

---

# TDA of large swarming dynamics

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 Swarms of birds or fishes that use internal chemical energy to form patterns of  
2 motion, such as biological swarms and cell groups, have attracted much attention in  
3 science and engineering in recent years. This study aims to capture the hierarchical  
4 structure formed by such swarms in large-degree-of-freedom systems through  
5 topological data analysis (TDA). A large-scale swarm simulation was conducted  
6 using the Reynolds' Boids model, which reproduces a swarm of birds by balancing  
7 the attraction and repulsion between individuals and the alignment force. As a  
8 result, they reported that filament-like, relatively small swarms driven by velocity  
9 fluctuations and relatively large swarms driven by density fluctuations occur, even  
10 though all particles have the same parameters. On the other hand, these analyses  
11 only analyze the internal fluctuation structure of the swarms but do not elucidate  
12 the relationships between the swarms. The aim of this study is to reveal the  
13 geometrical structure behind swarm pattern series by analyzing them using TDA  
14 and Wasserstein distance.

## 15 1 Introduction

16 Swarms of birds or fishes that use internal chemical energy to form patterns of motion, such as  
17 biological swarms and cell groups, have attracted much attention in science and engineering in  
18 recent years. Recently, it has been suggested that topological data analysis (TDA) is useful for  
19 analyzing such swarm behaviors [1, 2, 3]. These studies analyzed swarm behaviors with relatively  
20 small variability. This study aims to analyze large-scale swarming behavior through TDA and reveal  
21 the hierarchical structure caused by the scale nature of the swarming system.

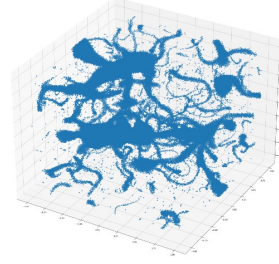
22 A large-scale swarm simulation was conducted using the Reynolds' Boids model, which reproduces  
23 a swarm of birds by balancing the attraction and repulsion between individuals and the alignment  
24 force. As a result, the authors reported the emergence of filament-like, relatively small swarms driven  
25 by velocity fluctuations and relatively large swarms driven by density fluctuations, even though all  
26 individuals have the same parameters [4, 5]. On the other hand, these analyses only analyze the  
27 internal fluctuation structure of the swarms but do not elucidate the relationships among the swarms.  
28 The aim of this study is to elucidate the hierarchical structure behind swarm pattern series by TDA  
29 [6] and Wasserstein distance [7].

## 30 2 Reynolds' boid model

31 In Reynolds' boid model [8], each individual moves accordingly to three forces, which are the forces  
32 attracting each other, separating each other, and aligning the orientation of each other:

$$\frac{d\mathbf{p}_j}{dt} = -W_{\text{ali}} \left( \mathbf{p}_j - \frac{\sum_k \mathbf{p}_k}{n_{\text{ali}}} \right) + W_{\text{sep}} \left( \sum_k \frac{(\mathbf{q}_j - \mathbf{q}_k)}{|\mathbf{q}_j - \mathbf{q}_k|} \right) + W_{\text{ali}} \left( \mathbf{q}_j - \frac{\sum_k \mathbf{q}_k}{n_{\text{att}}} \right),$$

where  $\mathbf{q} := (q_1, q_2, q_3)$ ,  $\mathbf{p} := (p_1, p_2, p_3)$ , and  $j, k$  represent the index of an individual. The attraction, separation, and alignment terms are represented by the first, second, and third terms, and the forces have the interaction ranges and the angles of view, respectively. The parameters  $W_{\text{att}}$ ,  $W_{\text{sep}}$ ,  $W_{\text{ali}}$ ,  $r_{\text{att}}$ ,  $r_{\text{sep}}$ ,  $r_{\text{ali}}$ ,  $\theta_{\text{att}}$ ,  $\theta_{\text{sep}}$ , and  $\theta_{\text{ali}}$  of the Reynolds' boid model can be tuned to simulate the collective motion of living things such as birds or fish [8, 9]. In this study, we focused on a parameter set that simulates the large-scale hierarchical structure described in [4, 5] (Snapshot of simulation is described in Fig. 1).



### 3 Analysis methods

TDA is a generic term for a group of data analysis methods that focus on the topological geometry of data. Topological features are different from those commonly used in pattern structure analysis, such as statistics and Fourier bases, in that they provide robust features that do not depend on small differences in the input data [6].

Figure 1: Swarm structure formed by large-scale simulation based on Reynolds' boids model

Persistent homology (PH) group is one of the topological features used in TDA. PH group represents a figure as a set of creation and annihilation of hole structures depending on a single parameter such as parameter related to scale of holes (Fig. 2). In general, topological information does not have distance information such as spatial scale, which is crucial for understanding physical phenomena, but in PH, scale information can be added via a single parameter in its features. PH is uniquely represented as a persistent diagram (PD) with the scale at which the hole occurred on the horizontal axis and the scale at which it disappeared on the vertical axis (Fig. 2):

$$PD_l = \{(b_k, d_k) \in \mathbf{R}^2 \mid k = 1, 2 \dots N_{\text{hole}}\}, \quad (1)$$

where  $l$  represents the dimension of holes and  $N_{\text{hole}}$  is the number of hole in the swarm structure. In this study, we focused on three-dimensional holes  $l = 3$ .

In this study, we first perform a TDA of a given snapshot according to the pattern dynamics analysis procedure proposed in [10]. In this analysis, the PD of the snapshot is calculated, and at the same time, the correspondence between the structure on the PD and the geometric structure of the swarm in real space is explored using the inverse analysis method [11, 12]. In the inverse analysis, the Death simplices method was used to extract the structure corresponding to the moment the hole died [11, 12]. Analysis of snapshots to help understand the correspondence between PD and swarm structure.

Next, we analyze the series of PDs that evolve in time according to the dynamics following [10], and combine this with the results of the snapshot analysis to understand the time evolution of the geometrical structure of the swarms. Before analyzing the PD series, we first check whether the PD series contain important information about swarm dynamics. For this purpose, we perform a regression analysis with the histogram of PD as the explanatory variable and the corresponding time

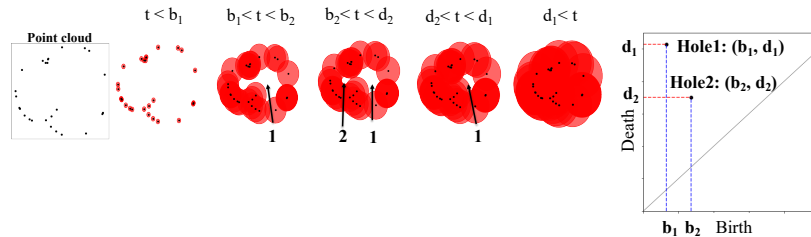


Figure 2: The left panel shows the evolution of the simplicial complex corresponding to the increase in radius, and the right panel shows the PD corresponding to the left panel's series of figures.

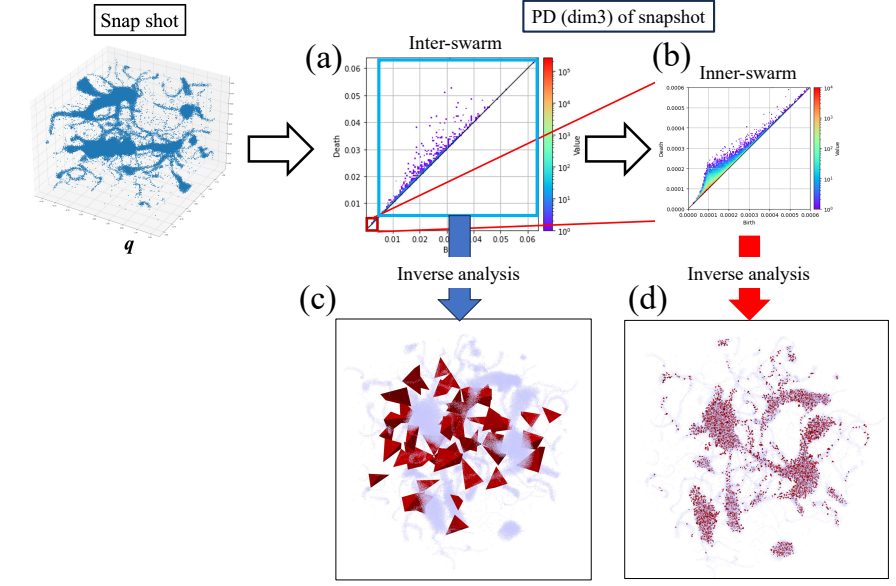


Figure 3: TDA of snapshot of the swarms' structure. (a) is a PD of the snapshot. (b) is a close-up of the PD's small radius region corresponding to the swarms' inner structure. (c) is the result of the inverse analysis of the PD in the large radius region (blue box). (d) is the result of the inverse analysis of PD in the small radius region.

as the objective variable. The histogram of PD is defined as,

$$h(r_{\text{birth}}, r_{\text{death}}) = \sum_{i=1}^{N_{\text{hole}}} \frac{1}{2\pi\sigma} \exp \left\{ -\frac{1}{4\sigma^2} [(r_{\text{birth}} - b_i)^2 + (r_{\text{death}} - d_i)^2] \right\}. \quad (2)$$

In this analysis, we estimate the predictive performance of the regression model using the cross-validation method [13]. If this performance is sufficiently high, we conclude that information on dynamics is embedded in the PD series.

After confirming that the PD series encompassed the information, the time variation of the distribution function,  $h(r_{\text{birth}}, r_{\text{death}})$ , was evaluated using the inter-distributional distance to assess the time variation of the geometric structure of the swarms. Concretely, using the mean distribution of the PD series,

$$h_{\text{mean}}(r_{\text{birth}}, r_{\text{death}}) := \frac{1}{N_{\text{time}}} \sum_{t=1}^{N_{\text{time}}} h_t(r_{\text{birth}}, r_{\text{death}}), \quad (3)$$

where  $h_t$  represents the histogram function computed from the PD at a given time  $t$ . In this study,  $N_{\text{time}}$  is set to 100. We use  $h_{\text{mean}}$  as a reference, the distances between each histogram  $h_t$  and the  $h_{\text{mean}}$  were evaluated using the 1-Wasser-Stein distance [7]. From the series of distances, information about the fluctuation structure of the swarm should be extracted.

## 4 Results and Discussion

By applying TDA to a snapshot data of the swarm, the PD as shown in Fig. 3(a) was obtained. The PD was observed to have a generator distribution (points on the PD) corresponding to long-lived holes in large radius regions (Fig. 3(a)). When the micro-radius region of the PD was enlarged, a peculiar structure with a relatively long lifetime was also observed in this region, as shown in Fig. 3(b). Applying the inverse analysis to the PD regions with these unique generators, blue and red rectangles in Fig. 3(a), it was confirmed that the long-lived generators in the large radius regions correspond to the inter-group hole structure (Fig. 3(c)), while the long-lived generators in the small radius regions correspond to the intra-group structure (Fig. 3(c)).

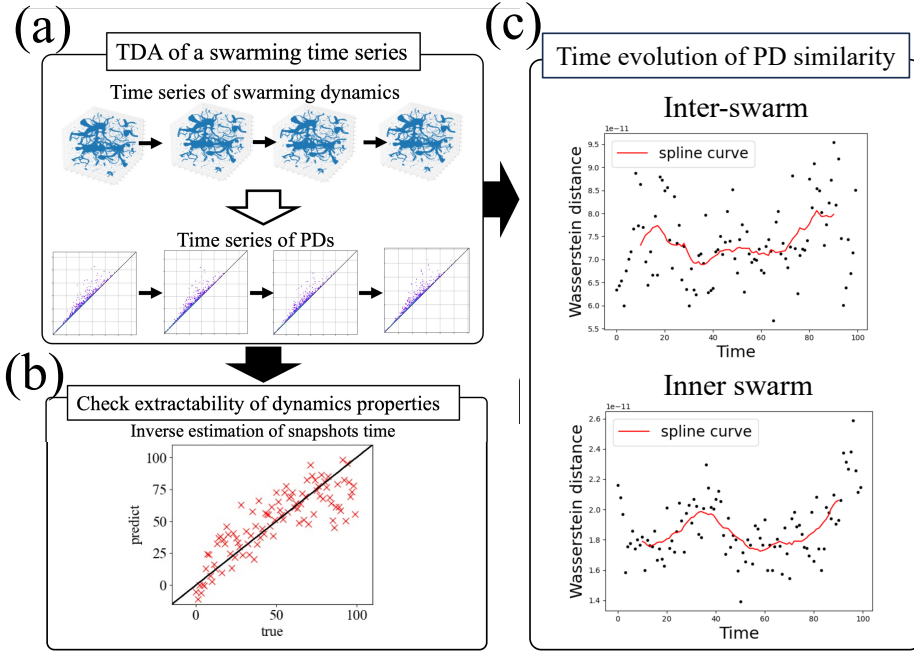


Figure 4: Analysis results of the time series of the swarming dynamics. (a) is an example of a time series of a swarm structure and the corresponding PD series. (b) is the regression result from the PD series with the corresponding time. The figure shows the predicted results for the test data using the cross-validation method. (c) is the time evolution of the Wasser Stein distance between the PD series and the average PD. The upper figure shows the results of the analysis of the PD region corresponding to a large radius, and the lower figure shows the results of the analysis of the PD region corresponding to a small radius. The red line is the curve obtained by smoothing at 20 time intervals.

A regression analysis was performed to estimate each time from a series of PD histograms  $\{h_t\}_{t=1}^{100}$ . As a result, a regression model with high accuracy was constructed for the prediction data (Fig. 4(b)). This suggests that the series of PD contains sufficient information about swarm dynamics. Note that, the resolution of the histogram used for this analysis was  $12800 \times 12800$ .

Finally, we analyzed the time evolution of the 1-Wasserstein distance between  $\{H_t\}_{t=1}^{100}$  and  $H_{\text{mean}}$ . The blue boxed area in Fig. 3(a) corresponding to the structure of the inter-swarm and the red boxed area in Fig. 3(a) corresponding to the structure of the intra-swarm, respectively, were analyzed. The results showed that the variation of the intra-swarm was less scattered with respect to the smoothing curve (red lines) compared to the inter-group variation (Fig. 4(c)). This result suggests that the internal structure of the swarm fluctuates while maintaining its average geometric structure. It seems consistent with the results obtained in previous studies that large swarms are driven by density fluctuations [4, 5]. Both inter- and intra-swarm smoothing curves varied systematically, which suggests that there are dynamics with similar time constants inter- and intra-swarm, and thus the possibility of interactions between inter- and intra-swarm structures is expected.

## 5 Summary

This study confirms the validity of analyzing swarm dynamics via a series of TDAs. In ordinary analysis of swarm dynamics, identification of individual swarm structure is a technical challenge, but the results suggest that TDA can be used to analyze swarm dynamics without going through such a preprocessing step [4, 5]. The development of the use of TDA for the analysis of swarm dynamics is expected in the future.

## References

- [1] Dhananjay Bhaskar, Angelika Manhart, Jesse Milzman, John T Nardini, Kathleen M Storey, Chad M Topaz, and Lori Ziegelmeier. Analyzing collective motion with machine learning and topology. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12), 2019.
- [2] Kyle C. Nguyen, Carter D. Jameson, Scott A. Baldwin, John T. Nardini, Ralph C. Smith, Jason M. Haugh, and Kevin B. Flores. Quantifying collective motion patterns in mesenchymal cell populations using topological data analysis and agent-based modeling. *Mathematical Biosciences*, 370:109158, 2024.
- [3] Chad M Topaz, Lori Ziegelmeier, and Tom Halverson. Topological data analysis of biological aggregation models. *PloS one*, 10(5):e0126383, 2015.
- [4] Yhoichi Mototake and Takashi Ikegami. A simulation study of large scale swarms. *Proc. SWARM*, pages 446–450, 2015.
- [5] Takashi Ikegami, Yohichi Mototake, Shintaro Kobori, Mizuki Oka, and Yasuhiro Hashimoto. Life as an emergent phenomenon: studies from a large-scale boid simulation and web data. *Phil. Trans. Royal. Soc. A-Math. Phys. Eng. Sci.*, 375(2109):20160351, 2017.
- [6] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
- [7] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.
- [8] Craig W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH ’87, pages 25–34, New York, NY, USA, 1987. Association for Computing Machinery.
- [9] Iain D Couzin, Jens Krause, Richard James, Graeme D Ruxton, and Nigel R Franks. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.*, 218(1):1–11, 2002.
- [10] Yoh ichi Mototake, Masaichiro Mizumaki, Kazue Kudo, and Kenji Fukumizu. Procedure to reveal the mechanism of pattern formation process by topological data analysis, 2024.
- [11] Ipei Obayashi, Yasuaki Hiraoka, and Masao Kimura. Persistence diagrams with linear machine learning models. *Journal of Applied and Computational Topology*, 1(3):421–449, 2018.
- [12] Robin Vandaele, Guillaume Adrien Nervo, and Olivier Gevaert. Topological image modification for object detection and topological image processing of skin lesions. *Scientific Reports*, 10(1):1–15, 2020.
- [13] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009.