

Unsupervised Manifold Learning of Collective Behavior

Mathew Titus^{1,a,b}, George Hagstrom^c, Zach Gelbaum^{a,b}, and James R. Watson^a

^aCollege of Earth, Ocean, and Atmospheric Sciences, 104 CEOAS Administration Building, Oregon State University, Corvallis, Oregon, 97331 USA; ^bThe Prediction Lab LLC, Corvallis, Oregon, 97331 USA; ^cDept. Ecology and Evolutionary Biology, 104a Guyot Hall, Princeton University, Princeton, New Jersey, 08544 USA

This manuscript was compiled on January 7, 2019

1 Collective behavior is an emergent property of numerous complex systems, from financial markets to cancer cells to predator-prey ecological
2 systems. Characterizing modes of collective behavior is often done through human observation (1, 2), training classifiers or generative
3 models (3), or other supervised learning techniques. In each of these cases one needs to have knowledge of and a method for characterizing
4 the macro-state(s) of the system. This presents a challenge for studying novel systems where there can be little to no *a priori* knowledge or
5 understanding. Here, we present a new unsupervised method of detecting emergent behavior in complex systems, and discerning between
6 distinct collective behaviors. We require only two agent-level metrics $d^{(1)}$ and $d^{(2)}$ defined on the set of agents, X , measuring interactions
7 or correlations in variables of interest. We apply the method of diffusion maps (4) to the systems ($X, d^{(i)}$) to recover efficient embeddings
8 of their interaction networks. Comparing these geometries, we formulate a measure of similarity between the two networks, called the
9 map alignment statistic (MAS). A large MAS is evidence that the two networks are codetermined in some fashion, indicating an emergent
10 relationship between the metrics $d^{(1)}$ and $d^{(2)}$. Additionally, the form of the macro-scale organization is partially encoded in the covariances
11 among the two sets of diffusion map components. Using these covariances one can discern between different modes of collective behavior
12 in a data-driven, unsupervised manner. This method is demonstrated on empirical movement data from schools of fish. We show that our
13 state classification subdivides the known behaviors of the school in a meaningful manner, leading to a finer description of the system's
14 behavior.

Complex Adaptive Systems | Micro, Macro, Multi-scale | Collective Behavior | Manifold Learning | State Identification | Unsupervised Learning

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

11 A key challenge in the study of complex systems is to identify when collective behavior emerges,
12 and in what way it manifests (11). 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. In this case, it is a non-trivial task to identify useful variables (at particular scales of organization) which define the possible emergent behaviors.

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 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 (?). 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 (?)

Significance Statement

Many complex systems in society and nature exhibit forms of collective behavior, where groups of individuals work together to solve the problems they are faced with. One challenge we face as scientists is to be able to identify and characterize different forms of collective behavior, and here we have developed a new method for analyzing data from individuals, to know when a given complex system is exhibiting system-wide organization. Importantly, our approach does not require 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.

All authors designed research and wrote the paper, G.H. created the flocking simulation, M.T. performed research and data analysis.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. E-mail: mtitus@ceoas.oregonstate.edu

33 ? ?). However, more complex types of collective phenomenon, such as topological phases, have
34 not been successfully identified using algorithms like PCA, which assume that the data has a linear
35 structure. In these more difficult cases, algorithms which can detect nonlinear structures such as
36 diffusion maps have shown much greater promise. For example, diffusion maps have been used to
37 find the Kosterlitz-Thouless phase transition in the XY-model from condensed matter physics(?).
38 The success of machine learning algorithms for physical problems suggests that machine learning
39 could be of great practical use for the study of collective phenomena in social and ecological systems,
40 which are usually much less amenable to theoretical analysis, and for which there often exists a
41 significant quantity of data at the individual level.

42 Here we have developed an approach to analyzing collective behavior in biological systems based
43 on dimensional reduction tools from manifold learning (12), specifically diffusion maps (4). Using
44 diffusion maps to formulate multiple data-driven coordinate systems, we compare their large-scale
45 structures to find latent global relationships between different variables observed at the level of
46 individual agents. We call this the map alignment between the two coordinate systems. When two
47 variables serve to globally organize the data along similar lines, their map alignment is high. One can
48 easily calculate the expected alignment between unrelated and independent variables, so this gives us
49 a test that detects a system's emergent behaviors but requires no prior knowledge of them. Moreover,
50 the ways in which the coordinates are correlated leave a signature for the relationship between
51 the macro-scale structures, and one may use this to distinguish various system-wide behaviors.
52 We apply this tool to modeled data produced from simulations of birds flocking, and to empirical
53 data on the activity of fish in a school. Using the micro-scale data to cast the agents' network
54 of interaction into multiple geometries, each evolving over time, we both detect the emergence
55 of macro-scale organization as well as perform a meaningful classification of distinct modes of
56 organization, identifying the macro-scale states of the system. Importantly, this analysis is based
57 purely on changes occurring at micro-scales, and makes no assumptions about which macro-scale
58 variables are important. This means that for any novel system of study, where there is a paucity of
59 knowledge regarding which macro-scale variable to measure and track, one could use this approach
60 to classify and observe the dynamics of the system at different levels of organization.

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

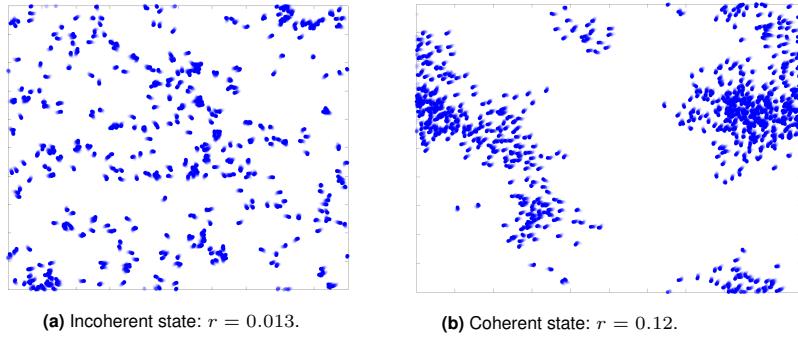


Fig. 1. Examples of an 800 agent flocking model on the torus, with and without long-range order, that is, in both incoherent and coherent macro-states.

64 Essentially we employ the algorithm in (13), with $\alpha = 1$ and $\epsilon = 1$. This involves

65 (i) constructing a similarity measure, $K(x_i, x_j)$, between pairs of data points,

66 (ii) normalizing K to create a Laplacian operator \hat{A} on the data,

67 (iii) computing eigenvector-eigenvalue pairs $\{(\phi_k, \lambda_k)\}$ for the Laplacian, and

68 (iv) constructing the diffusion map $\Phi(t, x) := (\lambda_k^t \phi_k(x))_{k=1}^n$.

69 The choice of similarity measure in (i) is highly system-dependent, for agent-wise interactions may
70 manifest themselves in many different ways and so require different metrics. In (ii), the choice of
71 normalization in constructing \hat{A} is subtler, and based on a choice of Laplace operator; this is not
72 sensitive to the data under observation, but rather reflects the assumption of Manifold Learning,
73 that the data points of X are drawn from a manifold embedded in euclidean space. For rigorous
74 results on when such a manifold exists for a given dataset, see (14).

75 Given n points $X = \{x_1, \dots, x_n\}$, we construct an $n \times n$ matrix of pairwise distances, $D =$
76 $(D_{i,j}) := (d(x_i, x_j))$, associated to a metric d . Writing $k(x)$ for the k^{th} nearest neighbor (kNN) of
77 $x \in X$, let $K(x, y)$ be the gaussian kNN (GkNN) kernel defined by

$$78 \quad K(x, y) = \exp\left(-\frac{d(x, y)^2}{d(x, k(x))d(y, k(y))}\right). \quad [1]$$

79 This kernel could be replaced with any other, as another function may better reflect the user's
80 impression of the similarity between two agents a distance $d(x, y)$ apart. This particular choice is
81 studied in (15), and serves to allow pairwise distances to scale according to their local geometries.
82 One could consider this the standard gaussian kernel applied to the modified distance function

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 $X \times X$
D	pairwise distance matrix
A	affinity matrix
\hat{A}	heat operator
$\mathcal{D}_t(\cdot, \cdot)$	diffusion distance on $X \times X$
$\{\phi_k\}$	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.

83 $d(x, y)/\sqrt{d(x, k(x))d(y, k(y))}$. The more sparse the point cloud (X, d) is near x or y , the smaller this
 84 distance becomes. This kernel also has the desirable property that it is scale-free, in that multiplying
 85 all distances by a fixed constant does not change the kernel K .

86 Let σ denote a row sum operator: if A is an $m \times n$ matrix, let $\sigma(A)$ be the m -vector whose i^{th}
 87 component is $\sigma(A)_i := \sum_{j=1}^n A_{i,j}$. We next calculate an affinity matrix $A = (A_{i,j}) := (K(x_i, x_j))$, its
 88 row sums $\sigma(A)$, and construct the matrix operator A' via

$$89 \quad A'_{i,j} = (\sigma(A)_i^{-1} A_{i,j} \sigma(A)_j^{-1}).$$

90 Normalizing by the row sums in this fashion is a particular choice of Laplace operator on the
 91 graph. This operator converges pointwise to the Laplace-Beltrami operator in the case that the
 92 points X are drawn from a probability distribution on a manifold (see (13)). We then define the
 93 Markov operator $\hat{A} = (\hat{A}_{i,j})$ by setting the row sums of A' to one:

$$94 \quad \hat{A}_{i,j} := \sigma(A')_i^{-1} A'_{i,j}. \quad [2]$$

95 This is the operator of primary interest to us, representing the evolution of a random walk on
 96 the graph, where the probability of transitioning from point x_i to x_j is given by $\hat{A}_{i,j}$. From the
 97 spectral structure of \hat{A} , written $\{(\phi_k, \lambda_k)\}_{k=1}^n$ with $\phi_k : X \rightarrow \mathbb{R}$ the unit eigenvectors and $\lambda_k \in \mathbb{R}$
 98 the eigenvalues, we construct the diffusion map,

$$99 \quad \Phi(t, x) = (\lambda_1^t \phi_1(x), \dots, \lambda_n^t \phi_n(x)), \quad t \geq 0. \quad [3]$$

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

$$103 \quad \mathcal{D}_t(x, y)^2 := \sum_{z \in X} (\hat{A}_{x,z}^t - \hat{A}_{y,z}^t)^2 \\ 104 \quad = \|\Phi(t, x) - \Phi(t, y)\|_2^2.$$

105 In essence, \mathcal{D}_t captures the geometric structure of the data by measuring the (dis)similarities of
 106 trajectories of random walks evolving on the data points from different starting points. The diffusion
 107 map is an isometry of (X, \mathcal{D}_t) into \mathbb{R}^n , with the remarkable benefit that truncating $\Phi(t, \cdot)$ to its
 108 first k coordinates gives the optimal k -dimensional embedding of X into \mathbb{R}^k . These truncated
 109 maps $\Phi_k(t, \cdot) := (\lambda_1^t \phi_1, \dots, \lambda_k^t \phi_k)$ are utilized below for detecting macro-scale relationships among
 110 variables.

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

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

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

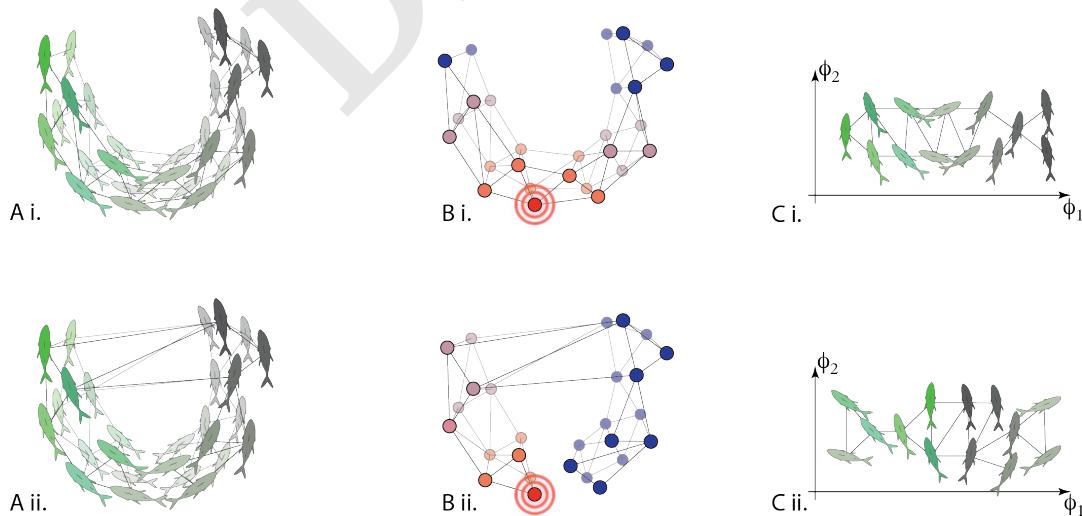


Fig. 2. Conceptual illustration of diffusion maps applied to data on fish schooling. A) For the same time period, different affinity operators can be calculated. These affinities could be correlations in simply spatial proximity (Ai) or correlations in velocities (Aii) for example. The operator is used 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. 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,

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

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

$$114 \quad d_i(x_j, x_k) := \|f_i(x_j) - f_i(x_k)\|_2. \quad [4]$$

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

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

$$125 \quad \Pi_k^{(2)}(\phi) = \sum_{j=1}^k \langle \phi, \phi_j^{(2)} \rangle \phi_j^{(2)}. \quad [5]$$

126 One attempts to choose k in a manner that ensures that the subspace spanned by $\{\phi_1^{(2)}, \dots, \phi_k^{(2)}\}$
 127 encompasses the relevant macro-scale coordinates induced by f_2 on X , but without k becoming so
 128 large that we begin to recover a significant portion of ϕ simply due to the large dimension of the
 129 image space. For example, if we take $k = n$ the projection $\Pi_n^{(2)}$ is the identity, simply expressing ϕ
 130 in the basis $\{\phi_j^{(2)}\}$. However, taking $k \ll n$ we recover only the portion of the vector which can be
 131 expressed as a linear combination of the top k diffusion coordinates of $\Phi^{(2)}$. We return to this issue

132 below.

133 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
134 the large-scale network structure induced on X by the variable f_2 .

135 Now we derive the expected projection size under a null hypothesis that the two variables f_1
136 and f_2 are entirely unrelated. Given the pair of diffusion bases, we fix $\phi \in \{\phi_j^{(1)}\} \subset \mathbb{R}^n$ and
137 calculate the distribution of the random variable $\|\Pi_k^{(2)}\phi\|$ assuming that the k -dimensional range,
138 $\text{Span}(\phi_1^{(2)}, \dots, \phi_k^{(2)})$, is chosen uniformly at random from all possibilities. One can show, and with
139 a little thought it is clear, that this is equivalent to asking for the magnitude of a random unit
140 vector of \mathbb{R}^n projected onto its leading k coordinates. When the vector is sampled uniformly at
141 random from $\{\psi \in \mathbb{R}^n : \|\psi\|_2 = 1\}$, the quantity $\|\Pi_k^{(2)}\phi\|^2$ is a beta distributed random variable with
142 parameters $(n - k)/2$ and $k/2$. Recognizing this, one can calculate the expectation and variance
143 to be k/n and $2(n - k)k/(n^3 + 2n^2)$, respectively. Normalizing the projection size accordingly, we
144 arrive at the statistic

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

145 which quantifies how well we can express the information of ϕ using only the subspace generated by
146 $\{\phi_1^{(2)}, \dots, \phi_k^{(2)}\}$.

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

155 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
156 the entries of \hat{u} give a normalized notion of the energy of each $\phi_j^{(1)}$, and so assigns a notion of how
157 important each mode $\phi_j^{(1)}$ is to the macrostructure derived from f_1 . Finally, we define the **map**
158 **alignment statistic** to be the sum of the $Z_k^{(2)}(\phi_j^{(1)})$, weighted by the contribution of $\phi_j^{(1)}$ to the

159 large-scale organization of f_1 :

160

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

161 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 $Z_k(f_1, f_2)$
162 is a measure of how much of the macrostructure of f_1 is reflected in the f_2 -based diffusion coordinates.
163 For example, the spectrum of the $\hat{A}^{(i)}$ may indicate that the top three diffusion coordinates of $\Phi_t^{(1)}$
164 carry a greater proportion of the macro-scale information of $\hat{A}^{(1)}$ than the top three coordinates of
165 $\Phi_t^{(2)}$ do for $\hat{A}^{(2)}$; in this case one would expect that $Z_3(f_1, f_2) > Z_3(f_2, f_1)$.

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

174 Alternatively, in using empirical data we found the bulk of the spectrum often follows an
175 approximate power-law decay. As a result, one may take k to be the smallest integer such that for
176 $j > k$, $\lambda_j \approx C j^{-\beta}$ for positive constants C, β .

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

183 **Unsupervised Macrostate Classification.** If the system under study exhibits multiple emergent
184 behaviors relating the two variables f_1 and f_2 , then coherence alone is not enough to identify system
185 behavior. In passing to the norm of the vector $\Pi_k^{(2)} \phi_j^{(1)}$, we clearly give up a significant amount of
186 information. In order to distinguish multiple coherent states, we inspect the individual dot products

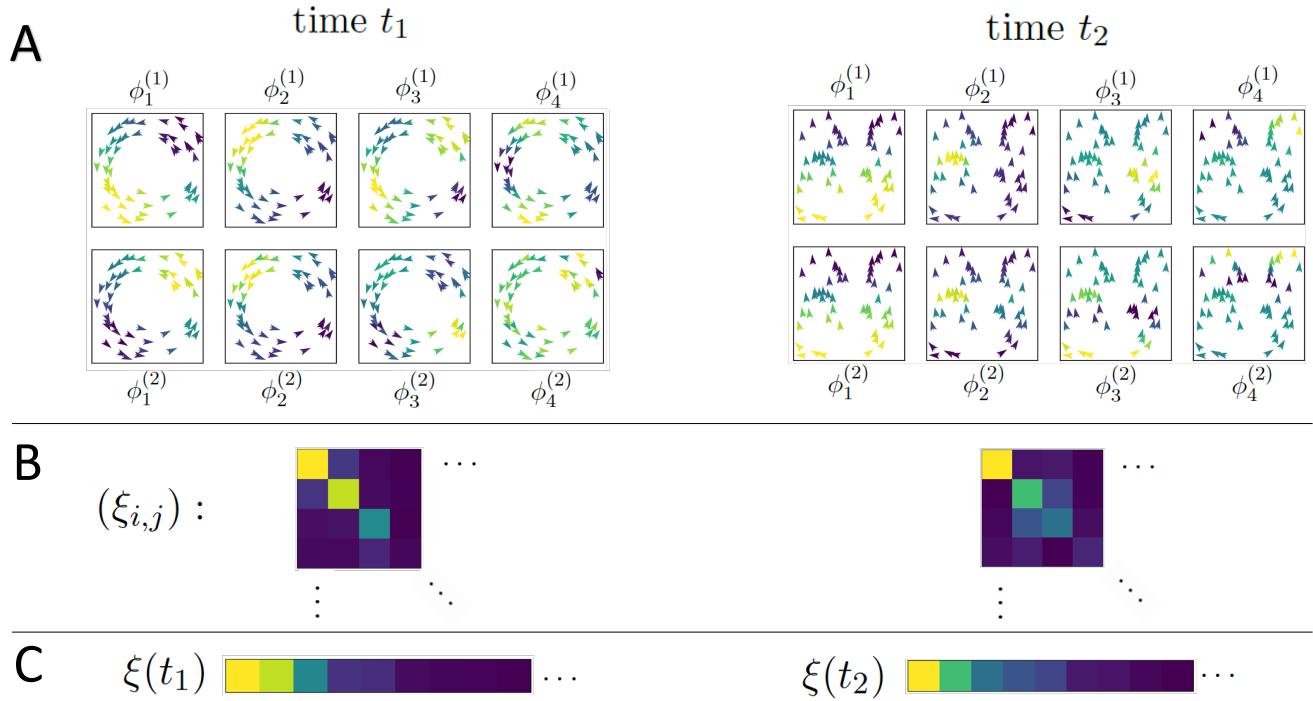


Fig. 3. A toy system of 50 agents; the left and right sides of the figure represent observations of the system at two different times when the system is in two different organized states. (A) Agents in each subpanel are colored by the indicated diffusion coordinate, (B) A color plot of the matrix $(\xi_{i,j})$ 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.

187 $\{\langle \phi_i^{(1)}, \phi_j^{(2)} \rangle\}$ that comprise the map alignment statistic.

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

$$190 \quad \xi_{i,j} := \langle \psi_i^{(1)}, \psi_j^{(2)} \rangle \\ 191 \quad = \lambda_i^{(1)} \lambda_j^{(2)} \langle \phi_i^{(1)}, \phi_j^{(2)} \rangle$$

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

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^{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^{th} eigenvectors ϕ_i and η_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 ξ (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.

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

Analysis of Bird Flocking Simulations. As a first test of our method we explored a well studied agent-based model of flocking behavior (16). 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 θ_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 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^{th} and j^{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 (SI3), (SI4) and (SI5) 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 Figures 4a and 4b 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 first set, the coupling parameter r is held at a low value

253 ($r = r_0 = 0.0065$) for 800 time steps before increasing linearly to a high value over 400 steps and
 254 remaining at that value ($r = r_1 = 0.06$) for the rest of the trial; see the lower panel of Figure 4a. The
 255 mean velocity $\langle v_j \rangle$ of the flock, averaged over the 100 trials, is plotted in orange in the upper panel.

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

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

266 Clearly $Z_{10}(p, v)$ indicates strongly the presence or absence of correlation. The map alignment
 267 statistic shows both a faster response to changing r and suffers from less attenuation as the speed of
 268 parameter change increases. Times $t = 800$ to 1000 especially highlight how the drop in Z_{10} is a
 269 precursor to the value of the macrovariable $\langle v_j \rangle$ falling. In this case and the previous case, we find
 270 we are able to track the transition from unorganized motion to flocking by the alignment between
 271 the velocity diffusion coordinates, $\Phi_{10}^{(1)}$, and the position diffusion coordinates, $\Phi_{10}^{(2)}$.

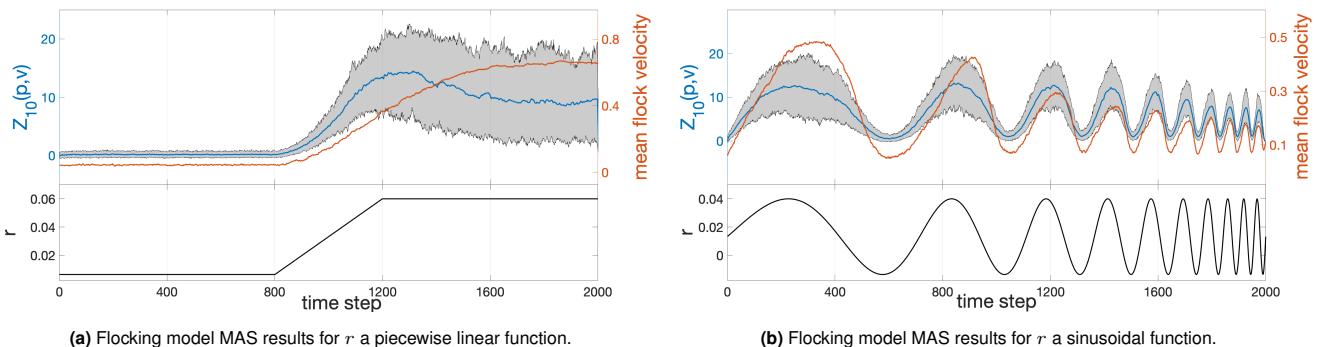


Fig. 4. Top: Map alignment statistic $Z_{10}(p, v)$ averaged over 100 flocking simulations of 2000 time steps (blue). The grey region denotes the 90% empirical confidence interval of $Z_{10}(p, v)$. The mean velocity $\langle v_j \rangle$ of the flock is averaged over the 100 trials and plotted in orange. 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.

272 **Analysis of Empirical Fish Schooling Data.** As an empirical application we use position and velocity
 273 data for a school of golden shiners (*notemigonus crysoleucas*); details of how the data were collected
 274 can be found in the Supplementary Information. These fish were studied in (17, 18). In (18)

275 the authors found that the school, constrained to an essentially two dimensional environment,
 276 would generally be found in one of three macro-states defined according to the school's collective
 277 polarization, O_p , and rotation, O_r : milling (circling), swarming (disordered, stationary), or polarized
 278 (translational). See Figure SI1 for example frames of these behaviors. Writing $v(x_i)$ for the velocity
 279 of fish i and $p(x_i)$ for its position, the chosen macrovariables are defined as

$$280 \quad O_p := \frac{1}{N} \left| \sum_{i=1}^N \hat{v}(x_i) \right|; \quad O_r = \frac{1}{N} \left| \sum_{i=1}^N \hat{q}(x_i) \times \hat{v}(x_i) \right|;$$

281 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 position
 282 of fish i . Hats indicate unit vectors and \times indicates the cross product. One can then demarcate the
 283 behaviors as follows:

$$\begin{aligned} & \text{Milling} && \text{if } O_p \leq 0.35 \text{ and } O_r \geq 0.65, \\ 284 \quad & \text{Swarming} && \text{if } O_p \leq 0.35 \text{ and } O_r \leq 0.35, \\ & \text{Polarized} && \text{if } O_p \geq 0.65 \text{ and } O_r \leq 0.35, \end{aligned}$$

285 with all other (O_p, O_r) values considered to be *transitional* states. The remainder of this section
 286 demonstrates that the unsupervised tools developed above are capable of discerning data-driven
 287 macrostates corresponding to and even refining those defined in (18).

288 For the study of fish schooling, as with the simulated data above, we built a pair of distance
 289 matrices at each time step. The first was calculated from the euclidean distances between agents
 290 (individual fish), $d^{(1)}$, defined as before. The second, $d^{(2)}$, using the velocity data only. In this case,
 291 our goal was to capture any interaction between agents, either instantaneous or at a lag of ≥ 1 time
 292 step due to one fish leading another. Therefore we define $d^{(2)}(x_i, x_j; t)$ as in the previous section
 293 (a weighted L^2 norm of the differences in the agents' recent velocity profile), but now allow one of
 294 the two agent's velocity profiles to begin at an earlier time step, $t - \ell$ with $0 \leq \ell \leq 60$. We take
 295 $d^{(2)}(x_i, x_j; t)$ to be the minimum of this collection of distances, as this choice of time translation
 296 gives the best alignment of the two fishes' trajectories.

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

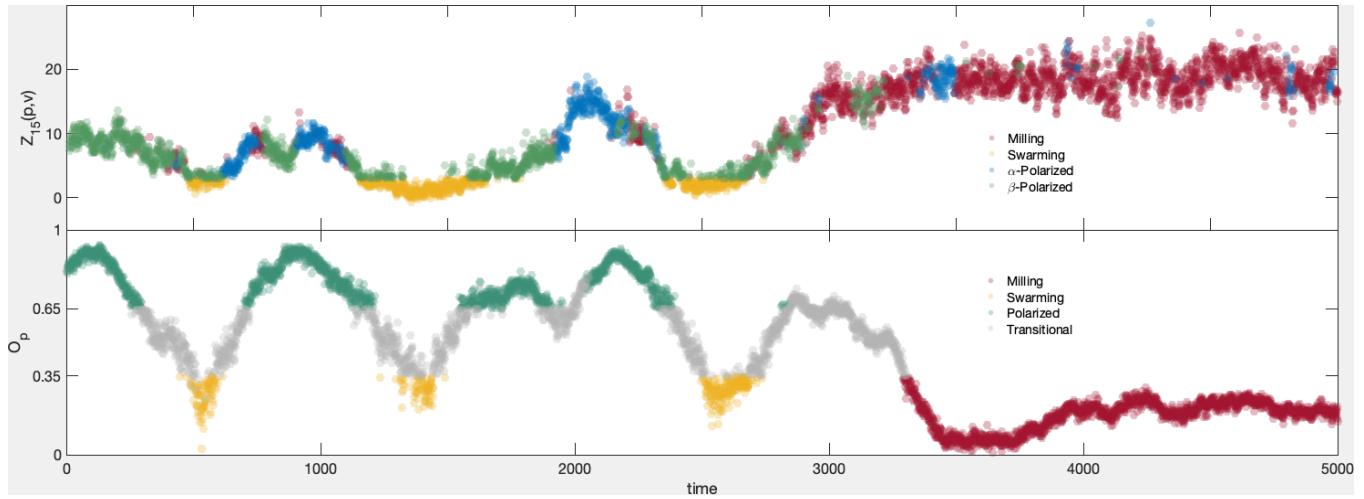


Fig. 5. Statistics calculated from 5000 frames of the golden shiner dataset. (Top): Map alignment $Z_{15}(p, v)$ calculated over 5000 video frames. The plots are colored according to our classification of macro-scale state; milling = red, swarming = yellow, α -polarized = blue, β -polarized = green. (Bottom): The group polarization macro-variable O_p calculated from the same 5000 frames. The coloring shows the states according to (17); milling = red, swarming = yellow, polarized = teal, unclassified/transitional = grey.

302 coordinate systems are unrelated, and that $Z_{15}(p, v)$ is correspondingly low.

303 Here again, the choice of $k = 15$ is driven by observing that outside of the top fifteen entries
 304 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
 305 for most frames (though this observation does break down for some frames when the eigenvalues
 306 become small enough). The results below prove to be robust to varying k from 10 to 25. Therefore
 307 the span of $\{\phi_1^{(1)}, \dots, \phi_{15}^{(1)}\}$ contains the majority of the structure within the data, while $k = 15$ is
 308 not so large that the signal in Z_k becomes diluted.

309 The two coherent behaviors we would like to separate are the polarized, or linear, motion (right
 310 column in SI Figure 1) and milling, or circling, motion (left column in SI Figure 1). Both states
 311 exhibit high levels of organization, and we must use a sharper tool to distinguish them.

312 To do this we employ the covariance vectors $\xi_{i,j}$, as defined in the Unsupervised Macrostate
 313 Classification section above, over the course of 5000 frames of video (see the Supplementary
 314 Information (SI) for more information on the dataset). First, those states with $Z_{15}(x, v) < 3$ are
 315 considered to exhibit incoherent behavior and to have no macro-scale dynamic, so we classify them
 316 as swarming frames and remove them from the set of frames to be classified. Then for each time
 317 t we compute the covariance vector $\xi(t) = (|\xi_{\sigma(i,j)}(t)|)$. We cluster the resulting vectors using the
 318 k -means algorithm, with $k = 3$. This leaves us with three groups representing distinct coherent
 319 behaviors, labelled G1, G2, and G3, as well as the incoherent group, which we label N . Comparing
 320 these group memberships to those given by calculating the group polarization and rotation, O_p and
 321 O_r , and classifying behavior as in (17) we find that group G1 contains 85.86% of the milling frames,

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				

322 G2 and G3 together comprise 87.47% of the polarized frames, while N holds 65.85% of the swarming
 323 frames (another 32.01% are relegated to G2). The full results of the classification appear in Table 1.

324 That the polarized frames are divided into two separate clusters suggests that there are two
 325 distinct modes of polarized behavior exhibited by the fish. Upon inspecting $\xi(t)$ for the frames
 326 from each cluster, we find that the two groups are essentially distinguished by their relative levels
 327 of organization; for example, the mean value of Z_{15} for cluster G2 was 6.3186, while for G3 the
 328 mean map alignment statistic is 12.0878. Practically, the frames of cluster G2 typically consist
 329 of polarized group motion with multiple subgroups of fish following distinct paths; we call this
 330 β -polarized behavior. This creates a macro-scale division in the network structure of $\hat{A}^{(v)}$ that is
 331 not present in the network structure of $\hat{A}^{(x)}$, which leads to a lower alignment between their top
 332 eigenvectors; here $\langle \xi_{1,1}(t) \rangle = 0.176$. On the other hand, the frames of G1 feature a single, unified
 333 group motion, leading to a high alignment ($\langle \xi_{1,1}(t) \rangle = 0.331$) between the top diffusion coordinates
 334 of each network. We refer to this as α -polarized schooling behavior.

335 One may approach the problem of classifying the covariance vectors $\{\xi(t)\}$ in numerous ways; we
 336 chose the k -means algorithm for its simplicity. However, it requires the user to fix the number of
 337 desired clusters beforehand. As the prior work of (17) only defined two separate coherent behaviors,
 338 milling and polarized, one may wonder what results we would find if we instead restricted the
 339 classification to two clusters. We did this and, labelling the groups G1* and G2*, found that G1*
 340 successfully captures all the milling frames, and G2* is predominantly composed of polarized frames
 341 (see Table 2 of SI). However, roughly 1 in 4 polarized frames are relegated to G1*. Table 3 in
 342 SI shows how the two k -means clusterings apportion the frames. In particular, the α -polarized
 343 frames tend to be classified as milling. By increasing the number of subgroups to $k = 3$, we allow
 344 the algorithm to perform a more sensitive clustering, which improved the correlation between our

345 groupings and those of (17), as well as distinguished between two forms of polarized behavior that
346 were grouped together in that work.

347 **Discussion**

348 To identify different states of collective behavior in complex systems, we have developed and applied
349 a new methodology based on measuring changes in the mutual geometry of a given system as it
350 is viewed through different lenses, using diffusion maps. The method is general and objectively
351 produces system-specific, macro-scale variables, Z_k and ξ for tracking the onset of and distinguishing
352 between different regimes of collective behavior. To test the method, we applied it to synthetic data
353 produced from a well-known model of birds flocking, and to empirical data on fish schooling. In both
354 cases the macro-scale quantities constructed using diffusion maps provided an accurate description
355 of the system in terms of the degree and type of collective behavior present.

356 The analysis, while data-driven, does suggest methods of controlling complex systems, or nudging
357 them towards a desired state, as in (19). For example, the top panel of Figure 5 suggests that
358 swarming behavior is typically bookended by a β -polarized state. Then to bring the school into a
359 swarming state from an α -polarized state, one might attempt to force the fish to break into multiple
360 polarized subgroups (by instituting an obstruction, say). Achieving a misalignment of the leading
361 diffusion coordinates from the spatial- and velocity-based networks of the school will clearly impede
362 any group consensus, making the school more likely to lose its coherence. This strategy is surely
363 obvious in the case of schooling fish, but an analogous tactic could be employed to inform control
364 over state change in markets or social networks as well.

365 Furthermore, the diffusion coordinates provide a quantitative basis for choosing how to align/misalign
366 a system; one can use *in silico* experiments to identify changes to the network structure which would
367 move the system towards the desired macro-state. Given a target covariance vector, the technique
368 could perhaps be automated. One approach would be to perform spectral clustering of the network
369 according to the present and target eigenfunctions. Agents belonging to the same cluster in both
370 cases could be expected to remain well-connected during the transformation, simplifying the search
371 for an appropriate perturbation to increasing or decreasing the influence between these submodules
372 of the system. Reducing the problem space in this way would be a first step towards an online
373 solution, advising system control in real time.

374 The modeled birds flocking and the empirical data on fish schooling were studied because the

375 collective behavior from each system is well known. That is, the scales at which collective behavior
376 emerges and the macro-scale variables that do a good job of describing group behavior are known.
377 This allowed for a clear demonstration of the new methodology, though one can apply this framework
378 to a variety of dynamical or static data sets, from subfields outside animal group dynamics, and
379 areas other than biology. In the two cases studied, familiarity with these particular systems and
380 their emergent properties made the choice to use position and velocity data snippets and their
381 respective metrics a straightforward decision. However, there are many other biological systems
382 where the organizational variables to choose may not be so obvious (20). Additionally, there are many
383 non-biological systems, such as financial and housing markets, city-transit systems, and power-grid
384 systems, which are known to exhibit emergent patterns at a range of scales, but for which the
385 relevant micro-scale variables are less well known. In these cases one may compute map alignment
386 in an uninformed fashion, testing a variety of metrics and variables, to discover latent emergent
387 relationships.

388 It is worth noting, however, that the approach that we have developed has its own challenges
389 relating to the nature of the data used. Diffusion maps inherently require a large number of data
390 points for the diffusion operator to exhibit the regularity of the limiting operator on the manifold from
391 which the data is sampled. With the fish data though, the intermittent nature of the observations
392 meant that the system size n varied between roughly 70 and 280 agents, and the still method proved
393 resilient across these fluctuations.

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

405 Measuring the map alignment between different networks induced by separate metrics on a graph

406 serves as an entry point for several other possible analyses. One could use a set of training data to
407 perform a classification of the various emergent behaviors of the system, then measure the stability of
408 the system (risk of state change) by the current distance from $\xi(t)$ to representative subgroups from
409 the various states. For example, if the distance from $\xi(t)$ to Group 1 is growing, and the distance
410 to Group 2 is shrinking, one might anticipate a shift in macrostate. In this way one may estimate
411 the risk of emergence, dissolution, or a large-scale shift of collective behavior. The authors believe,
412 based on preliminary analysis of the fish schooling data, that such an approach could provide a new
413 early warning metric for complex systems whose relevant macrovariables may be unknown, which
414 would inhibit the application of other early warning signals such as critical slowing down.

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

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

- 424 1. Maja J Mataric. From local interactions to collective intelligence. In *The Biology and Technology of Intelligent Autonomous Agents*, volume 144 of *Series F: Computer and Systems Sciences*.
425 Springer, 1995.
- 426 2. Peter Friedl, Joseph Locker, Erik Sahai, and Jeffrey E Segall. Classifying collective cancer cell invasion. *Nature Cell Biology*, 14, 2012. URL <https://doi.org/10.1038/ncb2548>.
- 427 3. Sachit Butail, Erik M Bollt, and Maurizio Porfiri. Analysis and classification of collective behavior using generative modeling and nonlinear manifold learning. *Journal of theoretical biology*, 336: 185–199, 2013.
- 428 4. Ronald R Coifman, Stephane Lafon, Ann B Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition
429 of data: Diffusion maps. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7426–7431, 2005.
- 430 5. Jens Krause, Graeme D Ruxton, and Graeme D Ruxton. *Living in groups*. Oxford University Press, 2002.
- 431 6. Iain D Couzin, Jens Krause, Nigel R Franks, and Simon A Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513, 2005.
- 432 7. Rama Cont and Jean-Philippe Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic dynamics*, 4(2):170–196, 2000.
- 433 8. Robert M May, Simon A Levin, and George Sugihara. Complex systems: Ecology for bankers. *Nature*, 451(7181):893, 2008.
- 434 9. Simon Levin, Tasos Xepapadeas, Anne-Sophie Crépin, Jon Norberg, Aart De Zeeuw, Carl Folke, Terry Hughes, Kenneth Arrow, Scott Barrett, Gretchen Daily, et al. Social-ecological systems as
435 complex adaptive systems: modeling and policy implications. *Environment and Development Economics*, 18(2):111–132, 2013.
- 436 10. Jianguo Liu, Thomas Dietz, Stephen R Carpenter, Marina Alberti, Carl Folke, Emilio Moran, Alice N Pell, Peter Deadman, Timothy Kratz, Jane Lubchenco, et al. Complexity of coupled human and
437 natural systems. *science*, 317(5844):1513–1516, 2007.
- 438 11. Simon Levin. Complex adaptive systems: exploring the known, the unknown and the unknowable. *Bulletin of the American Mathematical Society*, 40(1):3–19, 2003.
- 439 12. Ronen Talmon, Israel Cohen, Sharon Gannot, and Ronald R Coifman. Diffusion maps for signal processing: A deeper look at manifold-learning techniques based on kernels and graphs. *IEEE
440 signal processing magazine*, 30(4):75–86, 2013.
- 441 13. Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.
- 442 14. Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.
- 443 15. Tyrus Berry and Timothy Sauer. Consistent manifold representation for topological data analysis. In *Some Book...*, 2016.
- 444 16. John Toner and Yuhai Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Physical review E*, 58(4):4828, 1998.

- 446 17. Yael Katz, Kolbjørn Tunstrøm, Christos C Ioannou, Cristián Huepe, and Iain D Couzin. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, 108(46):18720–18725, 2011.
- 447 18. Kolbjørn Tunstrøm, Yael Katz, Christos C. Ioannou, Cristián Huepe, Matthew J. Lutz, and Iain D. Couzin. Collective states, multistability and transitional behavior in schooling fish. *PLOS Computational Biology*, 9(2):1–11, 02 2013. . URL <https://doi.org/10.1371/journal.pcbi.1002915>.
- 448 19. Marco A. Janssen. Targeting individuals to catalyze collective behavior in social networks. In *CSSSA 2011 Papers*, 2011. URL <http://computationalsocialscience.org/wp-content/uploads/2011/10/Janssen-CSSSA2011.pdf>.
- 449 20. Alessandro Attanasi et al. Collective behaviour without collective order in wild swarms of midges. *PLoS Comput Biol*, 10(7), 2014.
- 450
- 451
- 452

DRAFT

Unsupervised Manifold Learning of Collective Behavior

Mathew Titus^{1,a,b}, George Hagstrom^c, Zach Gelbaum^{a,b}, and James R. Watson^a

^aCollege of Earth, Ocean, and Atmospheric Sciences, 104 CEOAS Administration Building, Oregon State University, Corvallis, Oregon, 97331 USA; ^bThe Prediction Lab LLC, Corvallis, Oregon, 97331 USA; ^cDept. Ecology and Evolutionary Biology, 104a Guyot Hall, Princeton University, Princeton, New Jersey, 08544 USA

¹ Supplementary Information (SI)

Idealized Simulations of Bird Flocking. 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:

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

$$p_i(t+1) = p_i(t) + v(\cos(\theta_i(t)), \sin(\theta_i(t))) \quad [2]$$

where $\langle \theta_j(t) \rangle_{i,r}$ is the mean bearing of the agents within $B_i(r)$ and η 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.) Thus each agent's position is updated by flying a distance v in the direction determined (up to the perturbation η) 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 ?? 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. 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.

16 Image analysis was employed to locate the fish within each frame, then match them across frames
17 when possible. This allowed the research team to associate a unique identifier with each fish while
18 its trajectory was under observation.

19 These fish are highly social, and due to the difficulty of rendering individual fish when they are
20 packed tightly into a region, and the possible conflation of two separate fish swimming near one
21 another, individual fish are often ‘dropped’ from the data set. When this occurs the identifier
22 associated with that fish is retired, and the position and velocity data for that particular trajectory
23 terminates. However, when the fish again exceeds the liminal level of the image analysis it is
24 registered as a ‘new’ fish and given a new identifier. As a result, the data garnered from a typical
25 frame records position and velocity information for only 60 - 80% of the members, and individuals
26 are typically tracked for about 200 frames (6.7 seconds) before its image is ‘lost’ and its identifier
27 removed from future data. The resulting data set consists of 100298 frames (about 56 minutes) of
28 position and velocity data for the school. For the study in the manuscript we selected a 5000 frame
29 subset containing roughly an equal number of frames devoted to each of the behaviors observed
30 (frames 40001 to 45000). This enabled us to use the simple k -means clustering algorithm, which
31 otherwise suffers when the subgroups to be identified have widely varying sizes. (This is because
32 the classical k -means algorithm creates hyperspherical clusters, where the hyperspheres must be
33 centered within the convex hull of the data and have equal radii. Smaller clusters necessitate a
34 description via smaller hyperspheres. See (1)).

35 As discussed in (3), human observation of the schooling behavior identified three separate behavioral
36 regimes for the fish: milling, swarming, and polarized (see Figure 1). In the milling state, the fish
37 rotate as a group, circulating about a fixed point; this is a highly coherent group state. While
38 swarming, the fish trajectories are disordered, with very little discernible group alignment. In this
39 state the school is stationary, though individual fish may be moving in an uncoordinated fashion.
40 Last, the second coherent group structure is the polarized state, which is characterized by the group
41 having a coherent direction of motion, and strong local alignment among the agents.

42 As with the flocking model above, the variables of interest are the two-component vectors of fish
43 position, p_i , and fish velocity, v_i .

⁴⁴ **Metrics.** Let $x_i \in X$ be the set of observations for an agent (either a particle in the flocking simulation
⁴⁵ 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^{th}
⁴⁶ agent's position, heading, and velocity, respectively. The first metric, $d^{(1)}$, is simply the euclidean
⁴⁷ distance between the agents,

$$48 \quad d^{(1)}(x_i, x_j; t) := \|p_i(t) - p_j(t)\|_2. \quad [3]$$

⁴⁹ For the flocking simulation, the second metric is the following

$$50 \quad d^{(2)}(x_i, x_j; t) := \left(\sum_{k=0}^{30} g(k) \|v_i(t-k) - v_j(t-k)\|_2^2 \right)^{\frac{1}{2}} \quad [4]$$

⁵¹ where $v_i(t) = v(\cos(\theta_i(t)), \sin(\theta_i(t)))$ is the i^{th} velocity at time t . The smoothing term $g(k)$ is
⁵² proportional to $\exp(-k^2/30)$, but normalized so that $\sum_k g(k) = 1$. That is, $d^{(2)}$ is a gaussian
⁵³ smoothed time-average of the L^2 norm between velocities over the past 30 time steps.

⁵⁴ For the fish, we consider a generalization of the previous expression for $d^{(2)}$. Rather than comparing
⁵⁵ the two fishes' heading time series at simultaneous moments, we allow one time series to be shifted
⁵⁶ into the past by a lag of up to $L = 60$ video frames, or two seconds. Define

$$57 \quad \langle x_i, x_j \rangle_{L,t} := \min_{0 \leq l \leq L} \left(\sum_{k=0}^{30} g(k) \|v_i(t-k) - v_j(t-k-l)\|_2^2 \right)^{\frac{1}{2}},$$

⁵⁸ where $g(k)$ is again a gaussian smoothing kernel as in the previous paragraph. Thus l represents a lag
⁵⁹ allowing us to compare not only concurrent velocity profiles over 30 steps, but also cases where fish
⁶⁰ i is mimicking a trajectory of fish j up to two seconds in the past. Note that $\langle x_i, x_j \rangle_{L,t} \neq \langle x_j, x_i \rangle_{L,t}$,
⁶¹ as the righthand term calculates the distance between the trajectories under the assumption that j
⁶² is following i . We define our distance as

$$63 \quad d_L(x_i, x_j; t) := \min\{\langle x_i, x_j \rangle_{L,t}, \langle x_j, x_i \rangle_{L,t}\}. \quad [5]$$

⁶⁴ **Accessing the Data.** The data can be obtained by request to the corresponding author, and will be
⁶⁵ made publicly available in the future at the ScholarsArchive@OSU repository hosted by Oregon
⁶⁶ State University.

SI Tables.

67

Table 1. 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 2. 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, α -polarized, and β -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%)
$G2^*$	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%)

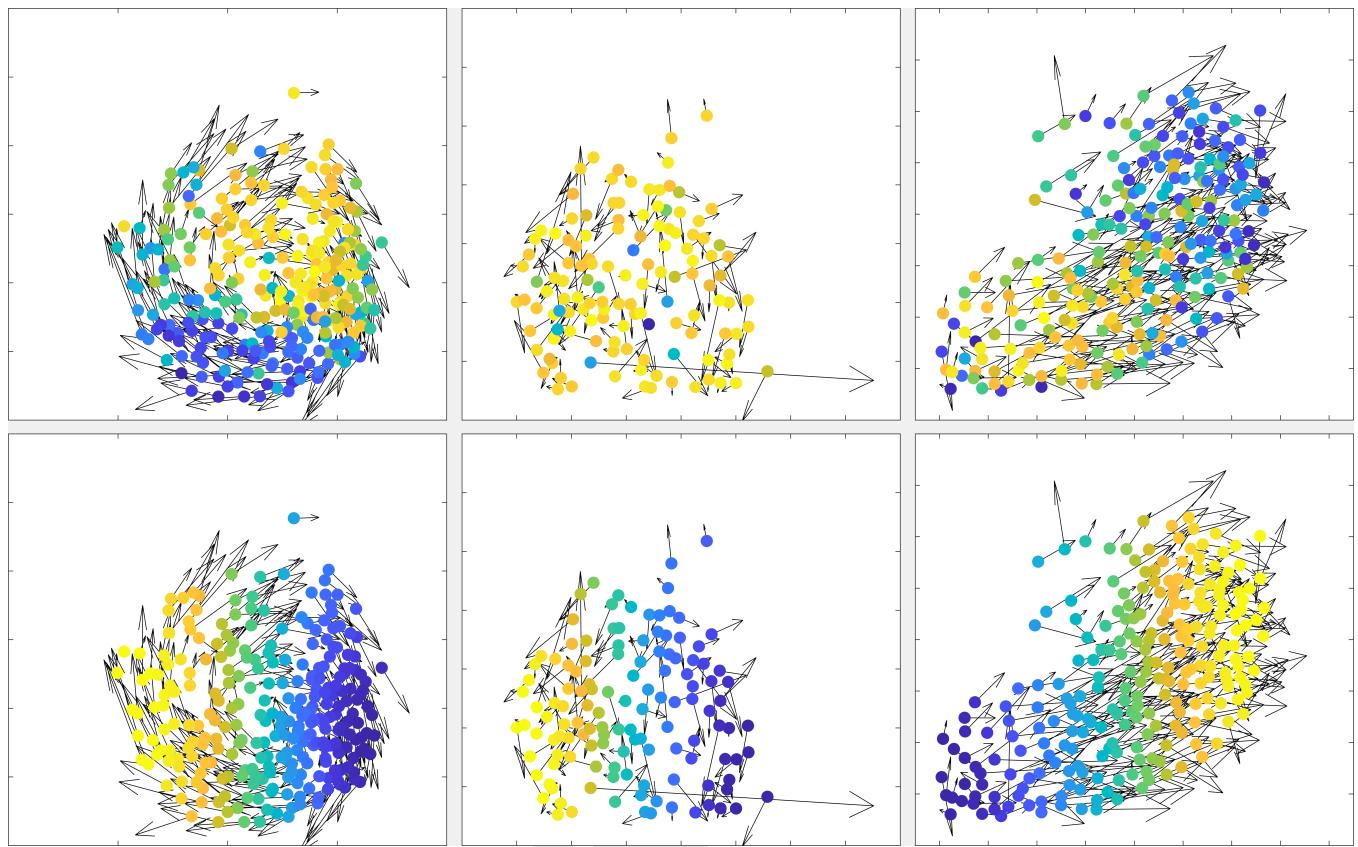


Fig. 1. Plots of fish position (colored markers) and velocity (arrows) at three example times, colored by diffusion map coordinates $\phi_1^{(v)}$ [top row; i.e. produced from analyzing correlations in velocities] and $\phi_1^{(x)}$ [bottom row; i.e. produced from analyzing spatial proximity]. The emergent behaviors from left to right are milling, swarming, and polarized motion.

- 69 1. Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323.
- 70 2. Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C., and Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*,
71 108(46):18720–18725.
- 72 3. Tunstrøm, K., Katz, Y., Ioannou, C. C., Huepe, C., Lutz, M. J., and Couzin, I. D. (2013). Collective states, multistability and transitional behavior in schooling fish. *PLOS Computational Biology*,
73 9(2):1–11.

DRAFT