, no. 3, pp. 257–279, 1998.
- [12] X. Yao and J.-C. Thill, "Spatial queries with qualitative locations in spatial information systems," *Comput., Environ. Urban Syst.*, vol. 30, no. 4, pp. 485–502, 2006.
- [13] J. T. Björke, "Topological relations between fuzzy regions: Derivation of verbal terms," *Fuzzy Sets Syst.*, vol. 141, no. 3, pp. 449–467, 2004.
- [14] S. Schockaert, M. De Cock, and E. E. Kerre, "Spatial reasoning in a fuzzy region connection calculus," *Artif. Intell.*, vol. 173, no. 2, pp. 258–298, 2009.
- [15] A. C. Carniel, M. Schneider, R. R. Ciferri, and C. D. de Aguiar Ciferri, "Modeling fuzzy topological predicates for fuzzy regions," in *Proc. 22nd ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2014, pp. 529–532.
- [16] E. Tøssebro and M. Nygård, "An advanced discrete model for uncertain spatial data," in *Proc. Int. Conf. Web-Age Inf. Manage.*, 2002, pp. 37–51.

- [17] E. Clementini and P. Di Felice, "Approximate topological relations," *Int. J. Approx. Reasoning*, vol. 16, no. 2, pp. 173–204, 1997.
- [18] E. Clementini and P. Di Felice, "A spatial model for complex objects with a broad boundary supporting queries on uncertain data," *Data Knowl. Eng.*, vol. 37, no. 3, pp. 285–305, 2001.
- [19] E. Clementini, "A model for uncertain lines," *J. Vis. Lang. Comput.*, vol. 16, no. 4, pp. 271–288, 2005.
- [20] L. Bejaoui, F. Pinet, Y. Bedard, and M. Schneider, "Qualified topological relations between spatial objects with possible vague shape," *Int. J. Geograph. Inf. Sci.*, vol. 23, no. 7, pp. 877–921, 2009.
- [21] A. G. Cohn and N. M. Gotts, "The 'egg-yolk' representation of regions with indeterminate boundaries," in *Geographic Objects With Indeterminate Boundaries*, vol. 2. London, U.K.: Taylor & Francis, 1996, pp. 171–187.
- [22] M. J. Egenhofer and M. P. Dube, "Topological relations from metric refinements," in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2009, pp. 158–167.
- [23] H. Liu and M. Schneider, "Tracking continuous topological changes of complex moving regions," in *Proc. ACM Symp. Appl. Comput.*, 2011, pp. 833–838.
- [24] M.-H. Jeong and M. Duckham, "Decentralized querying of topological relations between regions monitored by a coordinate-free geosensor network," *Geoinformatica*, vol. 17, no. 4, pp. 669–696, 2013.
- [25] P. Corcoran and C. B. Jones, "Modelling topological features of swarm behaviour in space and time with persistence landscapes," *IEEE Access*, vol. 5, pp. 18534–18544, 2017.
- [26] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [27] B. T. Fasy *et al.*, "Confidence sets for persistence diagrams," *Ann. Statist.*, vol. 42, no. 6, pp. 2301–2339, 2014.
- [28] H. Adams *et al.*, "Persistence images: A stable vector representation of persistent homology," *J. Mach. Learn. Res.*, vol. 18, no. 8, pp. 1–35, 2017.
- [29] H. Edelsbrunner and D. Morozov, "Persistent homology: Theory and practice," Ernest Orlando Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Tech. Rep. LBNL-6037E, 2012.
- [30] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*. Providence, RI, USA: American Mathematical Society, 2010.
- [31] O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*. Cambridge, MA, USA: Springer, 2010.
- [32] J. Reininghaus, S. Huber, U. Bauer, and R. Kwitt, "A stable multi-scale kernel for topological machine learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4741–4748.
- [33] R. Kwitt, S. Huber, M. Niethammer, W. Lin, and U. Bauer, "Statistical topological data analysis—a kernel perspective," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3070–3078.
- [34] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 77–102, 2015.
- [35] M. Egenhofer and K. Al-Taha, "Reasoning about gradual changes of topological relationships," *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*. Berlin, Germany: Springer, 1992, pp. 196–219.
- [36] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, pp. 723–773, Mar. 2012.
- [37] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL, USA: CRC press, 1994.
- [38] M. Heidi, "Heuristics for the generation of random polygons," in *Proc. Can. Conf. Comput. Geometry*, vol. 5. 1996, p. 38.
- [39] R. E. Megginson, *An Introduction to Banach Space Theory*, vol. 183. Springer, 2012.
- [40] P. Corcoran and C. B. Jones, "Spatio-temporal modeling of the topology of swarm behavior with persistence landscapes," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2016, pp. 1–4.
- [41] A. R. B. Shariff, "Natural-language spatial relations between linear and areal objects: The topology and metric of english-language terms," *Int. J. Geograph. Inf. Sci.*, vol. 12, no. 3, pp. 215–245, 1998.
- [42] J. Lee, *Introduction to Topological Manifolds*, vol. 453. New York, NY, USA: Springer, 2010.
- [43] H. Edelsbrunner and J. Harer, "Persistent homology—A survey," *Contemp. Math.*, vol. 453, pp. 257–282, Feb. 2008.
- [44] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete Comput. Geometry*, vol. 33, no. 2, pp. 249–274, 2005.



**PADRAIG CORCORAN** is currently a Lecturer with the School of Computer Science and Informatics, Cardiff University. He received the European Marie Curie International Outgoing Fellowship. During the outgoing phase of this fellowship, he was with the Massachusetts Institute of Technology; while during the incoming phase, he was with University College Dublin.



**CHRISTOPHER B. JONES** received the B.Sc. degree in geology from the University of Bristol and the Ph.D. degree from the School of Physics, Newcastle University, Newcastle upon Tyne, with a focus on the subject of periodicities in fossil growth rhythms. He is currently a Professor of geographical information systems with the School of Computer Science and Informatics, Cardiff University. He has also held academic positions, in computer science, at the University of South Wales

and, in geography, at the University of Cambridge. He was involved in software development with the British Geological Survey and BP Exploration. His research on cartographic label placement led to the development of the Maplex software, which is currently an ESRI product.

• • •