

Modelling Topological Features of Swarm Behaviour in Space and Time with Persistence Landscapes

Padraig Corcoran and Christopher B. Jones

Abstract—This paper presents a model of swarm behaviour that encodes the spatial-temporal characteristics of topological features such as holes and connected components. Specifically, the persistence of topological features with respect to time are computed using *zig-zag persistent homology*. This information is in turn modelled as a *persistence landscape* which forms a normed vector space and facilitates the application of statistical and data mining techniques. Validation of the proposed model is performed using a real data set corresponding to a swarm of fish. It is demonstrated that the proposed model may be used to perform retrieval and clustering of swarm behaviour in terms of topological features. In fact, it is discovered that clustering returns clusters corresponding to the swarm behaviours of *flock*, *torus* and *disordered*. These are the most frequently occurring types of behaviour exhibited by swarms in general.

Index Terms—Swarm Behaviour, Topology, Spatial-Temporal, Persistence Landscape

I. INTRODUCTION

A swarm is defined as a set of agents moving in close spatial proximity to each other. The set of agents may be robots [1] or animals of a single type, such as birds, fish or humans [2]. It has been demonstrated that swarms can accomplish complex tasks such as foraging [3], building complex structures [4] and navigation [5]. Furthermore, swarms can accomplish such tasks in a manner that is flexible, robust and scalable [6]. Given these attractive properties, developing accurate models of swarm behaviour is an active area of research.

To date a number of models of swarm behaviour have been proposed. For example, Tunström et al. [7] proposed to model swarm behaviour in terms of rotation order and polarization which measure the angular momentum and alignment of the agents respectively. Current models of swarm behaviour typically consider metric properties of the swarm such as the mean orientation of the agents. If one assumes the swarm to be samples from a topological space, one can infer this space using, for example, kernel density estimation. The topological features of this space, and in turn those of the swarm, can then be modelled. A fundamental approach to modelling the topological features of a topological space is to compute its corresponding Betti numbers. The p^{th} Betti number of a topological space intuitively equals the number of p -dimensional holes which it contains. Note that the 0^{th} Betti number of a topological space equals the number of

path-connected components it contains [8]. To illustrate this approach to modelling swarm behaviour consider the swarm of 300 Golden Shiner fish illustrated in Fig. 1(a) [7]. The fish in question are swimming in a shallow pool and therefore their positions can be specified using only the x and y Cartesian coordinates while ignoring the z Cartesian coordinate. The topological space in question is therefore embedded in \mathbb{R}^2 . This space appears to have a single path-connected component containing a single one dimensional hole. That is, the 0^{th} and 1^{st} Betti numbers of this topological space are both equal to 1. Topological features, such as path-connected components or holes, which persist over longer periods of time are usually considered of greater significance than those which persist over shorter periods [9]. Therefore when modelling the topological features of a swarm it is necessary to model the persistence of such features with respect to time.

This paper proposes a model of swarm behaviour that encodes the spatial-temporal characteristics of those topological features corresponding to Betti numbers. Specifically, the persistence of these topological features with respect to time are computed using *zig-zag persistent homology*. This gives a set of intervals representing the periods of existence of the topological features in question. These sets of intervals are subsequently converted into a normed vector space representation known as a *persistence landscape*. This space facilitates the application of statistical and data mining techniques. Validation of the model is performed using data corresponding to the swarm of fish introduced above. It is demonstrated that the proposed model can be used to perform retrieval and clustering of swarm behaviour in terms of topological features. The research presented here is an extension of previous work of the authors [10]; a greater discussion of the problem, literature review and evaluation of the proposed model is provided in the current article.

The remainder of this paper is structured as follows. Section II reviews related works on modelling swarm behaviour and topological features. Section III describes the proposed model of swarm behaviour in detail. Section IV presents a validation of the model. Finally section V draws conclusions from this research and discusses possible future research directions.

II. RELATED WORKS

This related works section is divided into two parts. Section II-A reviews related works on modelling swarm behaviour while II-B reviews related works on modelling the topological features of a topological space.

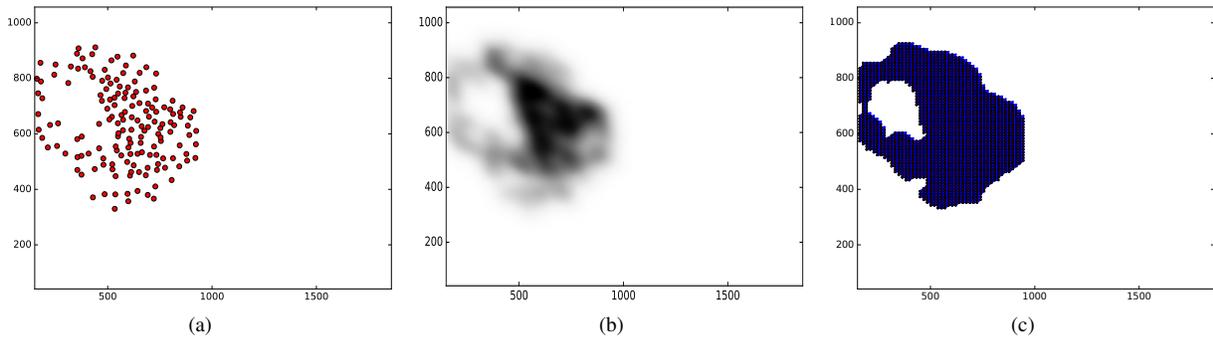


Fig. 1. A swarm of 300 fish is illustrated in (a) where each individual fish is represented by a red dot. The *Kernel Density Estimation* (KDE) f_h of the swarm is illustrated in (b). The simplicial complex of the upper-level set $f_h^{-1}[a, \infty)$ is illustrated in (c). The grid of points corresponding to this simplicial complex is of size 100×100 . The bandwidth h and threshold a of the upper-level set are equal to 0.23 and 1.25 respectively.

A. Modelling Swarm Behaviour

There exists an extensive literature of works which attempt to model the spatial-temporal characteristics of agents [11], [12]. In this section only that subset of works where the agents form a swarm are considered; that is, the agents move in close spatial proximity to each other. Of these works only the most relevant are reviewed and an interested reader is directed to the review article by Vicsek et al. [13] for a more extensive overview. Ballerini et al. [14] analysed the behaviour of a swarm of birds and discovered that agent interaction could not be modelled in terms of distance in the Cartesian coordinate space. Instead each agent was found to interact with on average a fixed number of neighbours. Two commonly used models of swarm behaviour are rotation order and polarization which measure the angular momentum and alignment of the agents respectively [7]. These models only consider swarm behaviour at a specific instance in time. In order to model the temporal features of swarm behaviour, Tunström et al. [7] proposed to model how rotation order and polarization vary as a function of time. Couzin et al. [2] used rotation order and polarization to validate a proposed model of swarm behaviour. Berger et al. [15] proposed a method for classifying swarm behaviour as *torus*, *flock* or *disordered* which are the most common behaviours exhibited by swarms. A behaviour of type torus occurs when the agents form a torus in three dimensions or annulus in two dimensions and move in a common circular motion. A behaviour of type flock occurs when the agents form a compact cluster and move in a common direction. Finally, a behaviour of type disordered occurs when the agents do not form a particular shape and their motion appears to be random.

Topaz et al. [16] proposed to model swarm behaviour by computing the Betti numbers corresponding to the swarm in question independently at each time step. These Betti numbers are subsequently plotted as a function of time. The authors found this approach to reveal characteristics of swarm behaviour not captured by other models that do not consider topological features. However this model does not consider the persistence with respect to time of those topological features corresponding to the Betti numbers. For example, consider the 0th Betti number which corresponds to the number of path-connected components. The model by Topaz et al. [16] computes the number of connected components at each

time step but it does not compute when these components first appeared and subsequently disappeared. It also does not compute if the connected components at different times are in fact the same or different connected components. In this paper a novel model of swarm behaviour that overcomes this limitation is proposed.

B. Modelling Topological Features

A number of models of topological relations between regions have been proposed within the GIS community [17]. The two most highly cited are the *Intersection Model* (IM) by Egenhofer [18] and *Region Connection Calculus* (RCC) by Randell et al. [19]. The IM is based on point set topology. The RCC is also based on point set topology but in addition provides a logic for reasoning. A number of works extended the above models to consider the temporal aspects of topological relations [20]. Egenhofer et al. [21] describe a model for reasoning about how topological relations between regions change over time. This model defines a partial order over relations which is in turn used to define a measure of similarity between different relations. The authors argue that this may be used to predict changes in such relations. Liu et al. [22], [23] proposed a closely related model which considers complex regions containing multiple path-connected components. Jiang et al. [24] present a tree model to represent topological relations and the temporal changes in such relations. Worboys et al. [25] proposed a method for detecting changes in topological features exhibited in triangulations corresponding to geosensor networks. The authors presented further developments in [26], [27].

The above models could potentially be used to model the topology of swarm behaviour but there are some significant differences relative to that proposed in this article. These models focus on modelling topological relations between regions and detecting changes in topological features. They are not concerned with modelling and analysing the persistence of topological features such as holes. Also these models assume that change over time is continuous [23] and when regions are represented by triangulations that change at each time step involves the addition or deletion of a single triangle [24]. Given this assumption, all of the above models consider a logic based approach where at each time step a set of conditions is

checked and used to infer changes in topological relations. In principle a very large number of changes could occur and need to be checked for. It is possible therefore that some will fail to be detected. The authors of [24] acknowledge that their experimental implementation did not detect all changes in topological features, though of course further rules could have been added. The model of swarm behaviour proposed in this article is formulated in terms of computing zig-zag persistent homology. This approach is algebraic, as opposed to logic, based. The zig-zag persistent homology may be computed exactly; therefore the method is correct and does not fail to detect changes in topological features. The approach presented here is similar to that of Worboys et al. [25] in that we also generate a triangulation (a simplicial complex), but here the triangulations are used to compute the zig-zag persistent homology, rather than computing changes with respect to the insertion and deletion of individual triangles. Notably, our use of persistence landscape leads to a normed vector space representation of topological features that facilitates spatial-temporal analysis.

III. MODEL OF SWARM BEHAVIOUR

The proposed model of swarm behaviour consists of the following three computational steps. Firstly the topological space in which the agents lie is inferred at each time step using *Kernel Density Estimation* (KDE). Next the persistence of topological features with respect to time is computed using *zig-zag persistent homology* [28]. This gives a set of intervals representing the periods of existence of the topological features in question. Finally these sets of intervals are converted into a normed vector space representation known as a *persistence landscape*. This space facilitates the application of statistical and data mining techniques [29], [30].

The following subsections describe each of these steps. Specifically, section III-A describes how the topological space is inferred. Section III-B briefly reviews homology theory which is used in the computation of the zig-zag persistent homology which is in turn described in section III-C. Section III-D describes how the output of this computation is converted into a persistence landscape representation.

A. Inferring the Topological Space

In this section we describe how the topological space in which the agents lie is inferred. It is necessary to represent this space using the combinatorial representation of a *simplicial complex* because this is what the zig-zag persistent homology computation requires as input. A simplicial complex \mathcal{K} is a collection of finite size subsets of a universal set where for each element σ of \mathcal{K} all subsets of σ are elements of \mathcal{K} . An element σ of \mathcal{K} is called a p -simplex if $|\sigma| = p + 1$ where $|\cdot|$ is the set size function. A simplex τ is a face of another simplex σ if $\tau \subset \sigma$. The intersection of any two elements of \mathcal{K} is either the empty set or a face of both elements [31].

As stated previously, it is assumed the agents are samples from a topological space. When attempting to infer this space it is necessary to do so in a manner that is robust to noise. In the context of modelling swarm behaviour noise equals a

minority of agents whose behaviour differs from that of the majority. A consequence of such noise is the introduction of topological artefacts. For example, consider Fig. 2(a) which displays the same swarm as that displayed in Fig. 1(a) but at a slightly later time. The topological features of the space at this time are similar to that of Fig. 1(a). However a single agent lies in the centre of the large path-connected component and in turn forms an additional path-connected component. Since this component is the consequence of the behaviour of a single agent, it is reasonable to classify it as topological noise.

There are a number of methods for clustering points robustly, notably DBSCAN [32]. There are also several methods for modelling the boundary of a set of points (e.g. [33]). Some of those methods might be applicable here, but with a view both to achieving robustness and to ensuring that holes are represented we use an approach which draws from recent works in the area of robust topological inference [34], [35], [36].

In the proposed model the following approach is employed. A *Kernel Density Estimation* (KDE) of agent locations is computed using a Gaussian kernel with bandwidth equal to h [37]. Let f_h denote this estimation. The upper-level set $f_h^{-1}[a, \infty)$ of this estimation can be considered a robust estimate of the topological space provided the threshold a is appropriately set. That is, the locations of those agents which can be considered to be noise will have a low density estimation. These locations will therefore not be represented in the inferred topological space. This upper-level set is subsequently represented using a simplicial complex which is denoted \mathcal{K} .

In the case of the topological space being embedded in the ambient space \mathbb{R}^2 , as is the case for the swarm of fish described above, one only needs to construct a simplicial complex containing simplices of dimension less than or equal to two. In this case the simplicial complex \mathcal{K} is constructed using the following approach. Firstly, the density for a grid of points over \mathbb{R}^2 is estimated. For each of these points a corresponding 0-simplex is included in \mathcal{K} if the density at that point is greater than the threshold a . For each pair of 0-simplices which are vertically, horizontally or main diagonally adjacent with respect to the grid, a corresponding 1-simplex is included in \mathcal{K} . For each triple of 0-simplices where all subsets of pairs are vertically, horizontally or main diagonally adjacent, a corresponding 2-simplex is included in \mathcal{K} . This construction produces a valid simplicial complex. Figure 3 illustrates this construction for a small grid where all points have a density greater than the threshold a .

Consider again the swarm illustrated in Fig. 1(a) and Fig. 2(a). The KDEs corresponding to this swarm are illustrated in Fig. 1(b) and Fig. 1(b) respectively. The simplicial complexes of the upper-level sets of these KDEs are illustrated in Fig. 1(c) and Fig. 2(c) respectively. Each simplicial complex contains a single path-connected component and a single one dimensional hole. They therefore accurately model the topological features of the swarm in a robust manner.

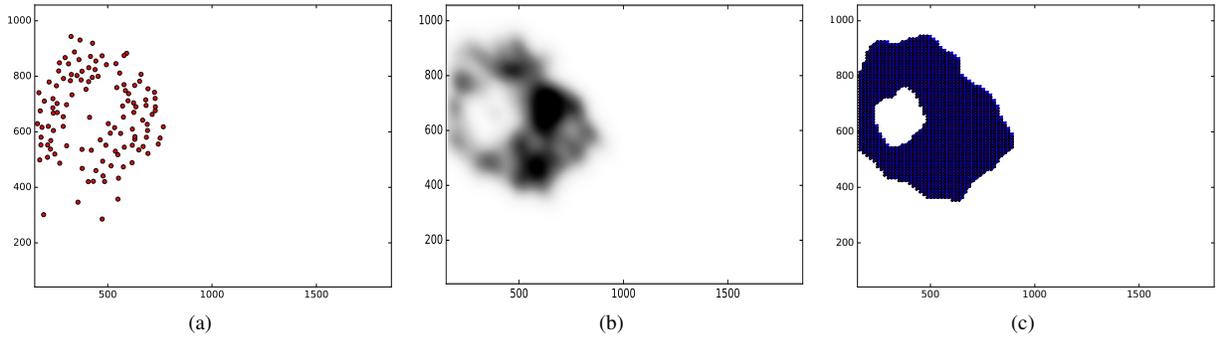


Fig. 2. A swarm of 300 fish is illustrated in (a) where each individual fish is represented by a red dot. The *Kernel Density Estimation* (KDE) f_h of the swarm is illustrated in (b). The simplicial complex of the upper-level set $f_h^{-1}[a, \infty)$ is illustrated in (c). The grid of points corresponding to this simplicial complex is of size 100×100 . The bandwidth h and threshold a of the upper-level set are equal to 0.23 and 1.25 respectively.

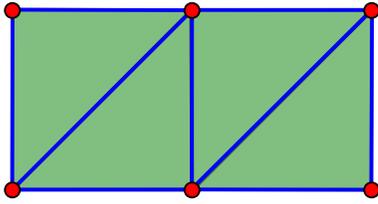


Fig. 3. For a grid of dimensions 2×4 a corresponding simplicial complex is illustrated where red dots represent 0-simplices, blue lines represent 1-simplices and green triangles represent 2-simplices.

B. Homology Theory

As discussed in the introduction to this article, a commonly used method to model the topology features of a topological space is to compute its corresponding Betti numbers. In this section we formally define the term Betti numbers. We also describe how the Betti numbers of a topological space are inferred from its corresponding simplicial complex representation.

Let \mathcal{K} be a simplicial complex. A p -chain on \mathcal{K} is defined in Equation 1 where each $\sigma_i \in \mathcal{K}$ is a p -simplex and each λ_i is an element in a specified field. The set of p -chains forms a group called a chain group $C_p(\mathcal{K})$. The *boundary map* ∂_p is a map from a p -simplex to the sum of its $(p-1)$ -simplex faces as defined in Equation 2. Here $[v_1, \dots, \hat{v}_i, \dots, v_{p+1}]$ is the $(p-1)$ -simplex obtained by deleting the 0-simplex v_i from the p -simplex $[v_1, \dots, v_{p+1}]$. This map is distributive and extends to the chain groups giving the sequence of chain groups in Equation 3. This sequence of groups is called a chain complex and is denoted C_* .

$$c = \sum \lambda_i \sigma_i \quad (1)$$

$$\partial_p \sigma = \sum_{i=1}^{p+1} [v_1, \dots, \hat{v}_i, \dots, v_{p+1}] \quad (2)$$

$$\dots \rightarrow C_{p+1}(\mathcal{K}) \xrightarrow{\partial_{p+1}} C_p(\mathcal{K}) \xrightarrow{\partial_p} C_{p-1}(\mathcal{K}) \rightarrow \dots \quad (3)$$

A p -chain $c \in C_p(\mathcal{K})$ is a p -boundary if there exists a $d \in C_{p+1}(\mathcal{K})$ where $c = \partial d$. Alternatively, it is a p -cycle if $\partial c = 0$. The sets of all p -boundaries and p -cycles

form corresponding groups which are denoted $B_p(\mathcal{K})$ and $Z_p(\mathcal{K})$ respectively. Each of these groups is a subgroup of $C_p(\mathcal{K})$. As a consequence of the fact $\partial_{p+1} \partial_p = 0$ it can be proved that $B_p(\mathcal{K}) \subseteq Z_p(\mathcal{K})$. The quotient group $H_p(\mathcal{K}) = Z_p(\mathcal{K})/B_p(\mathcal{K})$ is called the p -homology group of \mathcal{K} and its rank is called the p^{th} Betti-number. As described in the introduction to this article, the p^{th} Betti-number intuitively equals the number of p -dimensional holes in the simplicial complex \mathcal{K} . For a given simplicial complex there exists a number of methods for computing the corresponding Betti numbers [38], [8].

C. Zig-Zag Persistent Homology

In the interests of modelling swarm behaviour we wish to compute the persistence with respect to time of those topological features corresponding to the Betti numbers. This is accomplished by computing the zig-zag persistent homology of the sequence of simplicial complexes corresponding to the swarm in question. The approach is now described and its use is subsequently motivated.

Consider the sequence of simplicial complexes \mathcal{K} in Equation 4 which is a *zig-zag diagram* [28]. Each map \leftrightarrow in this equation equals either a forward inclusion map \rightarrow or a backward inclusion map \leftarrow . Forward and backward inclusion maps represent the addition and removal respectively of simplices. A zig-zag diagram induces the sequence of homology groups which is defined in Equation 5 and is known as a *zig-zag module*.

$$\mathcal{K} : \mathcal{K}_1 \leftrightarrow \mathcal{K}_2 \leftrightarrow \dots \leftrightarrow \mathcal{K}_n \quad (4)$$

$$H_p(\mathcal{K}) : H_p(\mathcal{K}_1) \leftrightarrow H_p(\mathcal{K}_2) \leftrightarrow \dots \leftrightarrow H_p(\mathcal{K}_n) \quad (5)$$

A zig-zag module can be decomposed into a direct sum of interval modules $\mathcal{I}_{[b,d]}$ [39]. The *zig-zag persistent homology* of the zig-zag diagram \mathcal{K} for dimension p is denoted $\text{Pers}_p(\mathcal{K})$ and is defined by Equation 6. It is the multiset of intervals $[b, d]$ corresponding to the set of interval summands $\mathcal{I}_{[b,d]}$ of $H_p(\mathcal{K})$. Each interval $[b, d]$ in $\text{Pers}_p(\mathcal{K})$ represents the persistence of a topological feature which appears at time b and persists until time d when it disappears. The topological features in question

are path-connected components in the case of $\text{Pers}_0(\mathcal{K})$ and one dimensional holes in the case of $\text{Pers}_1(\mathcal{K})$. The total persistence $\text{Pers}(\mathcal{K})$ of a zig-zag diagram \mathcal{K} is equal to the collection of $\text{Pers}_p(\mathcal{K})$ for each dimension p [39].

$$\text{Pers}_p(\mathcal{K}) = \{[b_j, d_j] | j \in J\} \quad (6)$$

When modelling the topological features of swarm behaviour, in most cases, one knows the agent locations at a sequence of discrete time steps. The corresponding sequence of simplicial complexes may not have the property that between each consecutive pair of simplicial complexes a forward or backward inclusion map exists. That is, the map between a consecutive pair of simplicial complexes may involve both the removal and addition of simplices. One cannot therefore compute the zig-zag persistent homology of such a sequence. To overcome this challenge, between each pair of consecutive simplicial complexes in the sequence an intermediate simplicial complex corresponding to the union of the simplicial complexes in question is introduced. This resulting sequence of simplicial complexes gives the zig-zag diagram \mathcal{K} of Equation 7 for which the zig-zag persistent homology may be computed. This computation is performed using the method of Carlsson et al. [39] which is implemented in the Dionysus software library.

$$\mathcal{K} : \mathcal{K}_1 \rightarrow (\mathcal{K}_1 \cup \mathcal{K}_2) \leftarrow \mathcal{K}_2 \rightarrow (\mathcal{K}_2 \cup \mathcal{K}_3) \leftarrow \mathcal{K}_3 \dots \mathcal{K}_n \quad (7)$$

To help illustrate the theory presented above, the following are some of the consequences of this formulation. If a hole (path-connected component) in \mathcal{K}_i intersects a single hole (path-connected component) in \mathcal{K}_{i+1} then the hole (path-connected component) in question persists from i until $i+1$. However if multiple holes (path-connected components) in \mathcal{K}_i intersect a single hole (path-connected component) \mathcal{K}_{i+1} all of the holes (path-connected components) in question die at $i+1$ apart from the hole (path-connected component) which appeared first.

D. Persistence Landscape

Recall that a total persistence $\text{Pers}(\mathcal{K})$ equals the collection of $\text{Pers}_p(\mathcal{K})$. We wish to convert this to a representation facilitating the application of statistical and data mining techniques. The most commonly used representation toward achieving this objective is the *persistence diagram*. This representation maps each interval in a given $\text{Pers}_p(\mathcal{K})$ to its endpoints [30]. A persistence diagram may be equipped with a metric, such as the Wasserstein or bottleneck metrics, to give a metric space [40]. However such a space does not facilitate many useful vector space operations, such as computing a mean [30]. One may compute the Fréchet mean using the above metrics but this may not be unique [41].

To overcome this limitation a representation which forms a normed vector space is used. This representation is known as a *persistence landscape* and is computed as follows [29]. First for a given interval $[a, b]$ a corresponding piecewise linear function $f_{[a,b]} : \mathbb{R} \rightarrow [0, \infty]$ is defined using Equation 8.

$$f_{[b,d]} = \begin{cases} 0, & \text{if } x \notin [b, d] \\ x - b, & \text{if } x \in [b, \frac{b+d}{2}] \\ -x + d, & \text{if } x \in (\frac{b+d}{2}, d] \end{cases} \quad (8)$$

For a given zig-zag persistent homology $\text{Pers}_p(\mathcal{K})$, which contains a multi-set of intervals, a corresponding persistence landscape is defined to be the sequence of functions $\lambda_k : \mathbb{R} \rightarrow [0, \infty]$ where $\lambda_k(x)$ is the k -th largest value in the multi-set of intervals at x . The persistence landscape is a function space which forms a normed vector space. In this work we use the L^2 norm which is defined in Equation 9 [42]. Using this norm one can compute the distance between two landscapes λ and λ' using Equation 10.

$$\|\lambda\|_2 = \left[\sum_{k=1}^{\infty} \int |\lambda_k|^2 dt \right]^{\frac{1}{2}} \quad (9)$$

$$\|\lambda - \lambda'\|_2 = \left[\sum_{k=1}^{\infty} \int |\lambda_k - \lambda'_k|^2 dt \right]^{\frac{1}{2}} \quad (10)$$

In order to form a normed vector space over $\text{Pers}(\mathcal{K})$ we adapt the following approach. As just discussed, each $\text{Pers}_p(\mathcal{K})$ in $\text{Pers}(\mathcal{K})$ forms an individual normed vector space. Let us denote this vector space \mathbb{L}_p . We take the *direct sum* of these spaces and equip the *direct sum norm* to form a new normed vector space [43]. Specifically let λ^p be a vector in the space \mathbb{L}_p . The norm in question is given by Equation 11 where m is the maximum dimension considered which in our case is 1.

$$\|(\lambda^0, \dots, \lambda^m)\|_2 = \left(\sum_{p=0}^m \|\lambda^p\|_2^2 \right)^{\frac{1}{2}} \quad (11)$$

IV. EXPERIMENTS

This section describes a set of experiments performed toward evaluating the proposed model of swarm behaviour. In section IV-A the data used within these experiments is described. In section IV-B the accuracy of the proposed model with respect to computing the total persistence $\text{Pers}(\mathcal{K})$ is evaluated. Finally sections IV-C and IV-D describe experiments which evaluate the ability of the proposed model to perform clustering and retrieval of swarm behaviour respectively.

A. Data

In all experiments presented, the data used corresponds to a swarm of 300 Golden Shiner fish and was described briefly in the introduction of this article [7]. The fish are swimming in a shallow pool (2.1m \times 1.2m, water depth 5cm). They were filmed for a duration of 56 minutes at a frame rate of 30 Hz and this framerate was down-sampled to a frame rate of 3 Hz. The pose (position and orientation) of each individual fish was tracked using a computer vision algorithm, details of which are described in [7]. Examples of the swarm in question at two different time steps are illustrated in Figure 1(a) and Figure 2(a). Here a time step equals a particular instant of

time. Given a frame rate of 3 Hz, the time duration between two consecutive time steps is 0.333 seconds.

Due to the time required to manually create ground truth data and computational complexity constraints, the entire dataset was not used in any of the experiments presented. Specifically the first 700 time steps were used to evaluate the accuracy of the model. On the other hand, time steps 1 to 3,010 were used to evaluate the ability of the proposed model to perform clustering and retrieval of swarm behaviour. Here, in the context of clustering and retrieval, a temporal window of 10 consecutive time steps was considered to give a total of 3,000 distinct but overlapping temporal windows. A temporal window of 10 consecutive time steps corresponds to a time duration of 3 seconds. From visually inspecting the data it was noted that most topological features exhibited by the swarm did not persist for a duration longer than this. Therefore this duration was determined to be sufficiently long to model the persistence of topological features.

As discussed in section III-A the simplicial complex \mathcal{K}_i at each time step i is computed using a grid of points. The size of this grid influences the resolution of topological features modelled. To aid the manual interpretation of the model output, a 10×10 grid was employed when evaluating the accuracy of the model in section IV-B. To allow topological features to be modelled at a finer resolution, a 100×100 grid was employed when evaluating the ability of the model to perform clustering and retrieval of swarm behaviour in sections IV-C and IV-D respectively.

Two additional model parameters which required specification were the bandwidth h of the Kernel Density Estimation f_h and the threshold a of the upper-level set $f_h^{-1}[a, \infty)$ (see section III-A). These parameters were specified such that the resulting upper-level set visually appeared to consistently accurately infer the topological space in question. Specifically the values used in all experiments were $h = 0.23$ and $a = 1.25$. Clearly there is a subjective element in this choice and alternative values may give different results.

B. Model Accuracy

Here we describe an experiment to determine how accurately our model computes the total persistence $\text{Pers}(\mathcal{K})$. Recall from section III-C that $\text{Pers}(\mathcal{K})$ is the collection of $\text{Pers}_p(\mathcal{K})$ where each $\text{Pers}_p(\mathcal{K})$ is in turn a multiset of intervals indicating the persistence of path-connected components for $p = 0$ and one dimensional holes for $p = 1$. Given that $\text{Pers}(\mathcal{K})$ is subsequently transformed into a persistence landscape representation, this represents an appropriate means to evaluate the accuracy of the proposed model.

We considered a temporal window of swarm behaviour of 700 consecutive time steps. By visually examining the corresponding sequence of simplicial complexes and determining the persistence of path-connected components and one dimensional holes, a ground truth $\text{Pers}(\mathcal{K})$ was constructed. To illustrate this process consider the sequence of simplicial complexes displayed in Figure 4 corresponding to a temporal window of 8 consecutive time steps. By visually inspecting this sequence it is evident that $\text{Pers}_0(\mathcal{K})$ contains the intervals [1, 8]

and [6, 7]. That is, a single path-connected component persists over the entire temporal window and never actually disappears. While a second path-connected component appears at time step six and disappears at time step seven; this path-connected component is represented by a single 0-simplex in Figure 4(f). Similarly, by visually inspecting this sequence it is evident that $\text{Pers}_1(\mathcal{K})$ contains the intervals [2, 3] and [5, 6]. The ground truth $\text{Pers}_0(\mathcal{K})$ and $\text{Pers}_1(\mathcal{K})$ corresponding to the temporal window containing 700 consecutive time steps contained 5 and 10 intervals respectively. The lengths of these intervals were (700, 1, 2, 4, 1) and (1, 2, 3, 1, 3, 2, 4, 2, 3, 2) respectively. The accuracy of $\text{Pers}(\mathcal{K})$ computed by the proposed model with respect to this ground truth was quantified in terms of precision and recall. It was found that 100% precision and 100% recall were achieved. This result validates the accuracy of the proposed model and is not surprising given that the method employed for computing zig-zag persistent homology is provably correct [39].

C. Clustering

This section presents a set of experiments which demonstrate that the proposed model can discover frequently occurring types of swarm behaviour in an automated manner. This is achieved by performing clustering of swarm behaviours. To perform this clustering the K-medoids data clustering method was employed [44]. Here the individual data points to be clustered correspond to 3,000 persistence landscape representations of swarm behaviour; one for each of the 3,000 temporal windows considered. K-medoids is an iterative method which iteratively determines K clusters by assigning each cluster a corresponding cluster centre, which is represented by an existing data point, such that the distance between each data point in that cluster and the cluster centre in question is minimized. As input the K-medoids algorithm takes a matrix of pairwise distances between data points. These distances are computed using the norm of Equation 11 and represent the pairwise similarity of swarm behaviour.

As a first experiment, clustering of swarm behaviour was performed using K-medoids with $K=2$. Figure 5 illustrates the cluster centres obtained. Recall swarm behaviour is modelled over a temporal window of length equal to 10 consecutive time steps. The left images of Figure 5(a) and 5(b) display the swarm at the midpoint of this window for each of the cluster centres. The centre and right images of Figure 5(a) and 5(b) illustrate the corresponding persistence diagrams of $\text{Pers}_0(\mathcal{K})$ and $\text{Pers}_1(\mathcal{K})$ respectively. Recall that a persistence diagram is constructed by mapping the intervals in question to their endpoints. For example, if a point exists at coordinates (1,9) in the persistence diagram of $\text{Pers}_0(\mathcal{K})$ this indicates that a path-connected component appeared at time 1 and disappeared at time 9. Similarly, if a point exists at coordinates (2,7) in the persistence diagram of $\text{Pers}_1(\mathcal{K})$ this indicates that a one dimensional hole appeared at time 2 and disappeared at time 7. Points which lie closer to the diagonal of a persistence diagram do not persist for a significant period and therefore are considered less significant topological features. In our figures the diagonal of a persistence diagram is represented

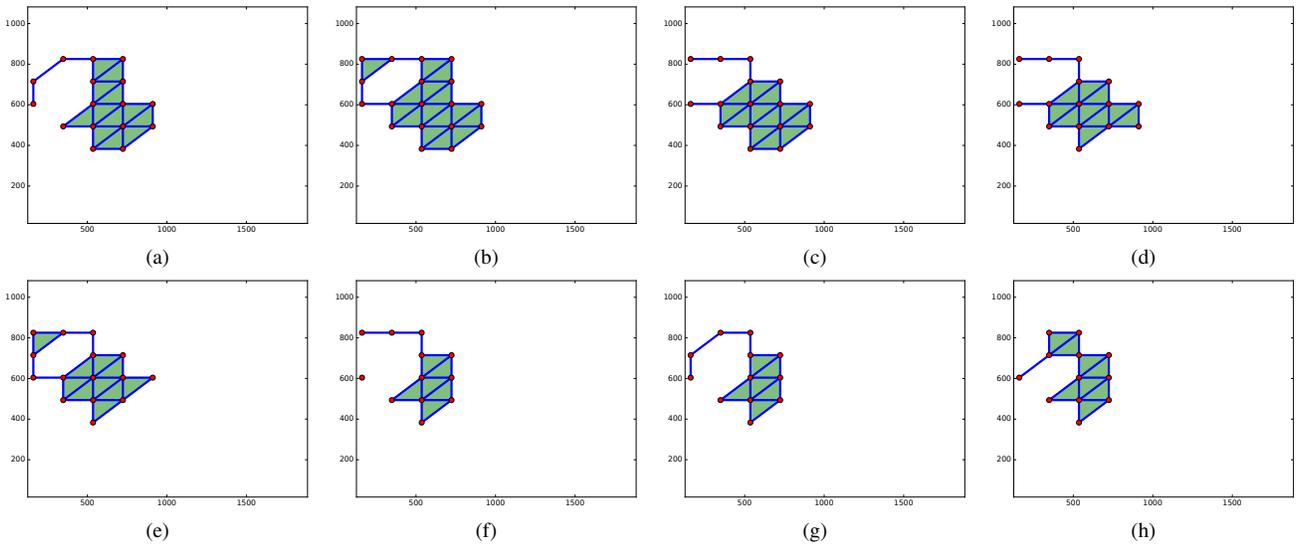


Fig. 4. A sequence of eight simplicial complexes are displayed in (a)-(h) where red dots represent 0-simplices, blue lines represent 1-simplices and green triangles represent 2-simplices. The grid of points corresponding to each simplicial complex is of size 10×10 .

by a blue line. Note that clustering was not directly applied to persistence diagrams but instead to corresponding persistence landscape representations.

Examining the persistence diagrams corresponding to a cluster centre reveals information regarding the set of swarm behaviours belonging to that cluster. Let us examine the persistence diagrams in Figure 5(a) corresponding to the first cluster centre. The persistence diagram of $\text{Pers}_0(\mathcal{K})$ contains a single point at coordinates (0,10) indicating that a single path-connected component persisted over the entire temporal window. The persistence diagram of $\text{Pers}_1(\mathcal{K})$ contains no points indicating that no one dimensional holes existed during the temporal window. Next let us examine the persistence diagrams in Figure 5(b) corresponding to the second cluster centre. The persistence diagram of $\text{Pers}_0(\mathcal{K})$ contains a point at coordinates (0,10) indicating that a path-connected component persisted over the entire temporal window. It also contains three points closer to the diagonal indicating that three path-connected components appeared only briefly and did not persist for a significant period. The exact times they appeared and disappeared can be determined from the diagram. The persistence diagram of $\text{Pers}_1(\mathcal{K})$ contains four points. These points are of varying distances away from the diagonal. The holes in question did not persist over the entire temporal window. From the above discussion it is evident that the clusters obtained using K-medoids for $K=2$ correspond to distinct swarm behaviours.

As a second experiment, clustering of swarm behaviour was performed using K-medoids with $K=3$. Relative to the previous clustering result, obtained using K-medoids with $K=2$, the same two cluster centres were obtained along with an additional cluster centre which is illustrated in Figure 6. For this cluster centre the persistence diagram of $\text{Pers}_0(\mathcal{K})$ contains a single point at coordinates (0,10) indicating that a single path-connected component persisted over the entire temporal window. The persistence diagram of $\text{Pers}_1(\mathcal{K})$ contains

a point at coordinates (0,10) indicating that a hole persisted over the entire temporal window. This persistence diagram also contains a point close to the diagonal. It is evident that the additional cluster obtained using K-medoids for $K=3$, as opposed to $K=2$, corresponds to an additional distinct swarm behaviour. For clusters obtained using K-medoids with K greater than 3, it was found that the additional clusters did not correspond to additional distinct swarm behaviours.

As discussed in the related works section of this article, *flock*, *torus* and *disordered* are considered the most frequently occurring types of behaviour exhibited by swarms. Examining the three clusters of behaviour obtained using our model we see that they in fact correspond specifically to these three types of behaviour. That is, a flock behaviour corresponds to a single path-connected component which persists for the entire duration. This behaviour is represented by the cluster in Figure 5(a). A disordered behaviour corresponds to a random number of path-connected components and a random number of holes where these features persist for a random duration. This behaviour is represented by the cluster in Figure 5(b). Finally, a torus behaviour corresponds to a single path-connected component with a single hole which persists for the entire duration. This behaviour is represented by the cluster in Figure 6. As such, our model was able to discover these three types of behaviour in an unsupervised manner.

D. Retrieval

This section describes a set of experiments which demonstrate that the proposed model may be used to perform retrieval of swarm behaviour with similar topological features to a given query swarm behaviour. Specifically, given one of the 3,000 swarm behaviours considered, the goal was to retrieve the most similar behaviours other than the query. Swarm behaviour was modelled using a persistence landscape representation and the similarity between two swarm behaviours was determined using the norm of Equation 11.

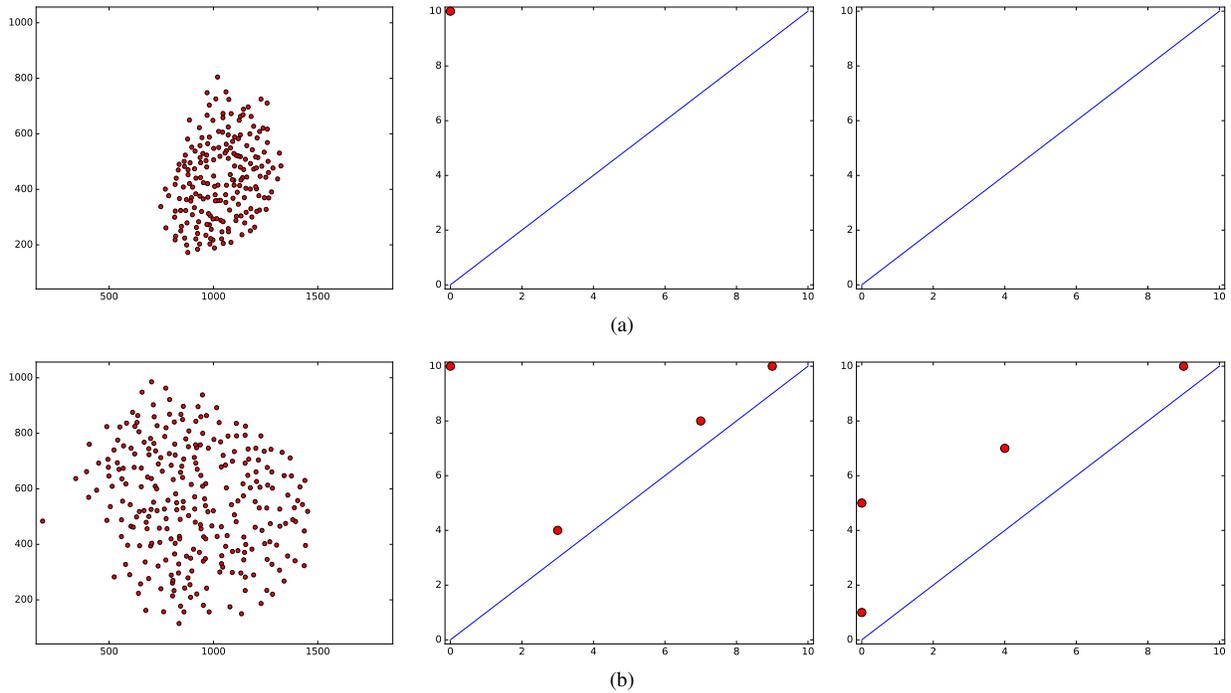


Fig. 5. The individual cluster centres obtained from clustering swarm behaviour with $K=2$ are illustrated in (a) and (b). The left image in each sub-figure displays the corresponding swarm at the midpoint of the temporal window over which behaviour is modelled. The centre and right images in each sub-figure illustrate the corresponding persistence diagrams of $\text{Pers}_0(\mathcal{K})$ and $\text{Pers}_1(\mathcal{K})$ respectively.

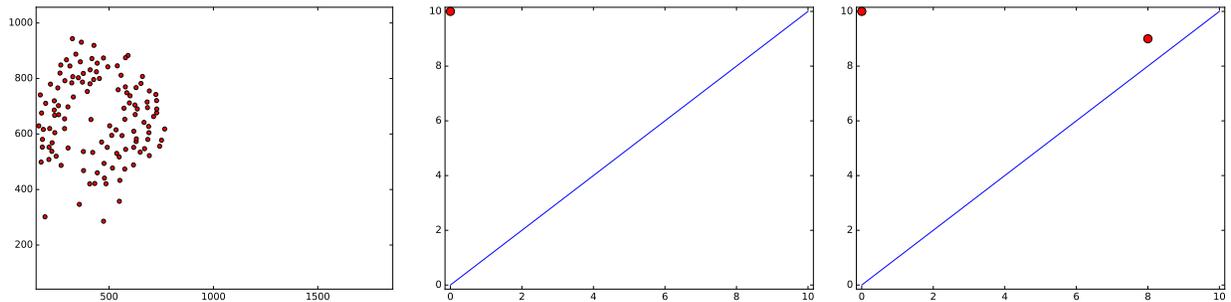


Fig. 6. The additional cluster centre obtained from clustering swarm behaviour with $K=3$ is illustrated. The left image displays the corresponding swarm at the midpoint of the temporal window over which behaviour is modelled. The centre and right images illustrate the corresponding persistence diagrams of $\text{Pers}_0(\mathcal{K})$ and $\text{Pers}_1(\mathcal{K})$ respectively.

For a query consisting of the swarm behaviour illustrated in Figure 5(a) the corresponding most and second most similar swarm behaviours, other than the query itself, are illustrated in Figure 7(a) and Figure 7(b) respectively. It is evident that the retrieved swarm behaviours are similar to that of the query. That is, in all three cases the swarm forms a single path-connected component with no holes which persists over the entire temporal window.

For a query consisting of the swarm behaviour illustrated in Figure 6 the corresponding most and second most similar swarm behaviours, other than the query itself, are illustrated in Figure 8(a) and Figure 8(b) respectively. Again it is evident that the retrieved swarm behaviours are similar to that of the query. That is, in all three cases the swarm forms a single path-connected component which persists over the entire temporal window. This component in turn contains one hole which persists over the entire temporal window and a another hole

which persists over a small portion of the temporal window.

V. CONCLUSIONS

This article presents a model of swarm behaviour which encodes spatial-temporal characteristics of topological features. To the authors' knowledge, this represents the first model of its kind. The experimental results presented demonstrate the proposed model may be used to perform the data mining tasks of clustering and retrieval of swarm behaviour in terms of topological features.

The authors believe that there exists much scope for future research and development. The following are some possible future research directions. The proposed model characterises topological features with respect to persistence over time. A possible research direction would be to characterise such features jointly with respect to both persistence over time and

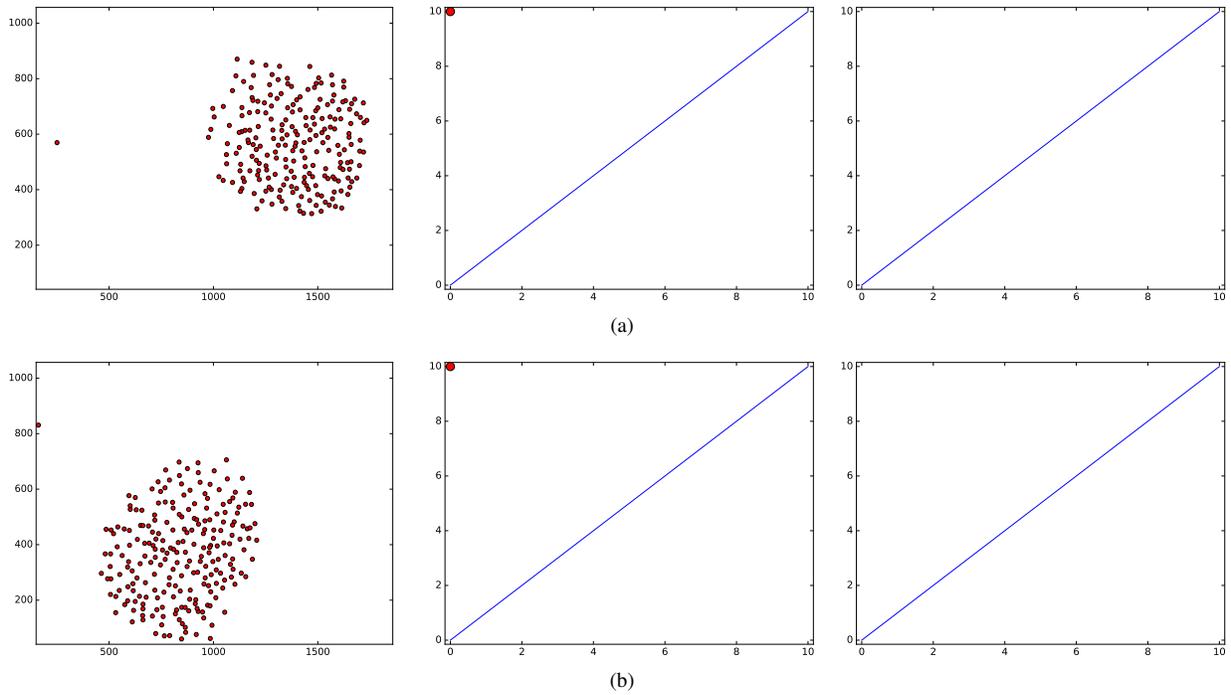


Fig. 7. For the swarm behaviour illustrated in Figure 5(a) the corresponding most and second most similar swarm behaviours are illustrated in (a) and (b) respectively.

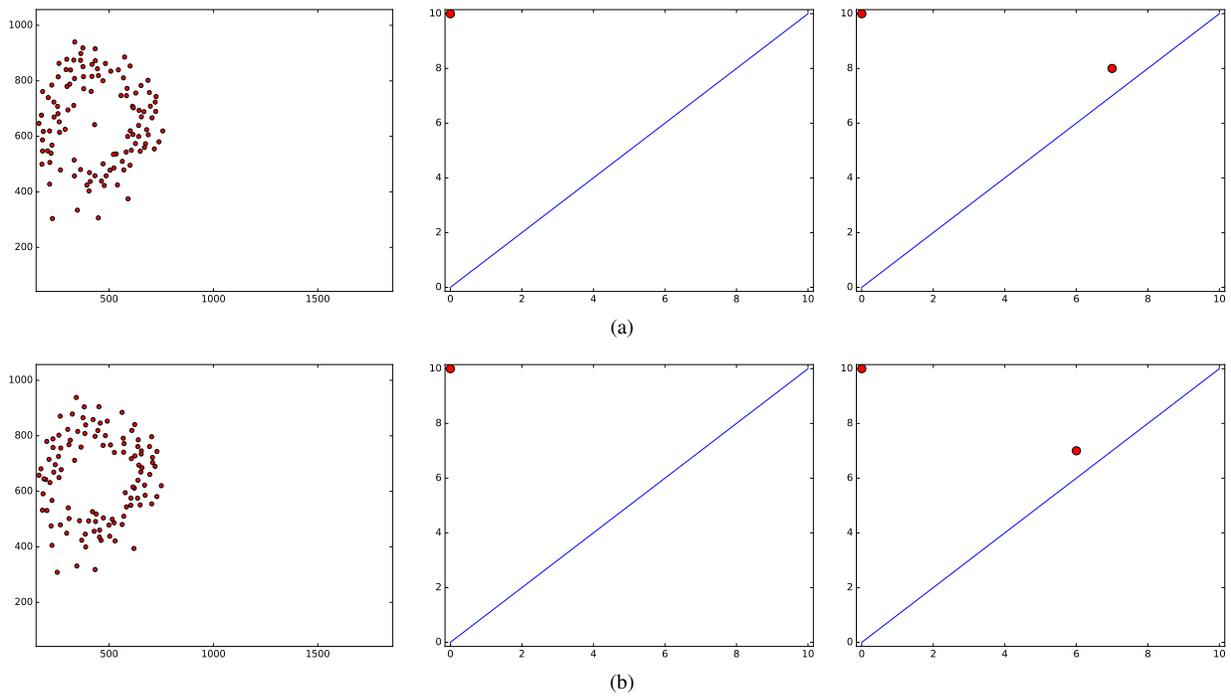


Fig. 8. For the swarm behaviour illustrated in Figure 6 the corresponding most and second most similar swarm behaviours are illustrated in (a) and (b) respectively.

scale using multi-dimensional persistent homology [45]. For the purpose of this article we didn't consider such an approach due to the fact that multi-dimensional persistent homology is still a developing research space and as a consequence the available tools are less mature.

In our experiments clustering of swarm behaviour was performed using the K-medoids method which requires the number of clusters to be specified. A possible research direction would be to employ an alternative method, such as DBSCAN, which does not have this requirement. A further issue is the selection of the KDE bandwidth, threshold and the grid resolution, all of which affect the distinction here between noise and signal. It would be possible to learn suitable values given manually annotated training data that explicitly identified significant and noise components. The proposed model only considers the topology of a swarm with respect to the locations of the agents in question. Considering the topology of a swarm with respect to additional dimensions, such as agent orientation, may provide a more accurate model. Finally, the proposed model uses a persistence landscape representation of persistence diagrams. Considering alternative representations, such as the persistence image by Adams et al. [40], is a possible research direction

The model proposed in this article has many potential applications other than modelling swarm behaviour. For example, the model could be employed to model the topological aspects of events in sensor networks.

REFERENCES

- [1] G. Wagner and H. Choset, "Gaussian reconstruction of swarm behavior from partial data," in *IEEE International Conference on Robotics and Automation*, 2015, pp. 5864–5870.
- [2] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, "Collective memory and spatial sorting in animal groups," *Journal of Theoretical Biology*, vol. 218, no. 1, pp. 1–11, 2002.
- [3] G. Pini, A. Brutschy, C. Pinciroli, M. Dorigo, and M. Birattari, "Autonomous task partitioning in robot foraging: an approach based on cost estimation," *Adaptive Behavior*, vol. 21, no. 2, pp. 118–136, 2013.
- [4] J. Werfel, K. Petersen, and R. Nagpal, "Designing collective behavior in a termite-inspired robot construction team," *Science*, vol. 343, no. 6172, pp. 754–758, 2014.
- [5] L. Bayindir, "A review of swarm robotics tasks," *Neurocomputing*, vol. 172, pp. 292–321, 2016.
- [6] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo, "Swarm robotics: a review from the swarm engineering perspective," *Swarm Intelligence*, vol. 7, no. 1, pp. 1–41, 2013.
- [7] K. Tunström, Y. Katz, C. C. Ioannou, C. Huepe, M. J. Lutz, and I. D. Couzin, "Collective states, multistability and transitional behavior in schooling fish," *PLoS Comput Biol*, vol. 9, no. 2, p. e1002915, 2013.
- [8] H. Edelsbrunner and J. Harer, *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [9] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.
- [10] P. Corcoran and C. B. Jones, "Spatio-temporal modeling of the topology of swarm behavior with persistence landscapes," in *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Burlingame, California, 2016.
- [11] P. Laube, *Computational movement analysis*. Springer, 2014.
- [12] P. Corcoran, P. Mooney, and G. Huang, "Unsupervised Trajectory Compression," in *IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, 2016.
- [13] T. Vicsek and A. Zafeiris, "Collective motion," *Physics Reports*, vol. 517, no. 3, pp. 71–140, 2012.
- [14] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini et al., "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1232–1237, 2008.
- [15] M. Berger, L. M. Seversky, and D. S. Brown, "Classifying Swarm Behavior via Compressive Subspace Learning," in *IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, 2016.
- [16] C. M. Topaz, L. Ziegelmeier, and T. Halverson, "Topological data analysis of biological aggregation models," *PLoS one*, vol. 10, no. 5, p. e0126383, 2015.
- [17] P. Corcoran, P. Mooney, and M. Bertolotto, "Spatial relations using high level concepts," *ISPRS International Journal of Geo-Information*, vol. 1, no. 3, pp. 333–350, 2012.
- [18] M. J. Egenhofer, "Reasoning about binary topological relations," in *Advances in Spatial Databases*. Springer, 1991, pp. 141–160.
- [19] D. Randell, Z. Cui, and A. Cohn, "A spatial logic based on regions and connection," in *International Conference on Knowledge Representation and Reasoning*. Morgan Kaufmann, 1992, pp. 165–176.
- [20] F. Wolter and M. Zakharyashev, "Spatio-temporal representation and reasoning based on RCC-8," in *Principles of Knowledge Representation and Reasoning Proceedings of the Seventh International Conference*, 2000.
- [21] M. J. Egenhofer and K. K. Al-Taha, "Reasoning about gradual changes of topological relationships," in *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*. Springer, 1992, pp. 196–219.
- [22] H. Liu and M. Schneider, "Detecting the topological development in a complex moving region," *Journal of Multimedia Processing and Technologies*, vol. 1, no. 3, pp. 160–180, 2010.
- [23] H. Liu and M. Schneider, "Tracking continuous topological changes of complex moving regions," in *ACM Symposium on Applied Computing*, 2011, pp. 833–838.
- [24] J. Jiang and M. Worboys, "Event-based topology for dynamic planar areal objects," *International Journal of Geographical Information Science*, vol. 23, no. 1, pp. 33–60, 2009.
- [25] M. Worboys and M. Duckham, "Monitoring qualitative spatiotemporal change for geosensor networks," *International Journal of Geographical Information Science*, vol. 20, no. 10, pp. 1087–1108, 2006.
- [26] L.-J. Guan and M. Duckham, "A unified framework for decentralized reasoning about gradual changes in topological relations," in *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2011, pp. 413–416.
- [27] L.-J. Guan and M. Duckham, "Decentralized reasoning about gradual changes of topological relationships between continuously evolving regions," in *Spatial Information Theory*. Springer, 2011, pp. 126–147.
- [28] G. Carlsson and V. De Silva, "Zigzag persistence," *Foundations of Computational Mathematics*, vol. 10, no. 4, pp. 367–405, 2010.
- [29] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 77–102, 2015.
- [30] P. Bubenik and P. Dłotko, "A persistence landscapes toolbox for topological statistics," *Journal of Symbolic Computation*, vol. 78, pp. 91–114, 2017.
- [31] R. Ghrist, *Elementary applied topology*. ed. 1.0, Createspace, 2014.
- [32] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1996, pp. 226–231.
- [33] A. Galton and M. Duckham, "What is the region occupied by a set of points?" in *International Conference on Geographic Information Science*. Springer, 2006, pp. 81–98.
- [34] J. M. Phillips, B. Wang, and Y. Zheng, "Geometric inference on kernel density estimates," *arXiv preprint arXiv:1307.7760*, 2013.
- [35] F. Chazal, B. T. Fasy, F. Lecci, B. Michel, A. Rinaldo, and L. Wasserman, "Robust topological inference: Distance to a measure and kernel distance," *arXiv preprint arXiv:1412.7197*, 2014.
- [36] P. Corcoran, "Topological path planning in GPS trajectory data," *Sensors*, vol. 16, no. 12, p. 2203, 2016.
- [37] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC Press, 1986, vol. 26.
- [38] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.
- [39] G. Carlsson, V. De Silva, and D. Morozov, "Zigzag persistent homology and real-valued functions," in *Symposium on Computational Geometry*, 2009, pp. 247–256.

- [40] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, "Persistence images: a stable vector representation of persistent homology," *Journal of Machine Learning Research*, vol. 18, no. 8, pp. 1–35, 2017.
- [41] Y. Mileyko, S. Mukherjee, and J. Harer, "Probability measures on the space of persistence diagrams," *Inverse Problems*, vol. 27, no. 12, p. 124007, 2011.
- [42] O. Christensen, *Functions, spaces, and expansions: mathematical tools in physics and engineering*. Springer Science & Business Media, 2010.
- [43] F. Albiac and N. J. Kalton, *Topics in Banach space theory*. Springer Science & Business Media, 2006, vol. 233.
- [44] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for k-medoids clustering," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3336–3341, 2009.
- [45] A. Cerri, M. Ethier, and P. Frosini, "The coherent matching distance in 2d persistent homology," in *International Workshop on Computational Topology in Image Context*, 2016, pp. 216–227.

Padraig Corcoran Padraig Corcoran is a lecturer in the School of Computer Science and Informatics at Cardiff University. Prior to this, he completed a European Marie Curie International Outgoing Fellowship. During the outgoing phase of this fellowship he worked at Massachusetts Institute of Technology (MIT) while during the incoming phase he worked at University College Dublin (UCD).

Christopher B. Jones Chris Jones took up his current position as Professor of Geographical Information Systems in the School of Computer Science & Informatics at Cardiff University in 2000. He has also held academic positions, in computer science, at the University of South Wales and, in geography, at the University of Cambridge and has worked in software development at the British Geological Survey and at BP Exploration. Research at the University of South Wales on cartographic label placement led to the development of the Maplex software which is now an ESRI product. He graduated in geology, from Bristol University, and received a PhD, from the School of Physics, University of Newcastle upon Tyne, for research on periodicities in fossil growth rhythms.