# Topological data analysis of large swarming dynamics

**Yoh-ichi Mototake***
Faculty of Social Data Science
Hitotsubashi University
y.mototake@r.hit-u.ac.jp

**Shinichi Ishida**
Faculty of Social Data Science
Hitotsubashi University
24034@edu.sds.hit-u.ac.jp

**Norihiro Maruyama**
Graduate School of Arts and Sciences
The University of Tokyo
maruyama@sacral.c.u-tokyo.ac.jp

**Takashi Ikegami**
Graduate School of Arts and Sciences
The University of Tokyo
ikeg@sacral.c.u-tokyo.ac.jp

## Abstract

In recent years, increasing attention has been directed toward the study of swarm dynamics in animals and cells in the fields of science and engineering. Notably, significant progress has been made in observing and analyzing large-scale swarming behavior, characterized by the collective movement of numerous individuals and spanning up to several kilometers in size. Traditional analyses of such large-scale swarming behavior have often focused on specific hierarchical levels, examining the dynamics within these levels. However, understanding the interactions between different hierarchical levels is equally critical, as swarming behavior emerges from both intra-group and inter-group dynamics. In this study, we analyzed large-scale swarm dynamics using topological data analysis based on persistent homology. This approach enables the extraction of topological features across multiple scales and provides a quantitative framework for understanding the relationships between these features.

## 1 Introduction

Swarms of animals or cells that use internal chemical energy to form swarm patterns have attracted much attention in science and engineering in recent years. Cavagna et al. [1] identified the presence of a scale-free property in swarming behavior. Specifically, they observed that the ratio of the correlation length of the velocity fluctuations among individuals in a flock to the flock size is constant regardless of the flock size. This finding, derived from measurements and analyse of bird flocks, , suggests that certain aspects of swarm dynamics exhibit characteristics of a critical system.

A series of models, known as boid models, have been proposed to reproduce swarming behaviors using simple rules between individuals [2, 3, 4, 5, 6]. For example, Vcsek modeled the collective motion of a flock of birds by drawing an analogy between the velocity vector of an individual bird and the spin in a ferromagnet [5]. In this model, the existence of giant density fluctuations, in which local particle density fluctuations in the collective motion layer exceed the values predicted by the central limit theorem, has been confirmed. This discovery has led to numerous studies on the model, as it exhibits various unique characteristic of non-equilibrium systems. However, the Vicsek model does not produce the well-defined clusters observed in real swarms. In fact, chate et al. reported that, in addition to the directional alignment effect of the Vicsek model, the addition of the effects of attraction, which causes individuals to move closer together, and repulsion, which prevents collisions

*Email: y.mototake@r.hit-u.ac.jp

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

by causing individuals to move away when too close, results in movements more similar to natural bird flocking in three-dimensional simulations [7].

Although many studies have explored swarming behavior, most have focused on the behavior of a single swarm with population sizes ranging from 100 to 1,000 individuals. Few studies have examined large populations where multiple swarms interact with each other simultaneously. In nature, large flocks interact with one another [8, 9]. For example, in western herring flocks, it has been observed that depending on their trophic state, the flocks separate into two groups, one in the upper layer of the ocean and the other in the lower layer, and that the members of these two groups switch between them as each individual's trophic state changes [9]. Also, a large-scale swarm simulation was conducted [10] using Reynolds' boids model [3], which reproduces a swarm of birds by balancing the attraction and repulsion between individuals and the alignment force. The study reported the emergence of filament-like, relatively small swarms driven by velocity fluctuations, as well as larger swarms driven by density fluctuations, despite all individuals sharing the same parameters [10, 11]. However, quantitative analyses of the hierarchical structures, such as interactions between swarms in large-scale systems, remain scarce in studies of such swarm models.

This study aims to analyze large-scale swarming behavior through topological data analysis (TDA) [12] and Wasserstein distance [13] to reveal the hierarchical nature of large-scale swarming systems. Recently, topological data analysis (TDA) has been suggested as a useful tool for analyzing swarming behaviors [14, 15, 16, 17], with prior studies focusing primarily on relatively small-scale swarms. TDA is a data analysis method that achieves multi-level scale hole feature extraction by adding hole scale information to the topological structure that extracts the number of holes. Therefore, the application of TDA to large-scale swarming behavior is expected to enable the quantification of swarming phenomena across multiple scales. TDA represents the extracted features as point cloud data, called a persistence diagram. In this study, we propose an analytical framework to extract the fluctuating structures of flock geometry by quantifying them as the Wasserstein distance between the distributions of the persistence diagram data. This framework is applied to large-scale flock simulation data to better understand the underlying hierarchical dynamics of swarming behavior.

## 2 Reynolds' boid model

For the simulation conducted in this study, because generating flocks with well-defined shapes is needed for multiple flocks to interact, a model incorporating three forces, as described in the Introduction section, is deemed appropriate. In this study, the boid model proposed by Reynolds et al. [3] was selected from among such models. Reynolds' boid model is one of the earliest attempts to replicate flock behavior based on interactions and has been shown to effectively reproduce complex bird flocking behavior [3]. In Reynolds' boid model, each individual moves accordingly to three forces: attraction, which draws individuals closer; separation, which prevents individuals from colliding; and alignment, which aligns an individual's orientation with those nearby:

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

where $\boldsymbol{q} := (q_1, q_2, q_3)$, $\boldsymbol{p} := (p_1, p_2, p_3)$, 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 $W_{\mathrm{ali}}$, $W_{\mathrm{sep}}$, $W_{\mathrm{att}}$, $r_{\mathrm{ali}}$, $r_{\mathrm{sep}}$, $r_{\mathrm{att}}$, $\theta_{\mathrm{ali}}$, $\theta_{\mathrm{sep}}$, and $\theta_{\mathrm{att}}$ of Reynolds' boid model can be tuned to simulate the collective motion of living things such as birds or fish [3, 18]. In this study, we focused on a specific parameter set (see Table 2) that simulates the large-scale hierarchical structure described in [10, 11]. A snapshot of the simulation is shown in Fig. 1. In general, a narrower effective interaction range reduces the computational cost of the simulation because there are fewer pairs of interactions to consider. However, in such a parameter region, the interaction in the whole system is weak

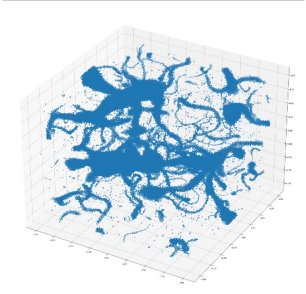

Figure 1: Swarm structure formed by large-scale simulation based on Reynolds' boids model.

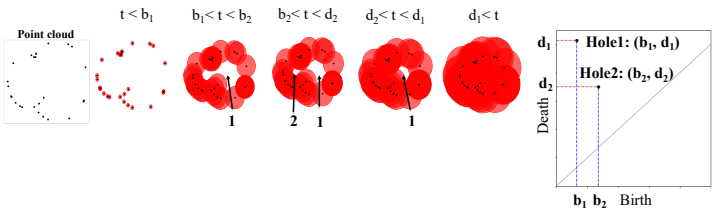


Figure 2: The left panel shows the evolution of the simplicial complex corresponding to the increase in radius, and the right panel shows the PD corresponding to the left panel's series of figures.

and the swarm structure cannot form well. This parameter, by taking an appropriately narrow range of interactions, allows for a large, hierarchical swarm structure with a small amount of computational cost. While there is no specific swarming phenomenon that corresponds to simulation of this parameter, it is excellent in that it allows discussion of the emergent nature of hierarchical structure in a swarming model, a model that reflects the nature of swarming.

## 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 [12]. One key feature used in TDA is persistent homology (PH). PH 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):

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

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

We first performed a TDA of a given snapshot according to the pattern dynamics analysis procedure proposed in [19]. The PD of the snapshot was calculated, and at the same time, the correspondence between the structure on the PD and the geometric structure of the swarm in real space was explored using the inverse analysis method [20, 21]. In the inverse analysis, the death simplices method was used to extract the structure corresponding to the moment the hole died [20, 21]. The analysis of snapshots helped understand the correspondence between PD and swarm structure.

Table 1: Parameters of this simulation

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| range of cohesion | 0.05 [unit] | angle of cohesion | $\pi/2$ |
| range of alignment | 0.05 [unit] | angle of alignment | $\pi/3$ |
| range of separation | 0.01 [unit] | angle of separation | $\pi/2$ |
| field size | 0.5 – 3.0 [unit] | number of individuals | 2,048 – 524,288 |
| max speed | 0.005 [unit/step] | min speed | 0.001 [unit/step] |
| $W_{\text{att}}$ | 0.008 | initial velocity | randomly distributed |
| $W_{\text{rep}}$ | 0.002 | initial position | randomly distributed |
| $W_{\text{ali}}$ | 0.06 | time step increment | 1 |
| density | 16,384 [num/unit$^3$] | boundary condition | 3D periodic |

Next, we analyzed the series of PDs that evolved over time according to the dynamics described in [19]. We combined 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 checked whether the PDs contained important information about swarm dynamics. For this purpose, we performed a regression analysis with the vectorized PD as the explanatory variable and the corresponding time as the objective variable. The vectorized PD, V(PD), is defined as,

$$
\begin{aligned}
V(PD) &:= \left\{ h(r_{\mathrm{birth}}, r_{\mathrm{death}}) |\ N_{\mathrm{grid}} r_{\mathrm{birth}}, N_{\mathrm{grid}} r_{\mathrm{death}} \in \mathbb{Z},\ r_{\mathrm{birth}}, r_{\mathrm{death}} \in \mathbb{Z}, [r_{\mathrm{min}}, r_{\mathrm{max}}] \right\}, \qquad (2) \\
h(r_{\mathrm{birth}}, r_{\mathrm{death}}) &:= \sum_{i=1}^{N_{\mathrm{hole}}} f\left(r_{\mathrm{birth}} - b_i, r_{\mathrm{death}} - d_i\right), \\
f(x, y) &= \begin{cases} 1, & \text{if } x \in \left[ -\dfrac{r_{\mathrm{max}} - r_{\mathrm{min}}}{N_{\mathrm{grid}}}, \dfrac{r_{\mathrm{max}} - r_{\mathrm{min}}}{N_{\mathrm{grid}}} \right] \text{ and } y \in \left[ -\dfrac{r_{\mathrm{max}} - r_{\mathrm{min}}}{N_{\mathrm{grid}}}, \dfrac{r_{\mathrm{max}} - r_{\mathrm{min}}}{N_{\mathrm{grid}}} \right] \\ 0, & \text{else} \end{cases}.
\end{aligned}
$$

where, $N_{\mathrm{grid}} = 128$ for intra-fluctuation analysis and $N_{\mathrm{grid}} = 12,800$ for inner-fluctuation analysis, which will be defined in the following section. The range $[r_{\mathrm{min}}, r_{\mathrm{max}}] = [\min r_{\mathrm{birth}}, \max r_{\mathrm{death}}]$ was used. We inversely regressed vectorized PDs to their corresponding time of simulation. The predictive performance of the regression model was evaluated using the fivefold cross-validation method [22]. If the model demonstrated sufficiently high performance, it suggests that dynamic information was embedded in the PD series.

After confirming that the PD series encompassed the information, the time variation of the distribution function, $h(r_{\mathrm{birth}}, r_{\mathrm{death}})$, 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, $h_{\mathrm{mean}}(r_{\mathrm{birth}}, r_{\mathrm{death}}) := \frac{1}{N_{\mathrm{time}}} \sum_{t=1}^{N_{\mathrm{time}}} h_t(r_{\mathrm{birth}}, r_{\mathrm{death}})$, where $h_t$ represents the function $h$ in Eq.(2) computed from the PD at a given time $t$. In this study, $N_{\mathrm{time}}$ was set to 100. We used $h_{\mathrm{mean}}$ as a reference, and the distances between each histgram $h_t$ and $h_{\mathrm{mean}}$ were evaluated using the 1-Wasserstein distance [13]. From the series of distances, information about the fluctuation structure of the swarm was extracted.

## 4 Results and discussion

By applying TDA to a snapshot of the swarm data, the PD 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). Upon

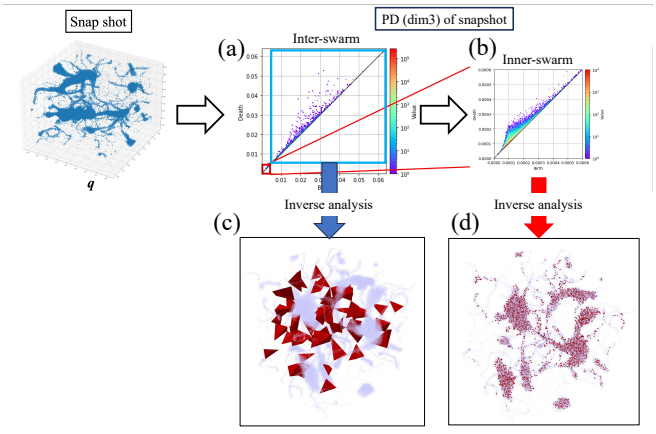


Figure 3: TDA of a snapshot of the swarms' structure. (a) PD of the snapshot. (b) Close-up of the PD's small radius region corresponding to the swarms' inner structure. (c) Results of the inverse analysis of the PD in the large radius region (blue box). (d) Results of the inverse analysis of PD in the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

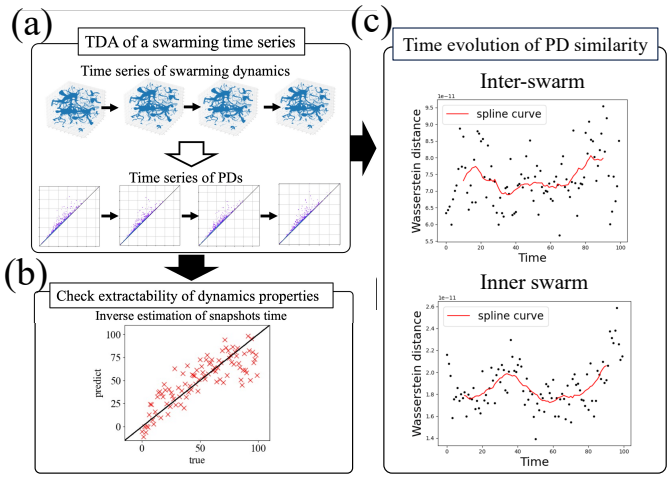


Figure 4: Analysis results of the time series of the swarming dynamics. (a) Example of a time series of a swarm structure and the corresponding PD series. (b) Regression result from the PD series with the corresponding time. The figure shows the predicted results for the test data using the cross-validation method. (c) Time evolution of the Wasserstein 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.

applying inverse analysis to the PD regions with these unique generators, highlighted by the blue and red rectangles in Fig. 3(a), it was confirmed that the long-lived generators in the large radius regions corresponded to the inter-group hole structure (Fig. 3(c)), while the long-lived generators in the small radius regions corresponded to the intra-group structure (Fig. 3(c)).

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 $12,800 \times 12,800$.

Finally, we analyzed the time evolution of the 1-Wasserstein distance between $\{H_t\}_{t=1}^{100}$ and $H_{\mathrm{mean}}$. 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. This finding is consistent with results from previous studies, which show that large swarms are driven by density fluctuations [10, 11]. 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 large-scale 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 [10, 11].

This study did not provide a statistically precise discussion. For example, while the presence of sufficient dynamical system information in the PDs was examined through regression analysis, the adequacy of the regression performance could have been assessed more robustly, such as by comparing it with random data. It is anticipated that refining these methods will facilitate the discovery of new insights in the analysis of large-scale swarm dynamics.

## Acknowledgment

This research was supported by the NEDO JPNP22100843-0, JST JPMJPR212A and JPMJCR2431, and JSPS 22K13979, 23H03460, 24K22309, and 24H00247.

## References

[1] Andrea Cavagna, Alessio Cimarelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107(26):11865–11870, 2010.

[2] Akira Okubo. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in biophysics*, 22:1–94, 1986.

[3] 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.

[4] Andreas Huth and Christian Wissel. The simulation of the movement of fish schools. *Journal of theoretical biology*, 156(3):365–385, 1992.

[5] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6):1226, 1995.

[6] Iain D Couzin, Jens Krause, et al. Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, 32(1):10–1016, 2003.

[7] Hugues Chaté, Francesco Ginelli, Guillaume Grégoire, Fernando Peruani, and Franck Raynaud. Modeling collective motion: variations on the vicsek model. *The European Physical Journal B*, 64:451–456, 2008.

[8] Shourav Pednekar, Ankita Jain, Olav Rune Godø, and Nicholas C Makris. Rapid predator-prey balance shift follows critical-population-density transmission between cod (gadus morhua) and capelin (mallotus villosus). *Communications Biology*, 7(1):1386, 2024.

[9] Bjørn Erik Axelsen, Leif Nøttestad, Anders Fernö, Arne Johannessen, and Ole Arve Misund. 'await'in the pelagic: dynamic trade-off between reproduction and survival within a herring school splitting vertically during spawning. *Marine Ecology Progress Series*, 205:259–269, 2000.

[10] Yhoichi Mototake and Takashi Ikegami. A simulation study of large scale swarms. *Proc. SWARM*, pages 446–450, 2015.

[11] 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.

[12] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[13] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[14] Golnar Gharooni-Fard, Morgan Byers, Varad Deshmukh, Elizabeth Bradley, Carissa Mayo, Chad M Topaz, and Orit Peleg. A computational topology-based spatiotemporal analysis technique for honeybee aggregation. *npj Complexity*, 1(1):3, 2024.

[15] 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.

[16] 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.

[17] Chad M Topaz, Lori Ziegelmeier, and Tom Halverson. Topological data analysis of biological aggregation models. *PloS one*, 10(5):e0126383, 2015.

[18] 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.

[19] Yoh ichi Mototake, Masaichiro Mizumaki, Kazue Kudo, and Kenji Fukumizu. Procedure to reveal the mechanism of pattern formation process by topological data analysis, 2024.

[20] Ippei Obayashi, Yasuaki Hiraoka, and Masao Kimura. Persistence diagrams with linear machine learning models. *Journal of Applied and Computational Topology*, 1(3):421–449, 2018.

[21] 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.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009.