

DSCC2018-9139

DIFFERENTIATION OF COLLECTIVE BEHAVIOR BASED ON AUTOMATED DISCOVERY OF DYNAMICAL KINDS

Amanda Hashimoto, Nicole Abaid*
Engineering Mechanics Program
Virginia Tech
Blacksburg, Virginia, 24061
Email: ahashimo, nabaid@vt.edu

Subhradeep Roy, Benjamin Jantzen
Department of Philosophy
Virginia Tech
Blacksburg, Virginia, 24061
Email: sdroy, bjantzen@vt.edu

Colin Shea-Blymyer
Department of Computer Science
Virginia Tech
Blacksburg, Virginia, 24061
Email: c0lin@vt.edu

ABSTRACT

In this paper, we explore a model of collective behavior using EUGENE, an algorithm for automated discovery of so-called “dynamical kinds”. Two systems are of the same dynamical kind if their underlying causal dynamics are similar, as defined using dynamical symmetry. We apply EUGENE to simulation data from a model capable of generating a range of qualitatively different collective behaviors, from aligned motion to circular milling. These behaviors are measured using both global and local order parameters, and this data is analyzed with EUGENE. We find that EUGENE is capable of differentiating between these systems when global order parameters are used, and can only identify more coarse characteristics when local order parameters are considered.

INTRODUCTION

Collective behavior refers to group-level complexity that can emerge from local interactions among multiple individuals. Besides being studied extensively in natural animal groups [1], such as bird flocks and fish schools, collective behavior is of interest to the larger scientific community for its ability to generate desired behaviors without directly controlling each individual in a group. This work finds application in engineering in particular, for example, in the design and control of robotic swarms, granular media, and engineered organisms [2–4].

Collective behavior can be generated in models for multi-agent systems that apply agent-based rules [5] or view the group

as a continuum [6]. These models have been shown capable of capturing a variety of behaviors that either replicate observations of nature [7] or are relevant to an engineering application [8]. As an example, the model in [9] defines interactions among a group of self-propelled particles through potential functions and includes an external source to diversify the types of collective behaviors the model can produce. These behaviors include aligned motion, unaligned swarming, and circular milling patterns, and are distinguished by selecting model parameters.

For the model in [9] and in the literature in general, collective behavior is quantified by defining order parameters, whose values capture the level of coordination in the group. Examples of order parameters include linear momentum [10], angular momentum [7], and a measure of closeness in space [9]. A detractor of such order parameters is that they are defined irrespective of the system dynamics and take a human-centered, global perspective on the group’s behavior. Not only does this ubiquitous view neglect the perspective of an individual (which would be relevant to an animal group, for example), there has not been a rigorous demonstration of the parameters’ optimality for measuring order or complexity as far as we know. This open question on what order parameters are appropriate to measure collective behavior may be answered by using model-free data-driven methods to directly study collective dynamics without imposing any model on the system.

Data-driven methods have been used to study collective behavior, at least in terms of pairwise relationships that can be used to build an interaction network. For example, information flow has been measured to detect animal group interactions among

*Address all correspondence to this author.

fish [11, 12] and insects [13] moving in groups, and time-delayed embeddings have been used for studying causality in economic systems [14] and neural networks [15]. These studies use a variety of methods, such as entropy (based in information theory) and manifold embedding (taken from dynamical systems). To a lesser extent, tools from the field of causal discovery have been used to learn about the causal structure of collectives in terms of both individual-level relationships and connections between high-level features of the collective [16, 17]. Recently, work on the automated discovery of scientifically relevant kinds has provided a set of tools for indirectly assessing the similarity of underlying dynamics in complex systems [18]. However, these tools have yet to be applied to the empirical investigation of collective behavior.

To gain a broad understanding of how data-driven analytical methods can be used to quantify collective behavior, it is necessary to test a system for which the dynamics can be precisely defined. To be useful for investigating collective behavior in real systems where it is difficult or impossible to obtain accurate information about the behavior of all members, we ideally require an analytical method that is sensitive to detailed causal structure and insensitive to possible confounding factors. Such factors include the choice of variables used to describe the collective, measurement noise, and stochasticity in the underlying dynamics. A new algorithm in the EUGENE collection of automated discovery tools [19] satisfies all of these desiderata.

In this paper, we seek to understand the potential of EUGENE to study collective behavior using an established model. We demonstrate the application of EUGENE to measure different behaviors generated by the self-propelled particle (SPP) model in [9]. We show that this algorithm is capable of distinguishing between different behaviors using low-dimensional parameters as input, rather than the high-dimensional full state of the system. In addition, we present a local version of such conventional order parameters, to which EUGENE can be analogously applied, and we find that results are less sensitive to behavioral differences. We find that EUGENE can differentiate behaviors in model data, even when the full state of the system is not given. This work supports EUGENE's use in future work to analyze field data from animal groups for which the full dynamics may be very difficult to observe and the rules governing interactions cannot be known.

SELF-PROPELLED PARTICLE MODEL

SPP model with external leader particle

The self-propelled particle model used in this study investigates the collective behavior of a multi-agent system interacting with an external leader particle (ELP). This model is developed and studied in [9] for the wide range of collective behaviors it produces. The agents in the group are N identical, self-propelled particles of mass m moving in two dimensions. The position

vector of the i -th particle is denoted as $\mathbf{x}_i(t)$, $i = 1, \dots, N$, and its velocity as $\mathbf{v}_i(t) = \dot{\mathbf{x}}_i(t)$, where $t \in \mathbb{R}$ is the time variable. Interactions between particles are governed by attractive and repulsive forces, and all agents also interact with a fixed ELP. The dynamics of the i -th agent, $i = 1, \dots, N$, are given by

$$m\dot{\mathbf{v}}_i = (\alpha - \beta\|\mathbf{v}_i\|^2)\mathbf{v}_i + \mathbf{F}_i + \mathbf{F}_i^0 \quad (1)$$

with $\alpha, \beta > 0$, $\alpha\mathbf{v}_i$ is a self-acceleration term, and $-\beta\|\mathbf{v}_i\|^2\mathbf{v}_i$ is a friction term. The force between agents is

$$\mathbf{F}_i = - \sum_{\substack{j=1 \\ j \neq i}}^N \nabla_{\mathbf{x}_i} \Phi(\|\mathbf{x}_{ij}\|) \quad (2)$$

where $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ and Φ is a generalized Morse potential given by $\Phi(z) = -c_a e^{-z/l_a} + c_r e^{-z/l_r}$ for a positive scalar z . The constants c_a , c_r , l_a , and l_r give the strength and cut-off length for the attractive and repulsive forces, labeled with a and r , respectively. The force between the ELP and agent i is

$$\mathbf{F}_i^0 = -\gamma \nabla_{\mathbf{x}_i} \Phi(\|\mathbf{x}_{i0}\|) \quad (3)$$

where $\mathbf{x}_{i0} = \mathbf{x}_i - \mathbf{x}_0$ is the relative position with respect to the ELP, and γ is the strength of interaction between particles ($\gamma \geq 0$). By varying γ and the self-acceleration term α in simulation, we analyze the generated behaviors with EUGENE.

Global order parameters

In line with the analysis in [9], we compute three physics-based order parameters using the state of the whole system of particles. These parameters capture linear and angular momenta, and hence alignment and rotation of the group with respect to the ELP, and cohesion with respect to the ELP's fixed position. At time t , the normalized group linear momentum (polarization) $P(t)$, the normalized group angular momentum $M(t)$, and a measure of cohesiveness $C(t)$ with respect to the ELP are defined as

$$P(t) = \frac{\|\sum_{i=1}^N \mathbf{v}_i(t)\|}{\sum_{i=1}^N \|\mathbf{v}_i(t)\|} \quad (4)$$

$$M(t) = \frac{\|\sum_{i=1}^N \mathbf{x}_{i0}(t) \times \mathbf{v}_i(t)\|}{\sum_{i=1}^N \|\mathbf{x}_{i0}(t)\| \|\mathbf{v}_i(t)\|} \quad (5)$$

$$C(t) = \frac{1}{N} \sum_{i=1}^N \exp \left[-\frac{\|\mathbf{x}_{i0}(t)\|}{4l_a} \right] \quad (6)$$

We comment that, while the polarization is independent of the ELP, the angular momentum and cohesion are defined based on

relative velocities or positions with respect to the ELP. In [9], the authors found that these parameters allowed for distinguishing between distinct states better than considering analogous order parameters defined with respect to the system's center of mass, which would be an intuitive alternate form. As a note, M and C in this work correspond to M^0 and C^0 in [9].

Local order parameters

While we observe in [9] that the group's state is qualitatively well-described by these order parameters, it is often extremely difficult to obtain such detailed information from a natural system, particularly with very large groups. To complement these global parameters, we also define local order parameters from the perspective of an individual agent, with motivation that this data may be more easily attainable and may in fact be useful to recover the group-level behavior.

We define local order parameters with respect to a randomly-chosen focal agent f , which is selected independently and used for all time steps in a single simulation. Once this focal agent is selected, we define a range R and consider all agents in the ball of radius R centered at the focal agent's position, $\mathbf{x}_f(t)$. We call this set of agents $N^L(t)$. The local linear momentum (local polarization) with this limited perspective is

$$P^L(t) = \frac{\|\sum_{i \in N^L(t)} \mathbf{v}_i(t)\|}{\sum_{i \in N^L(t)} \|\mathbf{v}_i(t)\|} \quad (7)$$

The local angular momentum is defined using the relative position of agents in $N^L(t)$ with respect to $\mathbf{x}_f$. Specifically,

$$M^L(t) = \frac{\|\sum_{i \in N^L(t)} \mathbf{x}_{if}(t) \times \mathbf{v}_i(t)\|}{\sum_{i \in N^L(t)} \|\mathbf{x}_{if}(t)\| \|\mathbf{v}_i(t)\|} \quad (8)$$

where $\mathbf{x}_{if} = \mathbf{x}_i - \mathbf{x}_f$ is the relative position of the agents with respect to the selected focal agent. The cohesion is similarly defined with respect to the focal agent, as

$$C^L(t) = \frac{1}{|N^L(t)|} \sum_{i \in N^L(t)} \exp \left[-\frac{\|\mathbf{x}_{if}(t)\|}{4l_a} \right] \quad (9)$$

where $|\bullet|$ gives the cardinality of a set.

Five distinct behavioral states

Since the goal of this paper is generating simulation data evidencing different behaviors in a system of particles, we select model parameters based on the parameter study in [9] that show a range of states. For these simulations, we take $N = 50$,

TABLE 1. SYSTEM BEHAVIORS AND THEIR ASSOCIATED α AND γ . THE TRIPLE $[P, M, C]$ CORRESPOND TO GENERALLY ORDERED (1) AND DISORDERED (0) STEADY STATES, AND INCLUDE A SHORT DESCRIPTION OF THE BEHAVIOR.

State	α	γ	$[P, M, C]$	Behavior
A	0.2	0.0001	[0, 0, 0]	No organized motion
B	0.02	0.3	[0, 0, 1]	Cohesive unaligned motion
C	0.8	0.8	[0, 1, 1]	Milling on a circle
D	0.5	4.0	[1, 0, 0]	Cluster on a linear path
E	0.2	6.0	[1, 1, 1]	Cluster on a circular path

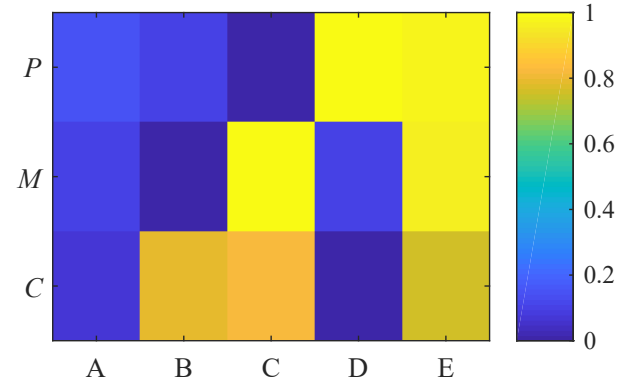


FIGURE 1. MEAN VALUES OF STEADY-STATE GROUP MOMENTUM (P), ANGULAR MOMENTUM (M), AND COHESIVENESS (C) WITHIN THE 5 BEHAVIORAL STATES.

$m = 1$, $\beta = 0.5$, $l_r = 0.5$, $l_a = 2$, $c_a = 0.5$, and $c_r = 1$. The parameters α and γ are varied to obtain different group behaviors. We use two different sets of initial conditions to generate random position and velocity vectors for the agents. The agents are initially dispersed uniformly within a square of side length $2l_a$ that is centered around the origin. For all simulations, the same initial conditions are used to allow for the comparison that EUGENE requires.

The simulation is run for the time interval $[0, 1000]$ and the entire run (including the initial transient) is used for analysis. This mostly transient data was selected since EUGENE relies on creating maps between systems at a variety of states. In contrast to most analyses for models of collective behavior, analysis with EUGENE is much less informative when the system is in steady-state.

We generate data for five different system behaviors (denoted A, B, C, D, and E) by varying α and γ , summarized in

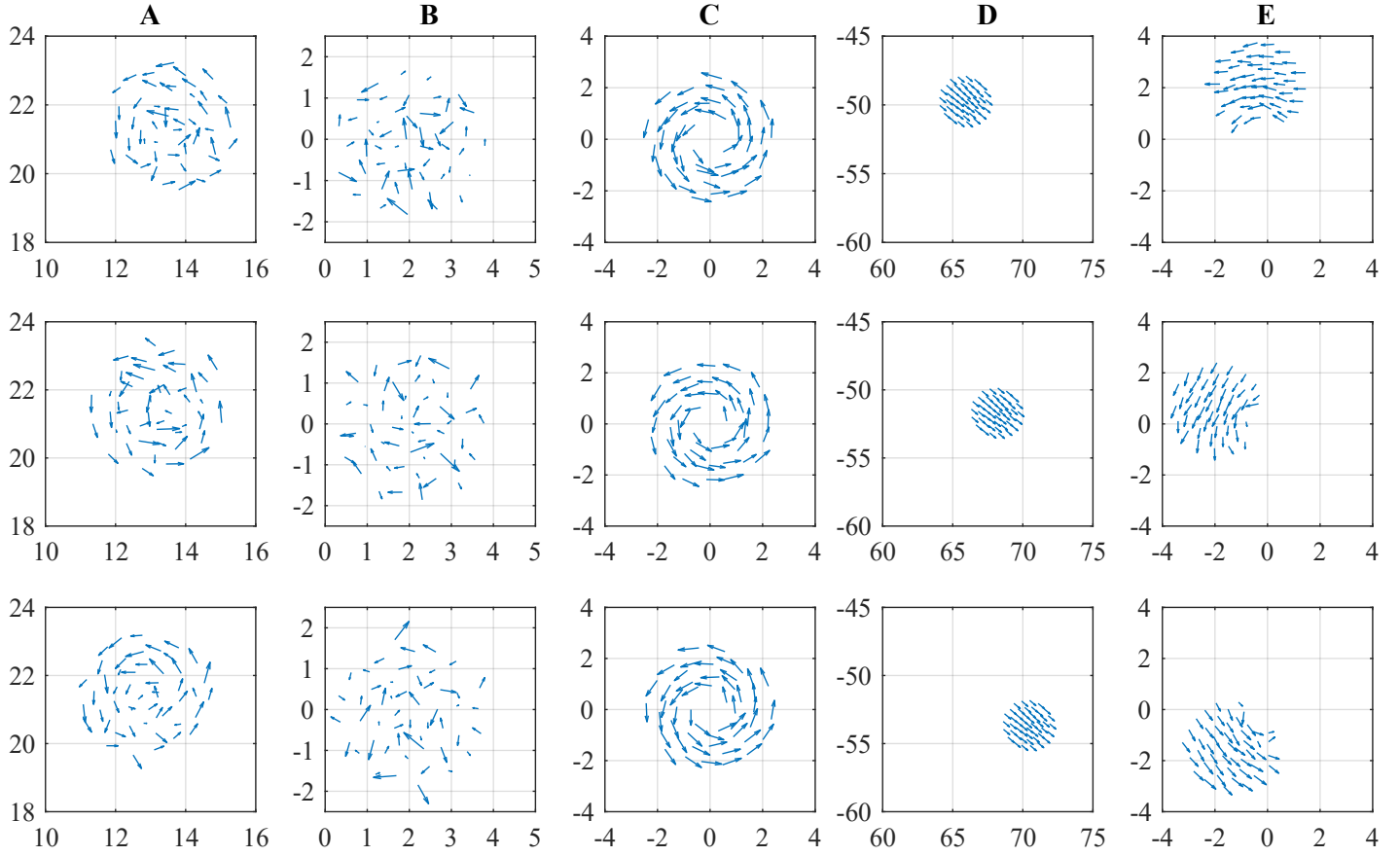


FIGURE 2. SNAPSHOTS OF GROUP MOTION FOR THE FIVE BEHAVIORAL STATES (LABELED FOR EACH COLUMN). FIGURES IN THE FIRST, SECOND, AND THIRD ROW ARE AT $t = 950$, 960 , AND 970 , RESPECTIVELY. ELP IS LOCATED AT THE ORIGIN.

table 1. These simulations capture a range of behaviors. State A acts as a control condition, since the particles show no strong alignment, rotation, or cohesion about the ELP. State B shows cohesive behavior, but with no alignment or organized milling. State C shows particles moving on an approximately circular trajectory, with agents distributed all along the circumference of the circle similarly to motion on the rim of a wheel. State D shows highly aligned motion of the group as a whole along a linear path. State E shows motion again on an approximately circular trajectory, but in contrast to State C, the particles are concentrated in a cohesive group and move together in an aligned fashion. To ensure that these model parameters generate the desired behaviors, we simulated the model for the time interval $[0,1000]$ and computed the order parameters over the last 750 seconds. These values are given in figure 1 and summarized generally as “ordered” (1) or “disordered” (0) in table 1.

Snapshots of representative steady-state group motion over time for these behaviors are shown in figure 2. Nevertheless, the analysis detailed below is performed on the transient data for

each simulation. Time series showing this data for each combination of order parameter and behavior are shown in figure 3, with both the global and local order parameters for polarization, angular momentum, and cohesion plotted over the time interval $[0,1000]$. These data are used as input for the analysis with EUGENE.

EUGENE ALGORITHM FOR AUTOMATED DISCOVERY

EUGENE is an algorithm for determining the similarity between two systems’ underlying causal dynamics [18]. Unlike examples in the existing literature which rely on low dimensional [20] or time delay [21] embeddings, it leverages dynamical symmetries - commutations between an evolution and transformation of a system - to verify that two systems are of different kinds. The original algorithm, however, is not designed for stochastic systems or cases of partial information, so we implement an extension of EUGENE that compares samples from carefully selected distributions over states of one system with the

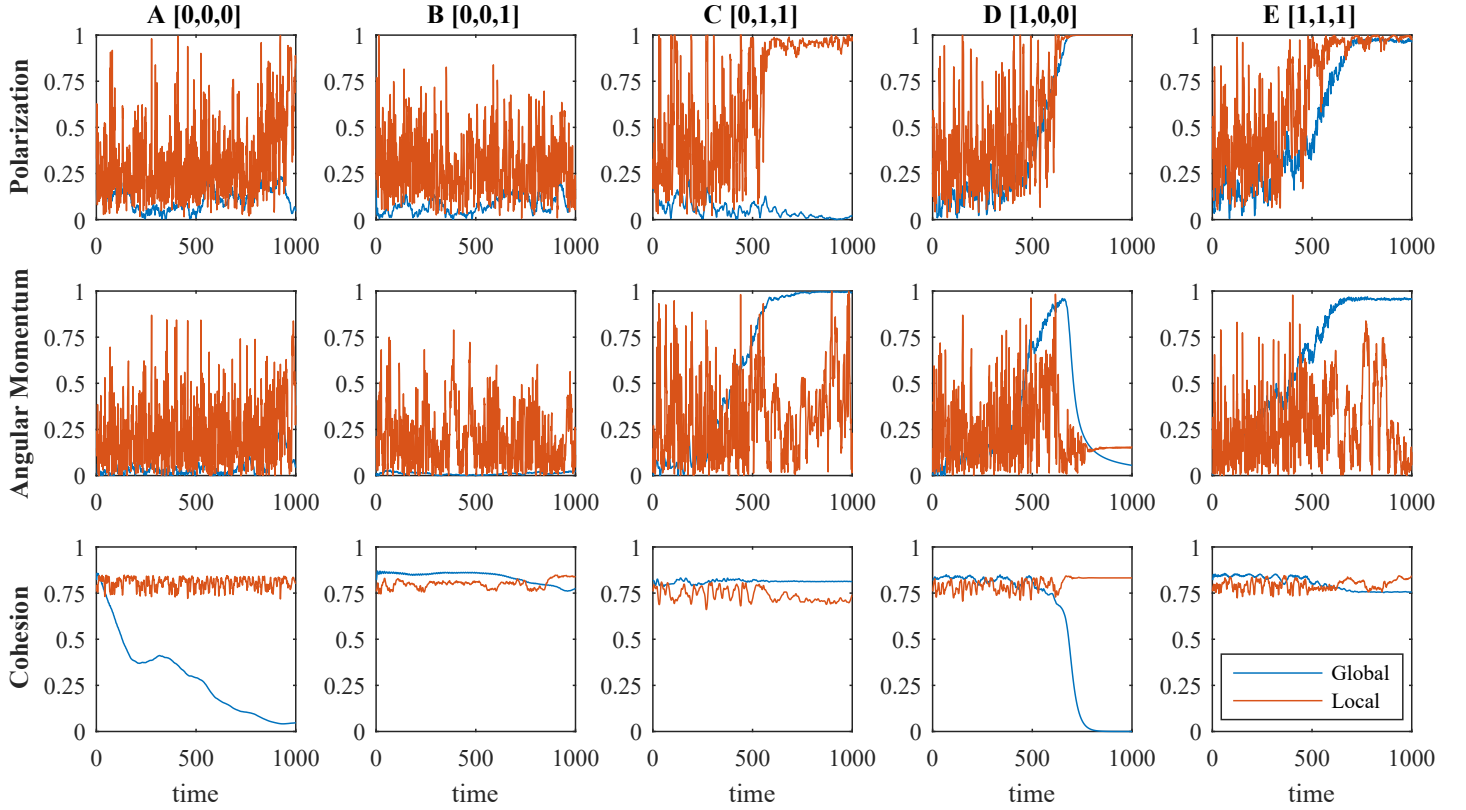


FIGURE 3. TIME SERIES DATA FOR EACH BEHAVIORAL STATE (LABELED FOR EACH COLUMN). THE FIGURES IN THE FIRST, SECOND, AND THIRD ROW CORRESPOND TO POLARIZATION, ANGULAR MOMENTUM, AND COHESION, RESPECTIVELY. THE TWO LINES ARE THE GLOBAL AND LOCAL ORDER PARAMETERS CALCULATED FOR EACH TIME SERIES.

corresponding distribution from another system. These distributions will be identical if and only if the two systems share the same dynamical symmetry, and the degree to which the distributions diverge provides an informative measure of how much the dynamics of one system diverges from that of the other.

Dynamical symmetries and kinds

As given in [22], a *dynamical symmetry* is a transformation σ on the variables of a system such that the final state of the system is identical regardless of whether σ is applied to the system before or after an intervention on the system's index variable (e.g. transformation over time). This definition was generalized in [18] to accommodate stochastic dynamics, which we adopt in the following form:

Definition 1 (Dynamical symmetry). Let V be a set of random variables. Let σ be an intervention on the variables in $Int \subset V$. The transformation σ is a dynamical symmetry with respect to some index variable $X \in V - Int$ if and only if σ has the following property: for all probability distributions f and g ,

the final probability distribution over V is the same whether σ is applied when the distribution of X is given by $p_x(x) = f(x)$ and then an intervention on X makes it such that $p_x(x) = g(x)$, or the intervention on X is applied first, changing its distribution from $f(x)$ to $g(x)$, and then σ is applied.

Dynamical symmetries are sensitive to both causal structure (i.e., the set of relations indicating which variables directly influence which others) and the functional form of causal influence. Thus, if the (possibly stochastic) dynamics underlying one system is described by, say, a differential equation of slightly different form than that of another system, the two will generally possess distinct yet similar dynamical symmetries. Thus, if one can directly compare the dynamical symmetries between systems with unknown dynamics, one can quantify their degree of dynamical similarity in a manner that is data-driven and model-free.

Current approach

The EUGENE algorithm that has been adapted from [19] can detect differences in dynamical symmetries in this general

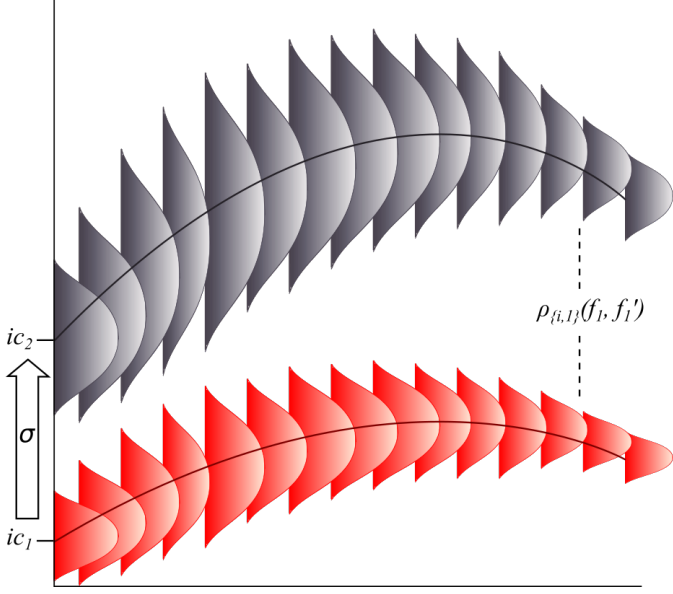


FIGURE 4. SCHEMATIC OF EUGENE ALGORITHM. f_1 AND f'_1 ARE COMPOSED OF MULTIPLE REALIZATIONS STARTING FROM INITIAL CONDITIONS ic_1 AND ic_2 , RESPECTIVELY, AND EVOLVING IN TIME HORIZONTALLY. RESULTING TIME SERIES ARE VIEWED AS THE RED AND GRAY DYNAMIC DISTRIBUTIONS (SNAPSHOTS SHOWN CHANGING OVER TIME). EUGENE TAKES THE POINTWISE CUMULATIVE JOINT DISTRIBUTION BETWEEN THESE TWO FUNCTIONS TO CHARACTERIZE SYSTEM A. THE SAME PROCESS IS PERFORMED FOR SYSTEM B, AND THE ENERGY DISTANCE BETWEEN THE TWO DISTRIBUTIONS DEFINES DIFFERENCE IN DYNAMICAL KIND.

sense. To do so, the algorithm requires two sets of time series for each of the systems A and B to be compared. Each set of time series contains a realization of the evolution of the system starting from the same initial condition. If the system is stochastic (or if the variables in terms of which it is described are incomplete), each realized time evolution in a set will be distinct. However, the distribution over states at any given time is fixed for each system (assuming the dynamics are unchanging). The function that maps this distribution for one initial condition (ic_1) to the distribution for another initial condition (ic_2) in the same system is a dynamical symmetry. By comparing the function for system A with that for system B, we can assess the similarity of their dynamics. Rather than estimate these functions directly, the EUGENE algorithm compares the joint distribution over states for system A at a time t after ic_1 and at a time t after ic_2 with the corresponding distribution from system B. Cumulative density functions are compared by computing the energy distance between them [23].

For analyzing data from the SPP model, the EUGENE al-

gorithm requires two realizations of the simulation for each set of parameter values of interest, where each realization is a multi-dimensional time-series spanning a fixed amount of time, Δt . The first realization is computed for a certain initial condition, ic_1 . Using the same simulation parameters, a second set of realizations is produced using a second initial condition, ic_2 . We refer to the time series of the first realization as $\vec{f}_1$, and that of the second realization as $\vec{f}'_1$, where each is of length n corresponding to the number of samples in the series. Then, using a new set of simulation parameters, two more sets of realizations are given using the same initial conditions used previously. These new sets of realizations will be referred to as $\vec{f}_2$ and $\vec{f}'_2$. Both $\vec{f}_1$ and $\vec{f}'_1$ belong to the first simulation configuration, while $\vec{f}_2$ and $\vec{f}'_2$ belong to the second. Similarly, $\vec{f}_1$ and $\vec{f}_2$ share initial conditions (ic_1), while $\vec{f}'_1$ and $\vec{f}'_2$ share a different set of initial conditions (ic_2). Each pair of time series elements $f_{1,i}$ and $f'_{1,i}$ for $i \in \{1, 2, \dots, n\}$ (that is, each pair of values at a given time for a given system) is treated as a realization of a pair random variables with a joint density function, $\rho_{1,i}(f_1, f'_1)$. The same is true for pairs $f_{2,i}$ and $f'_{2,i}$ which are presumed to be governed by a joint density function, $\rho_{2,i}(f_2, f'_2)$. Consider the cumulative distribution cdf_A over pairs of simultaneous values of f_1 and f'_1 , and cdf_B over pairs of values of f_2 and f'_2 , where $cdf_A(F_1, F'_1) = \sum_{i=1}^n \frac{1}{n} \int_{-\infty}^{F_1} \int_{-\infty}^{F'_1} \rho_{1,i}(f_1, f'_1) df_1 df'_1$ and $cdf_B(F_2, F'_2) = \sum_{i=1}^n \frac{1}{n} \int_{-\infty}^{F_2} \int_{-\infty}^{F'_2} \rho_{2,i}(f_2, f'_2) df_2 df'_2$. If the marginal distribution over values of f_1 is approximately the same as that over f_2 , then the difference between cdf_A and cdf_B is driven entirely by the difference in dynamical symmetries between systems A and B. We compute a metric to quantify this difference. Specifically, EUGENE computes an estimate of the energy distance between cdf_A and cdf_B [23]. A schematic for this algorithm is given in figure 4.

RESULTS AND DISCUSSION

By using the SPP model as an input into EUGENE, we confirm that EUGENE is able to identify the distinct behavioral states by distinguishing them as different dynamical kinds. In figure 5, the pairwise distances between the five behavioral states are shown as distance matrices for each of the global and local order parameters. The color axis on the right denotes the pairwise distances between the states, with distance = 0 corresponding to the two states being of the same dynamical kind and distance > 0 corresponding to the states being of different dynamical kinds. The magnitude of this difference is measured by the distance.

Despite the fact that the model has not converged to steady state as seen in figure 3, EUGENE is able to pick up trends for the global order parameters, which are shown in the first row of figure 5. The expected trends for the five system behaviors are the $[P, M, C]$ values in table 1. For global polarization P , for example, EUGENE finds that the behavioral States A, B, and C are the same, which is to be expected since their P values are low.

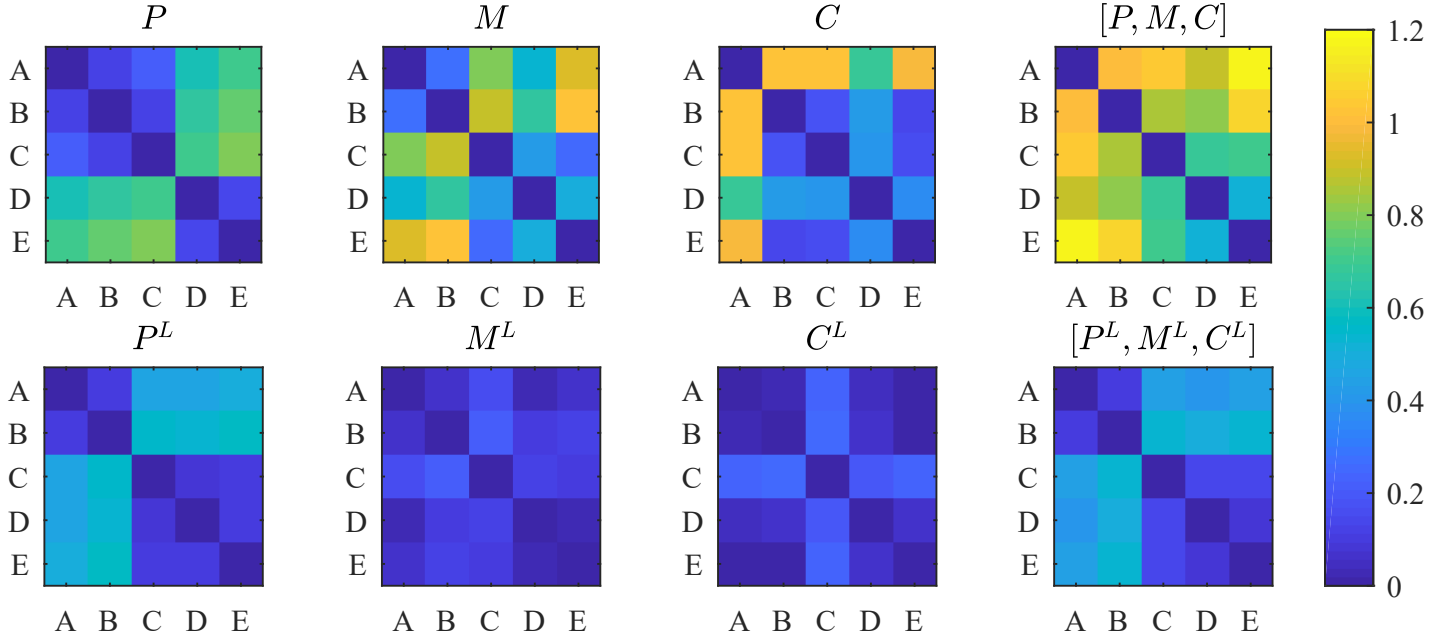


FIGURE 5. DISTANCE BETWEEN FIVE BEHAVIORAL STATES AS COMPUTED BY EUGENE. EACH SUBFIGURE SHOWS A DISTANCE MATRIX, WHERE THE PAIRWISE DISTANCE BETWEEN STATES LABELLED ON ROWS AND COLUMNS IS DENOTED BY THE COLOR CORRESPONDING TO THE COLOR AXIS ON THE RIGHT. THE TITLE OF EACH SUBFIGURE SHOWS WHICH ORDER PARAMETER WAS USED AS INPUT TO EUGENE; THE FIRST THREE COLUMNS EACH USE ONE ORDER PARAMETER AND THE FOURTH COLUMN USES ALL THREE EITHER LOCAL OR GLOBAL PARAMETERS.

The algorithm also detects that State D and E are the same (with high P values) and distinct from A, B, and C. When using angular momentum as the input, the States A and B are found to be similar with low angular momentum, and likewise with States C and E, which have high angular momentum. Interestingly, State D seems to be more alike the behavioral states with high angular momentum, despite having a low M value. For the cohesiveness parameter, behaviors with high cohesion, States B, C, and E, are found to be the same. State D, even though it has a small measure of C , appears closer to B, C, and E, likely due to its clustering behavior on a linear path. Importantly, when using all three global order parameters, EUGENE is able to distinguish that all the behavioral states are different (minimum value for off-diagonal entries > 0.5), which we expect to be true by design. However, these behaviors might be perceived differently if considered from the point of view of an individual rather than a group perspective.

When considering the local interactions, an individual might interpret the polarization of the agents around it much differently than if it could see the entirety of the group. EUGENE is able to detect these differences in the distance matrix for local polarization P^L . Similarly to the global polarization, States A and B are the same as each other, and likewise for States D and E. State

C, on the other hand, is now more similar to D and E. The difference in the local and global polarization parameter is seen in figure 3, where State C distinctly changes from near zero, globally, to one, locally. The switch in similarity is likely because an agent that is milling on a circle, as in State C, would see itself as being more aligned with the individuals nearest to it. In the local angular momentum M^L , the individual would have more difficulty detecting its angular momentum without knowing the entirety of the group, and thus EUGENE finds that all the behavioral states are nearly the same. For local cohesion C^L , the result is similar with the behaviors having pairwise distances close to zero. Finally, when considering all three local order parameters, the distance matrix shows similar structures to the local polarization plot, which means the local interactions are dominated by the P^L parameter.

Notably, EUGENE is able to distinguish between the five states using only low-dimensional descriptors, rather than the high-dimensional state of the system. However, even though we select the previously defined order parameters as input, the results are not identical to the binary distinctions of steady-state high or low values that are usually used to classify these parameters. This suggests that some of the emergent dynamics which occur in the transient may be different or absent when only the

asymptotic behavior of the system is considered. Since animal group behavior may arguably never reach steady-state, EUGENE may give a more realistic classification of these behaviors when applied to biological data. We note that the absolute distances when local parameters are used are low overall, and that the states are generally grouped into two kinds. This suggests that from the perspective of a single agent, these behaviors are not distinct, which is understandable when superimposing a sensing space (circle of radius 1) on any of the agents in figure 2. For recovering these different group-level behaviors, other inputs for EUGENE should be considered.

This work represents a first exploration of causal discovery to understand qualitatively different behaviors using a model-agnostic method for classifying the underlying dynamics driving the system. Here, we show that states known to be different can be distinguished from their time series data. In future work, this method will be used to study data from biological systems, whose dynamics may be governed by observable and unobservable variables, such as physiology and sociality, respectively.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation under grants CMMI-1708622 and SES-1454190, by the Institute for Critical Technology and Applied Science at Virginia Tech, and by the Graduate School at Virginia Tech through a Cunningham Doctoral Scholarship to Amanda Hashimoto.

REFERENCES

- [1] Krause, J., and Ruxton, G. D., 2002. *Living in groups*. Oxford University Press.
- [2] Aranson, I. S., and Tsimring, L. S., 2006. "Patterns and collective behavior in granular media: Theoretical concepts". *Reviews of Modern Physics*, **78**(2), p. 641.
- [3] Elgeti, J., Winkler, R. G., and Gompper, G., 2015. "Physics of microswimmers-single particle motion and collective behavior: a review". *Reports on Progress in Physics*, **78**(5), p. 056601.
- [4] Brambilla, M., Ferrante, E., Birattari, M., and Dorigo, M., 2013. "Swarm robotics: a review from the swarm engineering perspective". *Swarm Intelligence*, **7**(1), pp. 1–41.
- [5] Helbing, D., 2012. "Agent-based modeling". In *Social self-organization*. Springer, pp. 25–70.
- [6] Topaz, C. M., Bertozzi, A. L., and Lewis, M. A., 2006. "A nonlocal continuum model for biological aggregation". *Bulletin of Mathematical Biology*, **68**(7), p. 1601.
- [7] Couzin, I. D., Krause, J., James, R., Ruxton, G. D., and Franks, N. R., 2002. "Collective memory and spatial sorting in animal groups". *Journal of Theoretical Biology*, **218**(1), pp. 1–11.
- [8] Werfel, J., Petersen, K., and Nagpal, R., 2014. "Designing collective behavior in a termite-inspired robot construction team". *Science*, **343**(6172), pp. 754–758.
- [9] Aureli, M., and Porfiri, M., 2010. "Coordination of self-propelled particles through external leadership". *EPL (Europhysics Letters)*, **92**(4), p. 40004.
- [10] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., and Shochet, O., 1995. "Novel type of phase transition in a system of self-driven particles". *Physical Review Letters*, **75**(6), p. 1226.
- [11] Butail, S., Ladu, F., Spinello, D., and Porfiri, M., 2014. "Information flow in animal-robot interactions". *Entropy*, **16**(3), pp. 1315–1330.
- [12] Butail, S., Mwaffo, V., and Porfiri, M., 2016. "Model-free information-theoretic approach to infer leadership in pairs of zebrafish". *Physical Review E*, **93**(4), p. 042411.
- [13] Lord, W. M., Sun, J., Ouellette, N. T., and Boltt, E. M., 2016. "Inference of causal information flow in collective animal behavior". *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, **2**(1), pp. 107–116.
- [14] Dost, F., 2015. "A non-linear causal network of marketing channel system structure". *Journal of Retailing and Consumer Services*, **23**, pp. 49–57.
- [15] McBride, J. C., Zhao, X., Munro, N. B., Jicha, G. A., Schmitt, F. A., Kryscio, R. J., Smith, C. D., and Jiang, Y., 2015. "Sugihara causality analysis of scalp EEG for detection of early Alzheimer's disease". *NeuroImage: Clinical*, **7**, pp. 258–265.
- [16] Kiourktsidis, J. L., and Koulouras, G., 2011. "Adaptive routing provision by using Bayesian inference". *Mobile Ad Hoc Networks: Current Status and Future Trends*, p. 467.
- [17] Oakes, J. M., 2004. "The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology". *Social Science & Medicine*, **58**(10), pp. 1929–1952.
- [18] Jantzen, B. C., 2017. "Dynamical Kinds and their Discovery". *Proceedings of the UAI 2016 Workshop on Causation: Foundation to Application*. arXiv: 1612.04933.
- [19] Jantzen, B. C., 2017. The EUGENE project. Available at <https://github.com/jantzen/eugene> (accessed 9 March 2018).
- [20] Schaub, M. T., Delvenne, J.-C., Lambiotte, R., and Barahona, M., 2018. "Multiscale dynamical embeddings of complex networks". *arXiv preprint arXiv:1804.03733*.
- [21] Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., and Munch, S., 2012. "Detecting causality in complex ecosystems". *Science*, p. 1227079.
- [22] Jantzen, B. C., 2014. "Projection, symmetry, and natural kinds". *Synthese*, **192**(11), pp. 3617–3646.
- [23] Rizzo, M. L., and Székely, G. J., 2016. "Energy distance". *Wiley Interdisciplinary Reviews: Computational Statistics*, **8**(1), Jan., pp. 27–38.