

# Unsupervised manifold learning of collective behavior

Mathew Titus<sup>1,2\*</sup>, George Hagstrom<sup>3</sup>, James R. Watson<sup>1</sup>

**1** College of Earth, Ocean, and Atmospheric Sciences, 104 CEOAS Administration Building, Oregon State University, Corvallis, Oregon, 97331 USA

**2** The Prediction Lab LLC, Corvallis, Oregon, 97331 USA

**3** Dept. Ecology and Evolutionary Biology, 104a Guyot Hall, Princeton University, Princeton, New Jersey, 08544 USA

\*mat@thepredictionlab.com

## Abstract

Collective behavior is an emergent property of numerous complex systems, from financial markets to cancer cells to predator-prey ecological systems. Characterizing modes of collective behavior is often done through human observation, training generative models, or other supervised learning techniques. Each of these cases requires knowledge of and a method for characterizing the macro-state(s) of the system. This presents a challenge for studying novel systems where there may be little prior knowledge. Here, we present a new unsupervised method of detecting emergent behavior in complex systems, and discerning between distinct collective behaviors. We require only metrics,  $d^{(1)}$ ,  $d^{(2)}$ , defined on the set of agents,  $X$ , which measure agents' nearness in variables of interest. We apply the method of diffusion maps to the systems  $(X, d^{(i)})$  to recover efficient embeddings of their interaction networks. Comparing these geometries, we formulate a measure of similarity between two networks, called the map alignment statistic (MAS). A large MAS is evidence that the two networks are codetermined in some fashion, indicating an emergent relationship between the metrics  $d^{(1)}$  and  $d^{(2)}$ . Additionally, the form of the macro-scale organization is encoded in the covariances among the two sets of diffusion map components. Using these covariances we discern between different modes of collective behavior in a data-driven, unsupervised manner. This method is demonstrated on empirical fish schooling data. We show that our state classification subdivides the known behaviors of the school in a meaningful manner, leading to a finer description of the system's behavior.

## Author summary

Many complex systems in society and nature exhibit collective behavior where individuals' local interactions lead to system-wide organization. One challenge we face today is to identify and characterize these emergent behaviors, and here we have developed a new method for analyzing data from individuals, to detect when a given complex system is exhibiting system-wide organization. Importantly, our approach requires no prior knowledge of the fashion in which the collective behavior arises, or the macro-scale variables in which it manifests. We apply the new method to an agent-based model and empirical observations of fish schooling. While we have demonstrated the utility of our approach to biological systems, it can be applied widely to financial, medical, and technological systems for example.

# Introduction

1

Collective behavior is an emergent property of many complex systems in society and  
nature [1]. These behaviors range from the coordinated behaviors of fish schools, bird flocks,  
and animal herds [2–4], to human social dynamics such as those of traders in stock markets  
that have led to bubbles and crashes [5] in the past, and opinions on social networks [6].  
Collective behavior amplifies the transfer of information between individuals and enables  
groups to solve problems which would be impossible for any single group member alone [7].  
Furthermore, manifestations of collective behavior in both social and ecological systems  
regularly influence our individual welfare [8–10] through political and economic instability,  
disease spread, or changes in social norms. Building a better classification and  
understanding of collective behavior and collective states is crucial not only to basic science,  
but to designing a stable future.

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12

A key challenge in the study of complex systems is to identify when collective behavior  
emerges, and in what way it manifests [11, 12]. Statistical physicists first studied emergence  
in complex systems in the 1800s where they began to derive the physical laws governing the  
macroscopic behavior of systems with an extremely large number of degrees of freedom from  
the interactions between the microscopic, atomistic components of the system. Remarkably,  
there usually exist a small number of macroscopic variables or order parameters that  
accurately describe macroscale dynamics for numerous systems, and a small number of  
relationships between these variables characterize the different states or behavioral regimes  
of the system. In many systems, the key macroscopic variables are well known from either  
first principles or empirical study. However, there are many social, ecological, and even  
physical systems which elude simple description by macroscopic variables. Furthermore, we  
continually encounter novel systems for which we have no prior knowledge about multiscale  
dynamics, or models that give rise to unknown group dynamics [13]. In these cases, it is a  
non-trivial task to identify useful variables (at particular scales of organization) which define  
the possible emergent behaviors.

13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

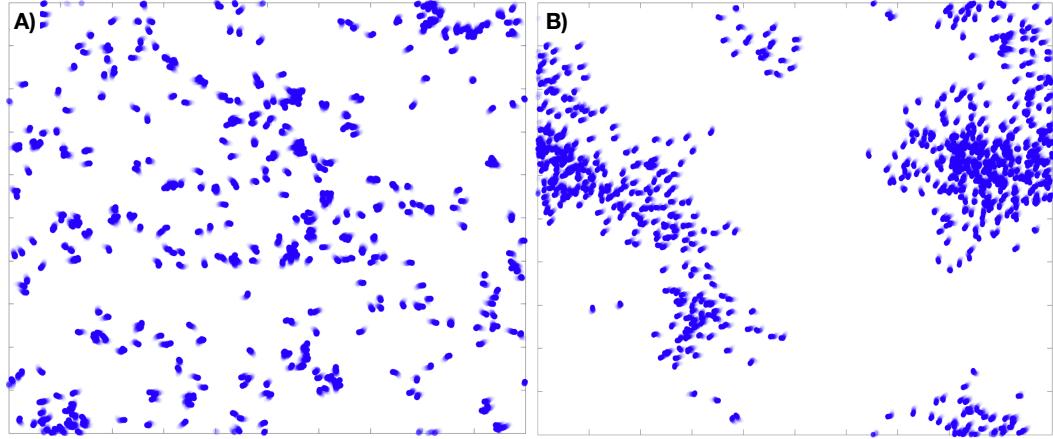
Machine learning and other statistical methods have emerged as key tools for finding  
macroscopic descriptions of complex systems with many degrees of freedom, in the physical,  
ecological, and social sciences. The application of these tools has been driven by high  
resolution observations of microscopic degrees of freedom (particularly in ecological and

28  
29  
30  
31

social systems), increases in computing power, and the development of new algorithmic tools. Dimensionality reduction, clustering algorithms, and other unsupervised learning algorithms have been particularly successful in discovering macroscopic variables and behavioral regimes from microscopic data on complex systems [14]. The paramagnetic/ferromagnetic phase transition in the classical, 2D Ising model has been discovered by algorithms ranging in complexity from principal component analysis to variational autoencoders [14–16]. However, more complex types of collective phenomenon, such as topological phases, have not been successfully identified using algorithms like PCA, which assume that the data has a linear structure. In these more difficult cases, algorithms which can detect nonlinear structures such as diffusion maps have shown much greater promise. For example, diffusion maps have been used to find the Kosterlitz-Thouless phase transition in the XY-model from condensed matter physics [17]. The success of machine learning algorithms for physical problems suggests that machine learning could be of great practical use for the study of collective phenomena in social and ecological systems, which are usually much less amenable to theoretical analysis, and for which there often exists a significant quantity of data at the individual level.

Here we have developed an approach to analyzing collective behavior in biological systems based on dimensional reduction tools from manifold learning [18], specifically diffusion maps [19]. Using diffusion maps to formulate multiple data-driven coordinate systems, we compare their large-scale structures to find latent global relationships between different variables observed at the level of individual agents. We call this the map alignment between the two coordinate systems. When two variables serve to globally organize the data along similar lines, their map alignment is high. One can easily calculate the expected alignment between unrelated and independent variables, so this gives us a test that detects a system’s emergent behaviors but requires no prior knowledge of them. Moreover, the ways in which the coordinates are correlated leave a signature for the relationship between the macro-scale structures, and one may use this to distinguish various system-wide behaviors. We apply this tool to modeled data produced from simulations of birds flocking, and to empirical data on the activity of fish in a school. Using the micro-scale data to cast the agents’ network of interaction into multiple geometries, each evolving over time, we both detect the emergence of macro-scale organization as well as perform a meaningful classification of distinct modes of organization, identifying the macro-scale states of the

system. Importantly, this analysis is based purely on changes occurring at micro-scales, and  
 makes no assumptions about which macro-scale variables are important. This means that  
 for any novel system of study, where there is a paucity of knowledge regarding which  
 macro-scale variable to measure and track, one could use this approach to classify and  
 observe the dynamics of the system at different levels of organization.



**Fig 1.** Examples of an 800 agent flocking model on the torus, with and without long-range order, that is, in both incoherent (A) and coherent (B) macro-states.

## Materials and methods

### Diffusion Maps

We begin with the basic definitions required to construct a diffusion map. This requires a set of data points,  $X$ , and a metric on those points,  $d$ . The data may be composed of many observables, making each  $x_k \in X$  a vector in a space of potentially very high dimension.

Essentially we employ the algorithm in [20], with  $\alpha = 1$  and  $\epsilon = 1$ . This involves

- (i) constructing a similarity measure,  $K(x_i, x_j)$ , between pairs of data points,
- (ii) normalizing  $K$  to create a Laplacian operator  $\hat{A}$  on the data,
- (iii) computing eigenvector-eigenvalue pairs  $\{(\phi_k, \lambda_k)\}$  for the Laplacian, and
- (iv) constructing the diffusion map  $\Phi(t, x) := (\lambda_k^t \phi_k(x))_{k=1}^n$ .

The choice of similarity measure in (i) is highly system-dependent, for agent-wise interactions may manifest themselves in many different ways and so require different metrics.

Symbol Reference	
$X = \{x_k\}_{k=1}^n$	data set
$n$	number of data points
$[k]$	the set of integers $1, 2, \dots, k$
$f : X \rightarrow \mathbb{R}^m$	micro-scale observable
$d(\cdot, \cdot)$	metric on $X$
$D$	pairwise distance matrix
$A$	affinity matrix
$\hat{A}$	heat operator
$\mathcal{D}_t(\cdot, \cdot)$	diffusion distance on $X$
$\{\phi_k\}$	eigenvectors of $\hat{A}$
$\{\lambda_k\}$	eigenvalues of $\hat{A}$
$\Phi(t, \cdot)$	diffusion map at scale $t$
$\{\psi_k(t, \cdot)\}$	diffusion coordinates of $\Phi(t, \cdot)$
$\xi_{i,j}$	inner product of $\psi_i^{(1)}$ and $\psi_j^{(2)}$

**Table 1.** List of symbols used.

In (ii), the choice of normalization in constructing  $\hat{A}$  is subtler, and based on a choice of Laplace operator; this is not sensitive to the data under observation, but rather reflects an assumption of Manifold Learning, that the data points of  $X$  are drawn from a manifold embedded in euclidean space. For rigorous results on when such a manifold exists for a given dataset, see [21].

Given  $n$  points  $X = \{x_1, \dots, x_n\}$ , we construct an  $n \times n$  matrix of pairwise distances,  $D = (D_{i,j}) := (d(x_i, x_j))$ , associated to a metric  $d$ . Writing  $k(x)$  for the  $k^{\text{th}}$  nearest neighbor (kNN) of  $x \in X$ , let  $K(x, y)$  be the gaussian kNN (GkNN) kernel defined by

$$K(x, y) = \exp\left(-\frac{d(x, y)^2}{d(x, k(x))d(y, k(y))}\right). \quad (1)$$

This kernel could be replaced with any other, as another function may better reflect the user's impression of the similarity between two agents a distance  $d(x, y)$  apart. This particular choice is studied in [22], and serves to allow pairwise distances to scale according to their local geometries. One could consider this the standard gaussian kernel applied to the modified distance function  $d(x, y)/\sqrt{d(x, k(x))d(y, k(y))}$ . The more sparse the point cloud  $(X, d)$  is near  $x$  or  $y$ , the smaller this distance becomes. This kernel also has the desirable property that it is scale-free, in that multiplying all distances by a fixed constant does not change the kernel  $K$ .

Let  $\sigma$  denote a row sum operator: if  $A$  is an  $m \times n$  matrix, let  $\sigma(A)$  be the  $m$ -vector

whose  $i^{\text{th}}$  component is  $\sigma(A)_i := \sum_{j=1}^n A_{i,j}$ . We next calculate an affinity matrix 98

$A = (A_{i,j}) := (K(x_i, x_j))$ , its row sums  $\sigma(A)$ , and construct the matrix operator  $A'$  via 99

$$A'_{i,j} = (\sigma(A)_i^{-1} A_{i,j} \sigma(A)_j^{-1}).$$

Normalizing by the row sums in this fashion is a particular choice of Laplace operator on 100  
the graph whose vertices represent the  $n$  agents and whose weighted edges are given by the 101  
entries of  $A'_{i,j}$  (see panel A of Figure 2). This operator converges pointwise to the 102  
Laplace-Beltrami operator in the case that the points  $X$  are drawn from a probability 103  
distribution on a manifold (see [20]). We then define the Markov operator  $\hat{A} = (\hat{A}_{i,j})$  by 104  
setting the row sums of  $A'$  to one: 105

$$\hat{A}_{i,j} := \sigma(A')_i^{-1} A'_{i,j}. \quad (2)$$

This is the operator of primary interest to us, representing the evolution of a random 106  
walk on the graph, where the probability of transitioning from point  $x_i$  to  $x_j$  is given by 107  
 $\hat{A}_{i,j}$ . This is visualized in panel B of Figure 2 as heat propagating through the network. 108  
From the spectral structure of  $\hat{A}$ , written  $\{(\phi_k, \lambda_k)\}_{k=1}^n$  with  $\phi_k : X \rightarrow \mathbb{R}$  the unit 109  
eigenvectors and  $\lambda_k \in \mathbb{R}$  the eigenvalues, we construct the diffusion map, 110

$$\Phi(t, x) = (\lambda_1^t \phi_1(x), \dots, \lambda_n^t \phi_n(x)), \quad t \geq 0. \quad (3)$$

The usefulness of this decomposition of the data  $(X, d)$  is that it gives an embedding of 111  
the data which reflects the *diffusion distance*,  $\mathcal{D}_t$ , defined as the  $L^2$  difference between the 112  
random walk's distributions when initiated from site  $x$  and site  $y$  after  $t$  time has passed: 113

$$\begin{aligned} \mathcal{D}_t(x, y)^2 &:= \sum_{z \in X} (\hat{A}_{x,z}^t - \hat{A}_{y,z}^t)^2 \\ &= \|\Phi(t, x) - \Phi(t, y)\|_2^2. \end{aligned}$$

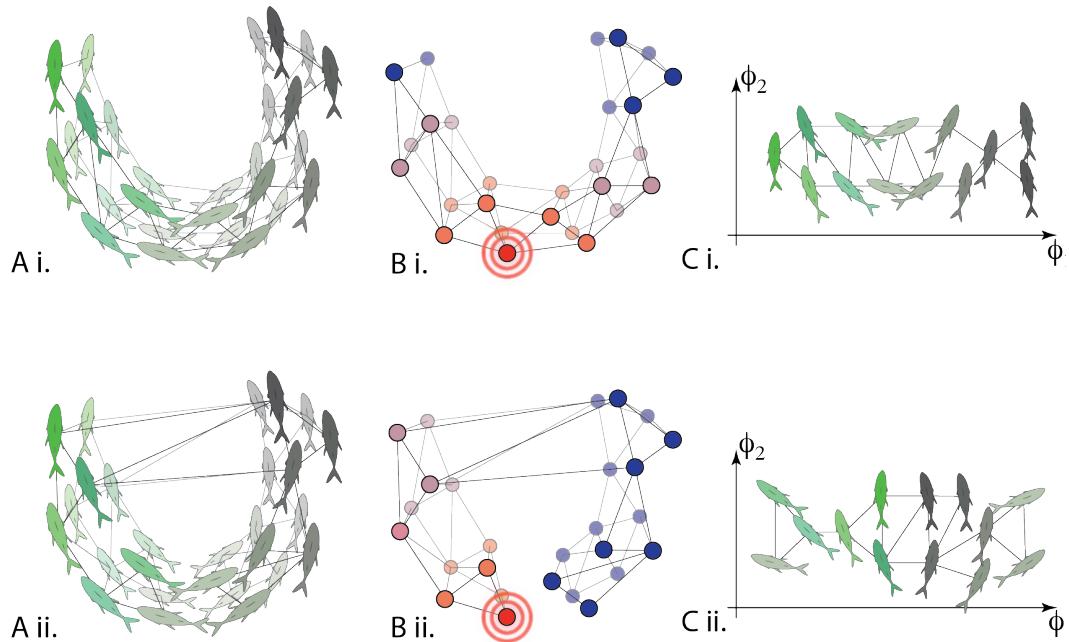
In essence,  $\mathcal{D}_t$  captures the geometric structure of the data by measuring the 114  
(dis)similarities of trajectories of random walks evolving on the data points from different 115  
starting points. The diffusion map is an isometry of  $(X, \mathcal{D}_t)$  into  $\mathbb{R}^n$ , with the remarkable 116

benefit that truncating  $\Phi(t, \cdot)$  to its first  $k$  coordinates gives the optimal  $k$ -dimensional embedding of  $X$  into  $\mathbb{R}^k$  (envisioned in panel C of Figure 2). These truncated maps 117  
 $\Phi_k(t, \cdot) := (\lambda_1^t \phi_1, \dots, \lambda_k^t \phi_k)$  are utilized below for detecting macro-scale relationships among 118  
variables. 119  
120

We remark that though increasing  $t$  decreases the resolution of small scale features in the embedding, in practice we take  $t = 1$  throughout and define  $\psi_k(x) = \lambda_k \phi_k$ , so that

$$\Phi_k(x) := \Phi_k(1, x) = (\psi_1(x), \dots, \psi_k(x)), \quad k \in \{1, \dots, n\},$$

and  $\mathcal{D} := \mathcal{D}_1$ . 121



**Fig 2.** Conceptual illustration of diffusion maps applied to data on fish schooling. A) For the same time period, different affinity operators can be calculated. These affinities could be correlations in simply spatial proximity (Ai) or correlations in velocities (Aii) for example. The operator is used to describe a network of agents. B) Once calculated, the heat diffusion operator allows us to quantify the geometry of the network and with this in hand C) diffusion map coordinates may be used to efficiently embed the fish in a low-dimensional space. Doing this for different metrics / affinities allows us to quantify the alignment between various diffusion map coordinate systems, which reveals changes in macro-scale behavior.

## Map Alignment

122

In this section we formulate a statistic to measure the degree to which the macro-scale organization of two observables measured on the micro-scale level are related. Let  $X = \{x_1, \dots, x_n\}$  be the collection of data for the  $n$  interacting agents making up the system, and let  $f_1$  and  $f_2$  be two vector-valued observables, such as position or velocity, defined on each agent,

$$f_i : X \rightarrow \mathbb{R}^{m_i}, m_i \in \mathbb{N}.$$

Suppose we are given metrics  $d_i : X \times X \rightarrow \mathbb{R}$  between agents depending only on the observed variable  $f_i$ , such as

$$d_i(x_j, x_k) := \|f_i(x_j) - f_i(x_k)\|_2. \quad (4)$$

In practice these are defined by the user, and a cogent choice will depend upon their knowledge of the agent-level interactions of the system. Using the methods of the previous subsection, we devise a pair of Markov operators,  $\hat{A}^{(i)}$  (taking  $i = 1, 2$ ), their associated eigenstructures,  $\left\{(\phi_j^{(i)}, \lambda_j^{(i)})\right\}_{j=1}^n$ , and the corresponding diffusion maps,  $\Phi^{(i)} = (\psi_j)_{j=1}^n$ .

For each  $i = 1, 2$  the diffusion coordinates  $\{\psi_j^{(i)}\}$  are ordered by decreasing contribution to the global organization of the network  $X$  under the heat operator  $\hat{A}^{(i)}$ . Therefore if there is a coherent large-scale organization on  $X$  manifest in the observables  $f_i$ , then the leading diffusion coordinates of  $\Phi^{(1)}$  should be determined in some manner by the leading coordinates of  $\Phi^{(2)}$ . To confirm this, we expand the unit vectors  $\phi_j^{(1)}$  in the subspace generated by the top  $k$  diffusion coordinates of  $\hat{A}^{(2)}$ , i.e. we project the eigenvectors into the subspace spanned by  $\{\phi_1^{(2)}, \dots, \phi_k^{(2)}\}$  via the linear operator

$$\Pi_k^{(2)}(\phi) = \sum_{j=1}^k \langle \phi, \phi_j^{(2)} \rangle \phi_j^{(2)}. \quad (5)$$

Then  $\|\Pi_k^{(2)}(\phi_j^{(1)})\|$  is a measure of how well the coordinate  $\phi_j^{(1)}$  associated to  $f_1$  is correlated with the large-scale network structure induced on  $X$  by the variable  $f_2$ . One attempts to choose  $k$  in a manner that ensures that the subspace spanned by  $\{\phi_1^{(2)}, \dots, \phi_k^{(2)}\}$  encompasses the relevant macro-scale coordinates induced by  $f_2$  on  $X$ , but

without  $k$  becoming so large that we begin to recover a significant portion of  $\phi$  simply due to the large dimension of the image space. For example, if we take  $k = n$  the projection  $\Pi_n^{(2)}$  is the identity, simply expressing  $\phi$  in the basis  $\{\phi_j^{(2)}\}$ . However, taking  $k \ll n$  we recover only the portion of the vector which can be expressed as a linear combination of the top  $k$  diffusion coordinates of  $\Phi^{(2)}$ . We return to this issue below.

Now we derive the expected projection size under a null hypothesis that the two variables  $f_1$  and  $f_2$  are entirely unrelated. Given the pair of diffusion bases, we fix  $\phi \in \{\phi_j^{(1)}\} \subset \mathbb{R}^n$  and calculate the distribution of the random variable  $\|\Pi_k^{(2)}\phi\|$  assuming that the  $k$ -dimensional range,  $\text{Span}(\phi_1^{(2)}, \dots, \phi_k^{(2)})$ , is chosen uniformly at random from the set of all  $k$ -dimensional subspaces. One can show, and with a little thought it is clear, that this is equivalent to asking for the magnitude of a random unit vector of  $\mathbb{R}^n$  projected onto its leading  $k$  coordinates. When the vector is sampled uniformly at random from  $\{\psi \in \mathbb{R}^n : \|\psi\|_2 = 1\}$ , the quantity  $\|\Pi_k^{(2)}\phi\|^2$  is a beta distributed random variable with parameters  $(n - k)/2$  and  $k/2$ . Recognizing this, one can calculate the expectation and variance to be  $k/n$  and  $2(n - k)k/(n^3 + 2n^2)$ , respectively. Normalizing the projection size accordingly, we arrive at the statistic

$$\mathcal{P}_k^{(2)}(\phi) := \left( \|\Pi_k^{(2)}\phi\|^2 - \frac{k}{n} \right) \sqrt{\frac{n^3 + 2n^2}{2(n - k)k}},$$

which quantifies how well we can express the information of  $\phi$  using only the random subspace generated by  $\{\phi_1^{(2)}, \dots, \phi_k^{(2)}\}$ .

Again, the leading diffusion coordinates of  $\Phi^{(i)}$  are data-driven macro-scale variables, expressing the large-scale organization of the network as viewed through  $\hat{A}^{(i)}$ . If the data obeys a (possibly nonlinear and multi-valued) relationship between  $f_1$  and  $f_2$  on the macro-scale, then it should be evidenced by a relationship between the bases  $\{\phi_j^{(1)}\}$  and  $\{\phi_j^{(2)}\}$ . Note that the observables  $f_i$  are used only in providing a metric on  $X$ . One can perform an identical analysis on the data given only the metrics  $d_1$  and  $d_2$ . In the context of this paper we will assume, however, that the metrics come from some micro-scale variable defined on  $X$ , as this is the approach most useful in analyzing a new complex system.

Let  $u = (\lambda_1^{(1)}, \dots, \lambda_k^{(1)})$  be the vector of decreasing eigenvalues of  $\hat{A}_1$ , and take  $\hat{u} := u/\|u\|_2$ . Then the entries of  $\hat{u}$  give a normalized notion of the energy of each  $\phi_j^{(1)}$ , and so gives a measure of how important each mode  $\phi_j^{(1)}$  is to the macrostructure derived from

$f_1$ . Finally, we define the **map alignment statistic** to be the sum of the  $Z_k^{(2)}(\phi_j^{(1)})$ ,  
 weighted by the contribution of  $\phi_j^{(1)}$  to the large-scale organization of  $f_1$ :  
169  
170

$$Z_k(f_1, f_2) := \sum_{j=1}^k \hat{u}_j Z_k^{(2)}(\phi_j^{(1)}). \quad (6)$$

We remark that this statistic is not in general symmetric;  $Z_k(f_1, f_2) \neq Z_k(f_2, f_1)$ . The MAS  
 $Z_k(f_1, f_2)$  is a measure of how much of the macrostructure of  $f_1$  is reflected in the  $f_2$ -based  
 diffusion coordinates. For example, the spectrum of the  $\hat{A}^{(i)}$  may indicate that the top three  
 diffusion coordinates of  $\Phi_t^{(1)}$  carry a greater proportion of the macro-scale information of  
 $\hat{A}^{(1)}$  than the top three coordinates of  $\Phi_t^{(2)}$  do for  $\hat{A}^{(2)}$ ; in this case one would expect that  
 $Z_3(f_1, f_2) > Z_3(f_2, f_1)$ .  
171  
172  
173  
174  
175  
176

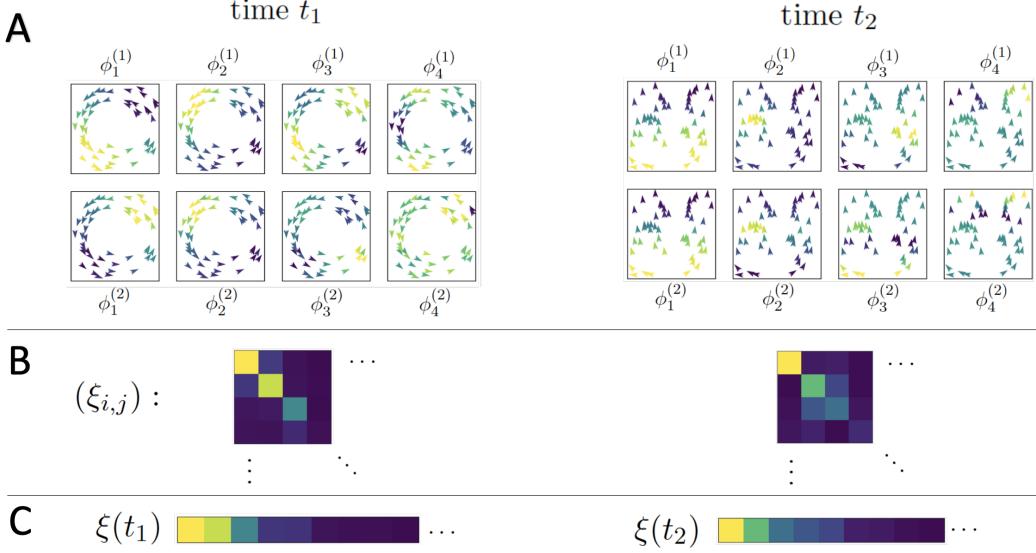
## Informative choice of $k$

177

Here we briefly discuss how to choose  $k$  so that  $Z_k$  gives an accurate measure of the degree  
 to which  $\Phi^{(1)}$  is determined by  $\Phi_k^{(2)}$ . Recall that the head of the spectrum of  $\hat{A}$  is sensitive  
 to large-scale changes, but the bulk of the spectrum is associated to high frequency  
 eigenfunctions, capturing small-scale organization often attributable to noise or local  
 idiosyncrasies in the data. As a result, given training data or simulated data one may  
 simply estimate a value of  $k$  after which the variance explained by the  $j^{\text{th}}$  eigenspace,  
 $\lambda_j / \sum_{i=1}^n \lambda_i$ , is approximately independent of the macrostate of the system. That is, choose  
 $k$  so that for  $j > k$ , the value  $\lambda_j$  makes up a fixed fraction of the sum of the eigenvalues.  
178  
179  
180  
181  
182  
183  
184  
185

Alternatively, in using empirical data we found the bulk of the spectrum often follows an  
 approximate power-law decay. As a result, one may take  $k$  to be the smallest integer such  
 that for  $j > k$ ,  $\lambda_j \approx C j^{-\beta}$  for positive constants  $C, \beta$ .  
186  
187  
188

These are necessarily rough guidelines, and in practice one often finds a range of  
 plausible choices for  $k$ . However, when increasing  $k$  the additional summands of  $\Pi_k^{(2)}$  must  
 eventually become essentially independent ( $\langle \phi_i^{(1)}, \phi_k^{(2)} \rangle \approx 0$ ) of the leading elements of  
 $\{\phi_i^{(1)}\}$ . It follows that after a point, a larger value of  $k$  has little influence on the computed  
 MAS,  $Z_k(f_1, f_2)$ . Therefore an overestimate of  $k$  has very little effect on the results, so it is  
 advised that in practice one err on the side of larger choices of  $k$ .  
189  
190  
191  
192  
193  
194



**Fig 3.** A toy system of 50 agents; the left and right sides of the figure represent observations of the system at two different times when the system is in two different organized states. (A) Agents in each subpanel are colored by the indicated diffusion coordinate, (B) A color plot of the matrix  $(\xi_{i,j})$  with yellow indicating larger values, blue indicating smaller values, (C) The values of the above inner product matrix, ordered by absolute magnitude; these are the vectors used to classify the system's regimes.

## Unsupervised Macrostate Classification

195

If the system under study exhibits multiple emergent behaviors relating the two variables  $f_1$  196 and  $f_2$ , then coherence alone is not enough to identify system behavior. In passing to the 197 norm of the vector  $\Pi_k^{(2)} \phi_j^{(1)}$ , we clearly give up a significant amount of information. In order 198 to distinguish multiple coherent states, we inspect the individual dot products  $\{\langle \phi_i^{(1)}, \phi_j^{(2)} \rangle\}$  199 that comprise the map alignment statistic. 200

However, to remove the effects of the trailing diffusion coordinates, which correspond to 201 smaller scale variations in the geometry of  $(X, \mathcal{D}^{(i)})$ , we consider the sequence of inner 202 products given by 203

$$\begin{aligned} \xi_{i,j} &:= \langle \psi_i^{(1)}, \psi_j^{(2)} \rangle \\ &= \lambda_i^{(1)} \lambda_j^{(2)} \langle \phi_i^{(1)}, \phi_j^{(2)} \rangle \end{aligned}$$

with  $1 \leq i, j \leq k \ll n$ . That is, we collect the covariances between coordinates of the 204 truncated diffusion maps  $\Phi_k^{(1)}$  and  $\Phi_k^{(2)}$ . Given time series of micro-scale observations, this 205 allows us to characterize each frame by the sequence  $(\xi_{i,j})_{1 \leq i, j \leq k}$ . We will add an argument 206

to indicate the timestep  $t$ , as  $\xi_{i,j}(t)$ , when necessary.

207

While the diffusion coordinates provide a natural organization of the data, the individual components of the map are eigenvectors, and so are computed via an iterated maximization problem: given the linear operator  $A$ , the  $i^{\text{th}}$  eigenvector is given by

$$\phi_i = \max_{v \in U_{i-1}^\perp} \langle v, Av \rangle,$$

where  $U_j^\perp := \{u \in \text{Span}\{\phi_1, \dots, \phi_j\} : \|u\| = 1\}$ . Such maximization problems are prone to discontinuities, in the sense that there exist, for any threshold  $\varepsilon > 0$ , linear operators  $A$  and  $B$  on  $\mathbb{R}^n$  with  $i^{\text{th}}$  eigenvectors  $\phi_i$  and  $\eta_i$ , respectively, such that  $\|A - B\| < \varepsilon$ , while  $\|\phi_i - \eta_i\| > 1/2$  for all  $1 \leq i \leq n$ . This means that the individual indices of the  $\xi_{i,j}$  may not correspond to a persistent structure in the data.

Therefore, for each collection  $\{\xi_{i,j} : 1 \leq i, j \leq k\}$  we define the map  $\sigma : [k] \times [k] \rightarrow [k^2]$  so that the components of  $(|\xi_{\sigma(i,j)}|)$  are ordered by decreasing magnitude. (Here  $[k]$  denotes the first  $k$  positive integers.) This removes the ambiguity in both index and sign for these products. We call the resulting vector the  $f_1 f_2$ -covariance vector (at time  $t$ ), abbreviated as  $\xi$  (resp.  $\xi(t)$ ). One may then use these vectors as a fingerprint for the macro-scale relationship exhibited between  $f_1$  and  $f_2$  at a given time.

Figure 3, panel A, gives an example of a toy flocking system comprised of 50 agents, exhibiting a circling behavior at time  $t_1$  and a (mostly) aligned state at time  $t_2$ ; its top four diffusion coordinates associated to both  $\Phi^{(1)}$  and  $\Phi^{(2)}$  are displayed for each time. (The choice of affinities  $\hat{A}^{(1)}$  and  $\hat{A}^{(2)}$  here are the same as in the flocking model below, being based on spatial proximity and velocity, respectively.) In panel B we see the absolute values of the inner products of the diffusion coordinates, where the matrix (represented by a color plot) has its  $(i, j)$ -entry given by  $\langle \psi_i^{(1)}, \psi_j^{(2)} \rangle$ . Finally, in panel C we see the covariance vectors at times  $t_1$  and  $t_2$ .

Covariance vectors allow us to embed the activity of the agents in a shared space,  $\mathbb{R}^{k^2}$ , and perform clustering analyses to group similar macro-scale relationships between  $f_1$  and  $f_2$  over time. In the next section we will perform a boilerplate clustering analysis to distinguish between distinct coherent group structures found in a data set of golden shiner schooling behavior, a task for which the map alignment statistic is insufficient on its own.

# Results

232

## Analysis of Bird Flocking Simulations

233

As a first test of our method we explored a well studied agent-based model of flocking behavior [23]. The model consists of a set of  $N$  agents,  $\{x_j\}$ , moving on the torus  $T^1 = S^1 \times S^1$ , each with position  $p_i$  and velocity determined by the angle  $\theta_i$  and the magnitude  $v_i$ . In this model each agent moves with a fixed speed,  $v_i = v_j$  for all  $i$  and  $j$ , so we simply write  $v$  for this speed.

Each agent's direction is updated at each time step by the average local direction plus a noise term, a normal random variable distributed as  $N(0, \epsilon^2)$ . Here local refers to the agents within a certain distance  $r$  of the focal agent. With all other parameters of the model fixed, the coupling parameter  $r$  governs the degree to which the system self-organizes. In the hydrodynamic limit, long-range order emerges for  $r$  above some critical value  $r_c(v, \epsilon) > 0$ , meaning that the mean velocity of the flock is nonzero:  $\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{i=1}^N (\cos \theta_i, \sin \theta_i) \right\| \neq 0$ . The direction of group travel is chosen randomly, and breaks the symmetry of the system. See the Supplementary Information (SI) for details of the model, video, and pseudocode.

In the discrete time, finite-agent setting we observe the mean velocity of the flock,  $\langle v_j \rangle = (v \cos \langle \theta_j \rangle, v \sin \langle \theta_j \rangle)$ , taking this as our ‘ground truth’ measurement of the level of emergent, coherent behavior in the system. As  $r$  increases past  $r_c(v, \epsilon)$ , the mean velocity (in the large  $N$  limit) transitions from zero to a nonzero value. However, this shift occurs continuously; as  $r$  grows we find the steady-state mean velocity of the flock increases towards its maximum value. Even with  $r = 0$  the global mean velocity will have magnitude  $O(1/\sqrt{N})$  and suffer fluctuations, so determining whether a preferred flock-wide direction of motion has been chosen in the finite case makes little sense. Instead, below we compare the value of  $Z_k(p, v)$  to  $\langle v_j \rangle$  to determine how well the MAS detects large-scale organization.

We chose two metrics to define a pair of time-evolving graphs as follows: Since we are interested in flocking behavior, we let  $d^{(1)}(x_i, x_j; t)$  be the euclidean distance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  agents at time  $t$ , while  $d^{(2)}(x_i, x_j; t)$  denotes the kernel-smoothed  $L^2$  norm of the two agents' relative speed profiles. (See (S3), (S4) and (S5) for the explicit definitions of the metrics used.) In particular,  $d^{(1)}$  captures the spatial information in the flock, while  $d^{(2)}$  measures the degree that agents are moving in concert over a short (on the order of 10 time steps) recent period.

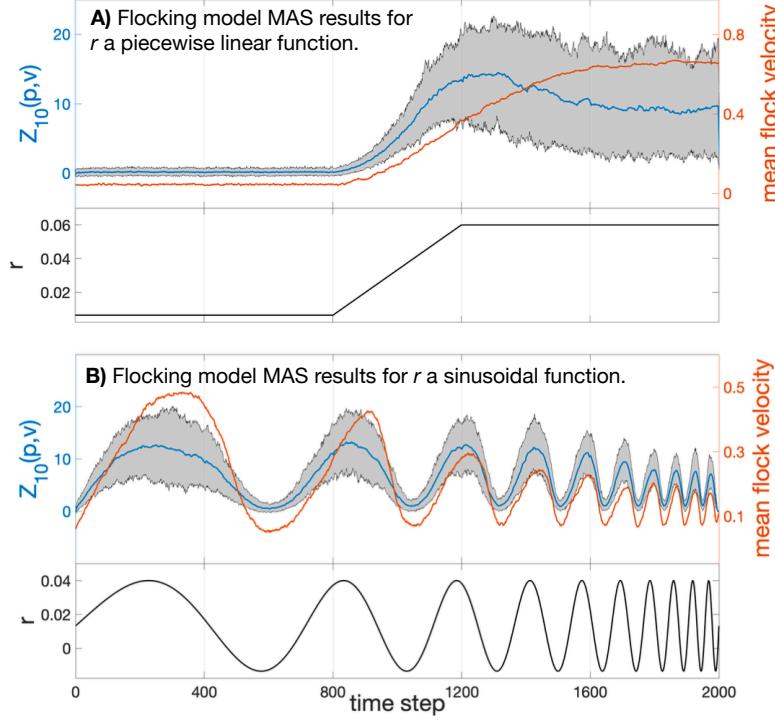
We then applied diffusion maps to the resulting sequence of graphs and calculated the map alignment between the two variables,  $p$  and  $v$ . The choice of subspace size to use,  $k = 10$ , was empirical. We observe that at various levels of organization the variance explained by the first five modes (i.e. the sum of the magnitudes of the first four components of  $\hat{u}$ ) is highly variable, while the variance explained by the higher modes ( $\phi_6^{(i)}, \phi_7^{(i)}, \dots, \phi_{500}^{(i)}$ ) shows little dependence on system organization, though we decided to increase  $k$  to 10 to be conservative. This suggests that the macro-scale information is well captured by  $\Phi_{10}^{(i)}$ , while components 11 through 500 are largely accounting for noise and small-scale features of  $\hat{A}^{(i)}$ . Note, however, that one may take  $k$  larger than 50 without any qualitative change in the analysis.

In Figure 4 below we plot the map alignment statistic  $Z_{10}(p, v)$  (blue) calculated from two sets of 100 independent simulations of the flocking model with 500 agents and fixed parameters ( $\epsilon = \pi/5, v = 1/320$ , see SI). In the left panels the coupling parameter  $r$  is held at a low value ( $r = r_0 = 0.0065$ ) for 800 time steps before increasing linearly to a high value over 400 steps and remaining at that value ( $r = r_1 = 0.06$ ) for the rest of the trial; see the lower panel of Figure 4(A). The mean velocity  $\langle v_j \rangle$  of the flock, averaged over the 100 trials, is plotted in orange in the upper panel.

One can see that the statistic  $Z_{10}(p, v)$ , measuring the dependence of position diffusion coordinates on the velocity diffusion coordinate system, has a strong correspondence with the velocity correlation statistic. That is, large-scale organization was clearly detected and quantified using only the two micro-scale inter-agent distance time series.

In a second trial we explored how quickly map alignment and long-range correlations vary in response to change in the micro-scale dynamics of the agents. This is done by allowing  $r$  to vary sinusoidally between  $r_0$  and  $r_1$ , with increasing frequency. The black line plotted in the lower panel of Figure 4(B) shows the coupling parameter as a function of time; in the upper panel the orange line shows the magnitude of the mean velocity of the flock, and the blue plot gives the mean value of  $Z_{10}$  across trials, with the filled area indicating the 90% confidence interval for  $Z_{10}$ .

Clearly  $Z_{10}(p, v)$  indicates strongly the presence or absence of correlation. The map alignment statistic shows both a faster response to changing  $r$  and suffers from less attenuation as the speed of parameter change increases. Times  $t = 800$  to 1000 especially highlight how the drop in  $Z_{10}$  is a precursor to the value of the macrovariable  $\langle v_j \rangle$  falling.



**Fig 4.** Top: Map alignment statistic  $Z_{10}(p, v)$  averaged over 100 flocking simulations of 2000 time steps (blue). The grey region denotes the 90% empirical confidence interval of  $Z_{10}(p, v)$ . The mean velocity  $\langle v_j \rangle$  of the flock is averaged over the 100 trials and plotted in orange. Bottom: The model's coupling constant  $r$  plotted over time. Notice that in (B) the MAS does not suffer the same hysteresis that the flock's speed does.

In this case and the previous case, we find we are able to track the transition from  
295  
unorganized motion to flocking by the alignment between the velocity diffusion coordinates,  
296  
 $\Phi_{10}^{(1)}$ , and the position diffusion coordinates,  $\Phi_{10}^{(2)}$ .  
297

### Analysis of Empirical Fish Schooling Data

As an empirical application we use position and velocity data for a school of golden shiners  
299  
(*notemigonus crysoleucas*); details of how the data were collected can be found in the  
300  
Supplementary Information. These fish were studied in [24, 25]. In [25] the authors found  
301  
that the school, constrained to an essentially two dimensional environment, would generally  
302  
be found in one of three macro-states defined according to the school's collective  
303  
polarization,  $O_p$ , and rotation,  $O_r$ : milling (circling), swarming (disordered, stationary), or  
304  
polarized (translational). See Figure S1 for example frames of these behaviors. Writing  $v(x_i)$   
305

for the velocity of fish  $i$  and  $p(x_i)$  for its position, the chosen macrovariables are defined as 306

$$O_p := \frac{1}{N} \left| \sum_{i=1}^N \hat{v}(x_i) \right|; \quad O_r = \frac{1}{N} \left| \sum_{i=1}^N \hat{q}(x_i) \times \hat{v}(x_i) \right|;$$

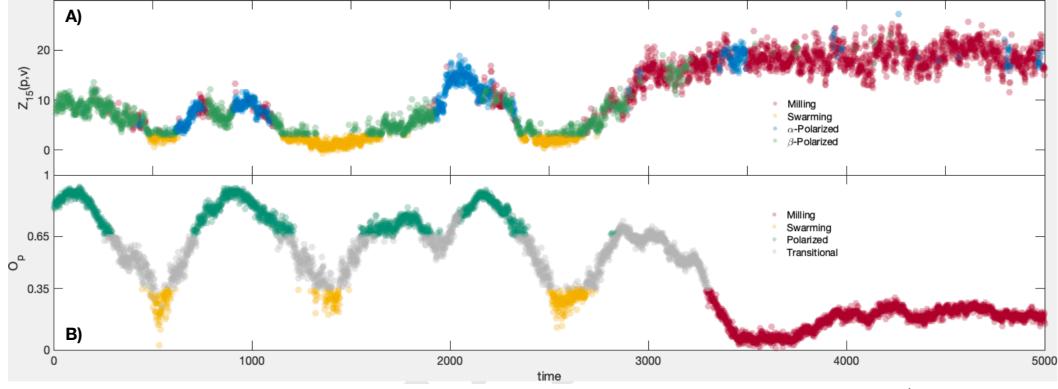
where  $q(x_i) := (p(x_i) - c)$  is the vector from the school's centroid,  $c = \frac{1}{N} \sum_{i=1}^N p(x_i)$ , to the 307  
position of fish  $i$ . Hats indicate unit vectors and  $\times$  indicates the cross product. One can 308  
then demarcate the behaviors as follows: 309

$$\begin{aligned} & \text{Milling} && \text{if } O_p \leq 0.35 \text{ and } O_r \geq 0.65, \\ & \text{Swarming} && \text{if } O_p \leq 0.35 \text{ and } O_r \leq 0.35, \\ & \text{Polarized} && \text{if } O_p \geq 0.65 \text{ and } O_r \leq 0.35, \end{aligned}$$

with all other  $(O_p, O_r)$  values considered to be *transitional* states. In the lower panel of 310  
Figure 5 the variable  $O_p$  is plotted for the 5000 frames used in this study; the coloration of 311  
the plot indicates the school's state as defined here (red for milling, yellow for swarming, 312  
dark green for polarized, and grey for transitional). The remainder of this section 313  
demonstrates that the unsupervised tools developed above are capable of discerning 314  
data-driven macrostates corresponding to, and even refining, those defined in [25]. 315

For the study of fish schooling, as with the simulated data above, we built a pair of 316  
distance matrices at each time step. The first was calculated from the euclidean distances 317  
between agents (individual fish),  $d^{(1)}$ , defined as before. The second,  $d^{(2)}$ , using the velocity 318  
data only. In this case, our goal was to capture any interaction between agents, either 319  
instantaneous or at a lag of  $\geq 1$  time step due to one fish leading another. Therefore we 320  
define  $d^{(2)}(x_i, x_j; t)$  as in the previous section (a weighted  $L^2$  norm of the differences in the 321  
agents' recent velocity profile), but now allow one of the two agent's velocity profiles to 322  
begin at an earlier time step,  $t - \ell$  with  $0 \leq \ell \leq 60$ . We take  $d^{(2)}(x_i, x_j; t)$  to be the 323  
minimum of this collection of distances, as this choice of time translation gives the best 324  
alignment of the two fishes' trajectories. 325

As with the flocking model, the first diffusion map's leading coordinates identify the 326  
large groups of fish who form spatially proximate clusters; the leading diffusion coordinates 327  
associated to  $d^{(2)}$  highlight groups whose velocity profiles are (up to a 2s, or 60 frame, 328  
translation in time) quite similar. The map alignment between these two systems,  $Z_{15}(p, v)$ , 329  
is plotted in the upper panel of Figure 5. When the fish are swarming (see Figure S1, center 330



**Fig 5.** Statistics calculated from 5000 frames of the golden shiner dataset. A): Map alignment  $Z_{15}(p, v)$  calculated over 5000 video frames. The plots are colored according to our classification of macro-scale state; milling = red, swarming = yellow,  $\alpha$ -polarized = blue,  $\beta$ -polarized = green. B): The group polarization macro-variable  $O_p$  calculated from the same 5000 frames. The coloring shows the states according to [24]; milling = red, swarming = yellow, polarized = teal, unclassified/transitional = grey.

column) we see that these two coordinate systems are unrelated, and that  $Z_{15}(p, v)$  is correspondingly low.

Here again, the choice of  $k = 15$  is driven by observing that outside of the top fifteen entries of  $\hat{u} = u/\|u\|$  for  $u = (\lambda_1^{(i)}, \dots, \lambda_n^{(i)})$ ,  $i \in \{1, 2\}$ , the spectral decay appears to follow a power law for most frames (though this observation does break down for some frames when the eigenvalues become small enough). The results below prove to be robust to varying  $k$  from 10 to 25. Therefore the span of  $\{\phi_1^{(1)}, \dots, \phi_{15}^{(1)}\}$  contains the majority of the structure within the data, while  $k = 15$  is not so large that the signal in  $Z_k$  becomes diluted.

The two coherent behaviors we would like to separate are the polarized, or linear, motion (right column in Figure S1) and milling, or circling, motion (left column in Figure S1). Both states exhibit high levels of organization, and we must use a sharper tool to distinguish them.

To do this we employ the covariance vectors  $\xi_{i,j}$ , as defined in the Unsupervised Macrostate Classification section above, over the course of 5000 frames of video (see the Supplementary Information (SI) for more information on the dataset). First, those states with  $Z_{15}(x, v) < 3$  are considered to exhibit incoherent behavior and to have no macro-scale dynamic, so we classify them as swarming frames and remove them from the set of frames to be classified. Then for each time  $t$  we compute the covariance vector  $\xi(t) = (|\xi_{\sigma(i,j)}(t)|)$ . We cluster the resulting vectors using the  $k$ -means algorithm, with  $k = 3$ . This leaves us with three groups representing distinct coherent behaviors, labelled G1, G2, and G3, as well

**Table 2.** Counts of classifications by our map alignment-based statistics (rows) and by the macro-scale variable classification (columns) for the 5000 frames of fish movement data. The percentage of the frames of a given state (milling, polarized, etc.) which is found in each of our unsupervised groups (G1, G2, G3, N) are displayed below the counts.

	Milling	Swarming	Polarized	Transitional
G1	1463 (85.9%)	5 (1.5%)	73 (5.1%)	348 (22.5%)
	0 (0%)	105 (32.0%)	738 (51.9 %)	454 (29.4%)
G2	241 (14.1%)	2 (0.6%)	505 (35.5%)	352 (22.8%)
	0 (0%)	216 (65.9%)	105 (7.4 %)	391 (25.3%)
N				

as the incoherent group, which we label  $N$ . These groups define the coloration of the top panel of Figure 5. Comparing these group memberships to those given by calculating the group polarization and rotation,  $O_p$  and  $O_r$ , and classifying behavior as in [24] we find that group G1 contains 85.86% of the milling frames, G2 and G3 together comprise 87.47% of the polarized frames, while  $N$  holds 65.85% of the swarming frames (another 32.01% are relegated to G2). The full results of the classification appear in Table 1.

That the polarized frames are divided into two separate clusters suggests that there are two distinct modes of polarized behavior exhibited by the fish. Upon inspecting  $\xi(t)$  for the frames from each cluster, we find that the two groups are essentially distinguished by their relative levels of organization; for example, the mean value of  $Z_{15}$  for cluster G2 was 6.3186, while for G3 the mean map alignment statistic is 12.0878. Practically, the frames of cluster G2 typically consist of polarized group motion with multiple subgroups of fish following distinct paths; we call this  $\beta$ -polarized behavior. This creates a macro-scale division in the network structure of  $\hat{A}^{(v)}$  that is not present in the network structure of  $\hat{A}^{(x)}$ , which leads to a lower alignment between their top eigenvectors; here  $\langle \xi_{1,1}(t) \rangle = 0.176$ . On the other hand, the frames of G1 feature a single, unified group motion, leading to a high alignment ( $\langle \xi_{1,1}(t) \rangle = 0.331$ ) between the top diffusion coordinates of each network. We refer to this as  $\alpha$ -polarized schooling behavior. It is possible that there is a correspondence between the empirically defined  $\beta$ -polarized activity and the *dynamic parallel* group behavior described in the family of models of [26].

One may approach the problem of classifying the covariance vectors  $\{\xi(t)\}$  in numerous ways; we chose the  $k$ -means algorithm for its simplicity. However, it requires the user to fix the number of desired clusters beforehand. As the prior work of [24] only defined two

separate coherent behaviors, milling and polarized, one may wonder what results we would 374  
find if we instead restricted the classification to two clusters. We did this and, labelling the 375  
groups G1\* and G2\*, found that G1\* successfully captures all the milling frames, and G2\* is 376  
predominantly composed of polarized frames (see Table S1). However, roughly 1 in 4 377  
polarized frames are relegated to G1\*. Table S2 shows how the two  $k$ -means clusterings 378  
apportion the frames. In particular, the  $\alpha$ -polarized frames tend to be classified as milling. 379  
By increasing the number of subgroups to  $k = 3$ , we allow the algorithm to perform a more 380  
sensitive clustering, which improved the correlation between our groupings and those of [24], 381  
as well as distinguished between two forms of polarized behavior that were grouped together 382  
in that work. 383

## Discussion 384

To identify different states of collective behavior in complex systems, we have developed and 385  
applied a new methodology based on measuring changes in the mutual geometry of a given 386  
system as it is viewed through different lenses, using diffusion maps. The method is general 387  
and objectively produces system-specific, macro-scale variables,  $Z_k$  and  $\xi$  for tracking the 388  
onset of and distinguishing between different regimes of collective behavior. To test the 389  
method, we applied it to synthetic data produced from a well-known model of birds flocking, 390  
and to empirical data on fish schooling. In both cases the macro-scale quantities constructed 391  
using diffusion maps provided an accurate description of the system in terms of the degree 392  
and type of collective behavior present. 393

The analysis, while data-driven, does suggest methods of controlling complex systems, or 394  
nudging them towards a desired state, as in [27]. For example, the top panel of Figure 5 395  
suggests that swarming behavior is typically bookended by a  $\beta$ -polarized state. Then to 396  
bring the school into a swarming state from an  $\alpha$ -polarized state, one might attempt to 397  
force the fish to break into multiple polarized subgroups (by instituting an obstruction, say). 398  
Achieving a misalignment of the leading diffusion coordinates from the spatial- and 399  
velocity-based networks of the school will clearly impede any group consensus, making the 400  
school more likely to lose its coherence. This strategy is surely obvious in the case of 401  
schooling fish, but an analogous tactic could be employed to inform control over state 402  
change in markets or social networks as well. 403

Furthermore, the diffusion coordinates provide a quantitative basis for choosing how to align/misalign a system; one can use *in silico* experiments to identify changes to the network structure which would move the system towards the desired macro-state. Given a target covariance vector, the technique could perhaps be automated. One approach would be to perform spectral clustering of the network according to the present and target eigenfunctions. Agents belonging to the same cluster in both cases could be expected to remain well-connected during the transformation, simplifying the search for an appropriate perturbation to those that only increase/decrease the connectivity of those clusters. Reducing the problem space in this way would be a first step towards an online solution, advising system control in real time.

Flocking models and empirical data on fish schooling were studied here because the collective behavior from each system is well known. That is, the scales at which collective behavior emerges and the macro-scale variables that do a good job of describing group behavior are known. This allowed for a clear demonstration of the new methodology, though one can apply this framework to a variety of dynamic or static data sets, from subfields outside animal group dynamics, and areas other than biology. There are many non-biological systems, such as financial and housing markets, city-transit systems, and power-grid systems, which are known to exhibit emergent patterns at a range of scales, but for which the relevant micro-scale variables are less well known. In these cases one may compute map alignment in an uninformed fashion, testing a variety of metrics and variables, to discover latent emergent relationships.

In practice, the decision of which micro-scale metric to use may not always be clear, but the map alignment method offers a resilience to these particulars. In the two cases studied, familiarity with animal collectives and their emergent properties made the choice to use position and velocity data snippets and their respective metrics a straightforward decision. However, there are many other biological systems where the organizational variables to choose may not be so obvious [28]. While the exact variables controlling information propagation in such systems may be unknown, as demonstrated in [29], the underlying structure may still be recovered through measurements of different, related variables. That is, we take advantage of the gauge-invariance of the diffusion map approach. As many previous publications have pointed out [24, 30, 31], fish do not interact with one another based purely on their euclidean distance; other factors including rank of nearness (closest

neighbors), relative alignment, and density. Therefore the metric  $d^{(1)}$  used above is a gross 436  
simplification of the role spatial organization plays. However, it captures enough of the true 437  
underlying organizational structure to allow us to detect emergent behavior. 438

It is worth noting, however, that the approach that we have developed has its own 439  
challenges relating to the nature of the data used. Diffusion maps inherently require a large 440  
number of data points for the diffusion operator to exhibit the regularity of the limiting 441  
operator on the manifold from which the data is, in theory, sampled. With the fish data 442  
though, the intermittent nature of the observations meant that the system size  $n$  varied 443  
between roughly 70 and 280 agents, and still the method proved resilient to these 444  
fluctuations. 445

Also, if one employs the GkNN kernel as we do here (see (1)) the scale-free property 446  
simplifies the task of resolving the macro-scale structure of the data; in particular, one does 447  
not need to adjust the kernel if the distances involved globally grow or shrink. But this 448  
entails a loss of information. As an example, if the system becomes highly organized when 449  
observed via the metric  $d$  (e.g.  $d$  corresponds to the velocity-based metric  $d^{(2)}$  above, and 450  
the fish all move as one) the corresponding data cloud  $(X, d)$  collapses towards a point mass. 451  
However, the GkNN kernel rescales distances in order to resolve their structure across scales, 452  
i.e. the rescaled distances only have meaning by virtue of their relative size. As a result the 453  
network's features can become a measurement of the lower-order stochastic fluctuations of 454  
individual agents' trajectories, rather than meaningful organization. This conflation of 455  
extremely organized and unorganized states could be overcome by tracking the mean 456  
inter-agent distance, for example, and may be important in application. 457

Measuring the map alignment between different networks induced by separate metrics on 458  
a graph serves as an entry point for several other possible analyses. One could use a set of 459  
training data to perform a classification of the various emergent behaviors of the system, 460  
then measure the stability of the system (risk of state change) by the current distance from 461  
 $\xi(t)$  to representative subgroups from the various states. For example, if the distance from 462  
 $\xi(t)$  to Group 1 is growing, and the distance to Group 2 is shrinking, one might anticipate a 463  
shift in macrostate. In this way one may estimate the risk of emergence, dissolution, or a 464  
large-scale shift of collective behavior. The authors believe, based on preliminary analysis of 465  
the fish schooling data, that such an approach could provide a new early warning metric for 466  
complex systems whose relevant macrovariables may be unknown, inhibiting the application 467

of other early warning signals such as critical slowing down.

468

## Conclusion

469

The method that we have developed is a new data-driven approach for detecting cross-scale emergent behavior in complex systems. It quantifies the dependence present between various micro-scale variables (those exhibited at the agent-level), and formulates a signature of the macro-scale behavior exhibited by the collective. The method can also detect the onset or loss of organization in an unsupervised fashion. With sensible choices of micro-scale variables, which are now measured routinely as Big Data, any complex system can be studied in this way. Doing so presents new opportunities for studying complex systems through analyzing their changing geometry.

470

471

472

473

474

475

476

477

## Supporting information

478

### Idealized Simulations of Bird Flocking

479

Let  $N$  agents be initiated at time 0 with position chosen uniformly at random on the torus  $\mathbb{T} = S^1 \times S^1$ , and with bearing chosen uniformly at random from  $S^1$ . Write

$B_i(r) = \{j : \|p_i - p_j\| < r\}$  for the set of agents within a radius  $r$  of agent  $b_i$ . The dynamics of our flocking model are written as follows:

$$\theta_i(t+1) = \langle \theta_j(t) \rangle_{i,r} + \eta \quad (\text{S1})$$

$$p_i(t+1) = p_i(t) + v(\cos(\theta_i(t)), \sin(\theta_i(t))) \quad (\text{S2})$$

where  $\langle \theta_j(t) \rangle_{i,r}$  is the mean bearing of the agents within  $B_i(r)$  and  $\eta$  is a Gaussian noise with standard deviation  $\epsilon > 0$ . (Note that this mean bearing is calculated by finding the local mean velocity, then setting  $\langle \theta_j(t) \rangle_{i,r}$  equal to its angle with the abscissa.) Thus each agent's position is updated by flying a distance  $v$  in the direction determined (up to the perturbation  $\eta$ ) by the average of the directions of all other agents within  $r$  units of the focal agent. Allowing  $r$  to vary over time, transitions between flocking and disordered behavior can be observed. Figure 1 in the main text shows typical system states with  $N = 800$  agents in the incoherent and coherent phases. Code for this simulation can be found below.

In the figures and analysis of the manuscript, the datasets are generated with  $N = 500$ , taking the noise to be  $\epsilon = \pi/5$ , velocity  $v = 1/320$ , with minimum and maximum values of the interaction radius given by  $r_0 = 0.0065$  and  $r_1 = 0.06$ .

### Empirical Data on Collective Behavior in Fish

491

The fish behavioral data was gathered by Tonstrøm et al. by recording video at 30 frames per second of a group of 300 golden shiners moving about a tank of shallow water. The water level was low enough to essentially limit the fish to planar motion. Image analysis was employed to locate the fish within each frame, then match them across frames when possible. This allowed the research team to associate a unique identifier with each fish while its trajectory was under observation.

These fish are highly social, and due to the difficulty of rendering individual fish when they are packed tightly into a region, and the possible conflation of two separate fish

swimming near one another, individual fish are often ‘dropped’ from the data set. When  
 500  
 this occurs the identifier associated with that fish is retired, and the position and velocity  
 501  
 data for that particular trajectory terminates. However, when the fish again exceeds the  
 502  
 liminal level of the image analysis it is registered as a ‘new’ fish and given a new identifier.  
 503  
 As a result, the data garnered from a typical frame records position and velocity information  
 504  
 for only 60 - 80% of the members, and individuals are typically tracked for about 200 frames  
 505  
 (6.7 seconds) before its image is ‘lost’ and its identifier removed from future data. The  
 506  
 resulting data set consists of 100298 frames (about 56 minutes) of position and velocity data  
 507  
 for the school. For the study in the manuscript we selected a 5000 frame subset containing  
 508  
 roughly an equal number of frames devoted to each of the behaviors observed (frames 40001  
 509  
 to 45000). This enabled us to use the simple  $k$ -means clustering algorithm, which otherwise  
 510  
 suffers when the subgroups to be identified have widely varying sizes. (This is because the  
 511  
 classical  $k$ -means algorithm creates hyperspherical clusters, where the hyperspheres must be  
 512  
 centered within the convex hull of the data and have equal radii. Smaller clusters  
 513  
 necessitate a description via smaller hyperspheres. See [32].)  
 514

As discussed in [25], human observation of the schooling behavior identified three  
 515  
 separate behavioral regimes for the fish: milling, swarming, and polarized (see Figure S1).  
 516  
 In the milling state, the fish rotate as a group, circulating about a fixed point; this is a  
 517  
 highly coherent group state. While swarming, the fish trajectories are disordered, with very  
 518  
 little discernible group alignment. In this state the school is stationary, though individual  
 519  
 fish may be moving in an uncoordinated fashion. Last, the second coherent group structure  
 520  
 is the polarized state, which is characterized by the group having a coherent direction of  
 521  
 motion, and strong local alignment among the agents. As with the flocking model above, the  
 522  
 variables of interest are the two-component vectors of fish position,  $p_i$ , and fish velocity,  $v_i$ .  
 523

## Metrics

524

Let  $x_i \in X$  be the set of observations for an agent (either a particle in the flocking  
 simulation or a golden shiner found in the school) at a fixed time  $t$ , while  $p_i(t)$ ,  $\theta_i(t)$ , and  
 $v_i(t)$  are the  $i^{\text{th}}$  agent’s position, heading, and velocity, respectively. The first metric,  $d^{(1)}$ ,

is simply the euclidean distance between the agents,

$$d^{(1)}(x_i, x_j; t) := \|p_i(t) - p_j(t)\|_2. \quad (\text{S3})$$

For the flocking simulation, the second metric is the following

$$d^{(2)}(x_i, x_j; t) := \left( \sum_{k=0}^{30} g(k) \|v_i(t-k) - v_j(t-k)\|_2^2 \right)^{\frac{1}{2}} \quad (\text{S4})$$

where  $v_i(t) = v(\cos(\theta_i(t)), \sin(\theta_i(t)))$  is the  $i^{\text{th}}$  velocity at time  $t$ . The smoothing term  $g(k)$  is proportional to  $\exp(-k^2/30)$ , but normalized so that  $\sum_k g(k) = 1$ . That is,  $d^{(2)}$  is a gaussian smoothed time-average of the  $L^2$  norm between velocities over the past 30 time steps.

For the fish, we consider a generalization of the previous expression for  $d^{(2)}$ . Rather than comparing the two fishes' heading time series at simultaneous moments, we allow one time series to be shifted into the past by a lag of up to  $L = 60$  video frames, or two seconds.

Define

$$\langle x_i, x_j \rangle_{L,t} := \min_{0 \leq l \leq L} \left( \sum_{k=0}^{30} g(k) \|v_i(t-k) - v_j(t-k-l)\|_2^2 \right)^{\frac{1}{2}},$$

where  $g(k)$  is again a gaussian smoothing kernel as in the previous paragraph. Thus  $l$  represents a lag allowing us to compare not only concurrent velocity profiles over 30 steps, but also cases where fish  $i$  is mimicking a trajectory of fish  $j$  up to two seconds in the past. Note that  $\langle x_i, x_j \rangle_{L,t} \neq \langle x_j, x_i \rangle_{L,t}$ , as the righthand term calculates the distance between the trajectories under the assumption that  $j$  is following  $i$ . We define our distance as

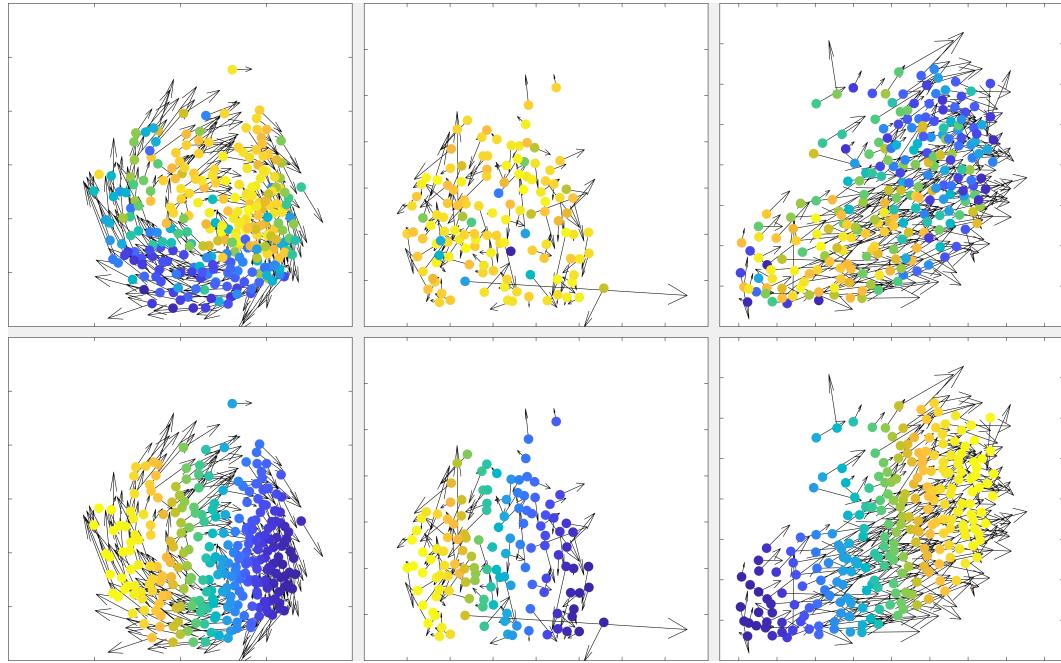
$$d_L(x_i, x_j; t) := \min\{\langle x_i, x_j \rangle_{L,t}, \langle x_j, x_i \rangle_{L,t}\}. \quad (\text{S5})$$

**Table S1.** Counts of classifications by our map alignment-based statistics (rows) and by the macro-scale variable classification (columns) for 5000 frames of fish movement data. The percentage of the frames of a given state (milling, polarized, etc.) which is found in each of our unsupervised groups ( $G1^*$ ,  $G2^*$ , N) are displayed below the counts.

	Milling	Swarming	Polarized	Transitional
$G1^*$	1704 (100.0%)	6 (1.8%)	375 (26.4%)	606 (39.2%)
$G2^*$	0 (0%)	106 (32.3%)	941 (66.2 %)	548 (35.5 %)
N	0 (0%)	216 (65.9%)	105 (7.4 %)	391 (25.3%)

**Table S2.** Comparison of the two  $k$ -means classifications for 5000 frames of fish movement data. The 2-group clustering gave G1\* and G2\*, corresponding to frames of high and low macro-scale organization, respectively. The 3-group clustering gave G1, G2, and G3, corresponding to milling,  $\alpha$ -polarized, and  $\beta$ -polarized behaviors, respectively. The percentages are with respect to all the frames in the entry's column.

	G1	G2	G3	N
G1*	1886 (99.95%)	803 (72.87%)	2 (0.15%)	0 (0.0%)
	1 (0.05%)	299 (27.13%)	1295 (99.85 %)	0 (0.0 %)
N	0 (0.0%)	0 (0.0%)	0 (0.0 %)	712 (100.0%)



**Fig S1.** Plots of fish position (colored markers) and velocity (arrows) at three example times, colored by diffusion map coordinates  $\phi_1^{(v)}$  [top row; i.e. produced from analyzing correlations in velocities] and  $\phi_1^{(x)}$  [bottom row; i.e. produced from analyzing spatial proximity]. The emergent behaviors from left to right are milling, swarming, and polarized motion.

## Supplementary Movies

533

**Movie 1:** A sample run of the flocking model ( $\epsilon = \pi/5$ ,  $v = 1/320$ ) where the interaction radius  $r(t)$  is equal to  $r_0 = 0.0065$  until time  $t = 800$ , is equal to  $r_1 = 0.06$  for times after  $t = 1200$ , and increases linearly from  $r_0$  to  $r_1$  over time steps 800 to 1200. Each subplot is colored according to one of the top two diffusion coordinates, as labeled. The top two subplots display the first and second coordinates from the  $d^{(1)}$ -derived affinity matrix (see equation (S3)), while the bottom subplots are colored by the diffusion coordinates from the  $d^{(2)}$ -based affinity matrix (see equation (S4)).

**Movie 2:** A sample run of the flocking model ( $\epsilon = \pi/5$ ,  $v = 1/320$ ) where the interaction radius  $r(t)$  varies sinusoidally between  $r_0 = 0.0065$  and  $r_1 = 0.06$ , with the period of oscillation increasing linearly for the duration of the time series:  $r(t) = 0.0$ . Each subplot is colored according to one of the top two diffusion coordinates, as labeled. The top two subplots display the first and second coordinates from the  $d^{(1)}$ -derived affinity matrix (see equation (S3)), while the bottom subplots are colored by the diffusion coordinates from the  $d^{(2)}$ -based affinity matrix (see equation (S4)).

**Movie 3:** This video displays a subset of the fish schooling data used, colored by various diffusion coordinates. We restrict the frames displayed to keep the file size reasonable while still showcasing the various modes of behavior described in the main text. The time step displayed in each frame corresponds with the time steps shown on the abscissa of Figure 5. The top two subplots display the first and second coordinates from the  $d^{(1)}$ -derived affinity matrix (see equation (S3)), while the bottom subplots are colored by the diffusion coordinates from the  $d^{(2)}$ -based affinity matrix (see equation (S5)).

## Acknowledgments

555

The authors would like to acknowledge support from the DARPA YFA project N66001-17-1-4038 and Kolbjørn Tunstrøm for providing the fish schooling data

556

557

## References

558

1. Krause J, Ruxton GD, Ruxton GD. Living in groups. Oxford University Press; 2002.

559

2. Butail S, Boltt EM, Porfiri M. Analysis and classification of collective behavior using generative modeling and nonlinear manifold learning. *Journal of theoretical biology*. 2013;336:185–199. 560  
561  
562
3. Deutsch A, Theraulaz G, Vicsek T. Collective motion in biological systems. *Interface Focus*. 2012;2(6):689–692. doi:10.1098/rsfs.2012.0048. 563  
564
4. Couzin ID, Krause J, Franks NR, Levin SA. Effective leadership and decision-making in animal groups on the move. *Nature*. 2005;433(7025):513. 565  
566
5. Cont R, Bouchaud JP. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic dynamics*. 2000;4(2):170–196. 567  
568
6. Watts DJ. The “New” Science of Networks. *Annual Review of Sociology*. 2004;30(1):243–270. doi:10.1146/annurev.soc.30.020404.104342. 569  
570
7. Galton F. Vox Populi. *Nature*. 1907;75:450. 571
8. May RM, Levin SA, Sugihara G. Complex systems: Ecology for bankers. *Nature*. 2008;451(7181):893. 572  
573
9. Levin S, Xepapadeas T, Crépin AS, Norberg J, De Zeeuw A, Folke C, et al. Social-ecological systems as complex adaptive systems: modeling and policy implications. *Environment and Development Economics*. 2013;18(2):111–132. 574  
575  
576
10. Liu J, Dietz T, Carpenter SR, Alberti M, Folke C, Moran E, et al. Complexity of coupled human and natural systems. *science*. 2007;317(5844):1513–1516. 577  
578
11. Levin S. Complex adaptive systems: exploring the known, the unknown and the unknowable. *Bulletin of the American Mathematical Society*. 2003;40(1):3–19. 579  
580
12. Friedl P, Locker J, Sahai E, Segall JE. Classifying collective cancer cell invasion. *Nature Cell Biology*. 2012;14. 581  
582
13. Mataric MJ. From Local Interactions to Collective Intelligence. In: *The Biology and Technology of Intelligent Autonomous Agents*. vol. 144 of Series F: Computer and Systems Sciences. Springer; 1995. 583  
584  
585

14. Mehta P, Bukov M, Wang CH, Day AGR, Richardson C, Fisher CK, et al. A high-bias, low-variance introduction to Machine Learning for physicists. arXiv e-prints. 2018; p. arXiv:1803.08823. 586  
587  
588
15. Wetzel SJ. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Phys Rev E*. 2017;96:022140. 589  
590  
doi:10.1103/PhysRevE.96.022140. 591
16. Wang L. Discovering phase transitions with unsupervised learning. *Phys Rev B*. 2016;94:195105. doi:10.1103/PhysRevB.94.195105. 592  
593
17. Rodriguez-Nieva JF, Scheurer MS. Identifying topological order via unsupervised machine learning. arXiv e-prints. 2018; p. arXiv:1805.05961. 594  
595
18. Talmon R, Cohen I, Gannot S, Coifman RR. Diffusion maps for signal processing: A deeper look at manifold-learning techniques based on kernels and graphs. *IEEE signal processing magazine*. 2013;30(4):75–86. 596  
597  
598
19. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(21):7426–7431. 599  
600  
601  
602
20. Coifman RR, Lafon S. Diffusion maps. *Applied and computational harmonic analysis*. 2006;21(1):5–30. 603  
604
21. Fefferman C, Mitter S, Narayanan H. Testing the manifold hypothesis. *Journal of the American Mathematical Society*. 2016;29(4):983–1049. 605  
606
22. Berry T, Sauer T. Consistent Manifold Representation for Topological Data Analysis. In: Some Book...; 2016. 607  
608
23. Toner J, Tu Y. Flocks, herds, and schools: A quantitative theory of flocking. *Physical review E*. 1998;58(4):4828. 609  
610
24. Katz Y, Tunstrøm K, Ioannou CC, Huepe C, Couzin ID. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*. 2011;108(46):18720–18725. 611  
612  
613

25. Tunstrøm K, Katz Y, Ioannou CC, Huepe C, Lutz MJ, Couzin ID. Collective States, 614  
Multistability and Transitional Behavior in Schooling Fish. PLOS Computational 615  
Biology. 2013;9(2):1–11. doi:10.1371/journal.pcbi.1002915. 616
26. Couzin ID, Krause J, James R, Ruxton GD, Franks NR. Collective memory and 617  
spatial sorting in animal groups. Journal of theoretical biology. 2002;218(1):1–11. 618
27. Janssen MA. Targeting individuals to catalyze collective behavior in social networks. 619  
In: CSSSA 2011 Papers; 2011. Available from: 620  
<http://computationalsocialscience.org/wp-content/uploads/2011/10/> 621  
Janssen-CSSSA2011.pdf. 622
28. Attanasi A, et al. Collective Behaviour without Collective Order in Wild Swarms of 623  
Midges. PLoS Comput Biol. 2014;10(7). 624
29. Kemeth FP, Haugland SW, Dietrich F, Bertalan T, Li Q, Boltt EM, et al. An Equal 625  
Space for Complex Data with Unknown Internal Order : Observability , Gauge 626  
Invariance and Manifold Learning; 2017. 627
30. Huth A, Wissel C. The simulation of the movement of fish schools. Journal of 628  
Theoretical Biology. 1992;156(3):365 – 385. 629  
doi:[https://doi.org/10.1016/S0022-5193\(05\)80681-2](https://doi.org/10.1016/S0022-5193(05)80681-2). 630
31. Cambuí DS, Rosas A. Density induced transition in a school of fish. Physica A: 631  
Statistical Mechanics and its Applications. 2012;391(15):3908 – 3914. 632  
doi:<https://doi.org/10.1016/j.physa.2012.03.009>. 633
32. Jain AK, Murty MN, Flynn PJ. Data Clustering: A Review. ACM Comput Surv. 634  
1999;31(3):264–323. doi:10.1145/331499.331504. 635