



*Special Issue: Singularity Biology and Beyond
Review (Invited)*

Inferring the roles of individuals in collective systems using information-theoretic measures of influence

Sulimon Sattari¹, Udo S. Basak², M. Mohiuddin^{3,4}, Mikito Toda^{1,5,6}, Tamiki Komatsuzaki^{1,3,7,8,9}

¹ Research Institute for Electronic Science, Hokkaido University, Sapporo, Hokkaido 0010020, Japan

² Pabna University of Science and Technology, Pabna 6600, Bangladesh

³ Graduate School of Chemical Sciences and Engineering, Hokkaido University, Sapporo, Hokkaido 0600812, Japan

⁴ Comilla University, Cumilla 3506, Bangladesh

⁵ Faculty Division of Natural Sciences, Nara Womens University, Nara 6308506, Japan

⁶ Graduate School of Information Science, University of Hyogo, Kobe, Hyogo 6500047, Japan

⁷ Institute for Chemical Reaction Design and Discovery (WPIICReDD), Hokkaido University, Sapporo, Hokkaido 0010021, Japan

⁸ Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Osaka 5650871, Japan

⁹ The Institute of Scientific and Industrial Research, Osaka University, Osaka, Ibaraki 5670047, Japan

Received January 1, 2024; Accepted March 18, 2024;

Released online in J-STAGE as advance publication March 22, 2024

Edited by Hiroko Bannai

In collective systems, influence of individuals can permeate an entire group through indirect interactions complicating any scheme to understand individual roles from observations. A typical approach to understand an individuals influence on another involves consideration of confounding factors, for example, by conditioning on other individuals outside of the pair. This becomes unfeasible in many cases as the number of individuals increases. In this article, we review some of the unforeseen problems that arise in understanding individual influence in a collective such as single cells, as well as some of the recent works which address these issues using tools from information theory.

Key words: causal inference, information theory, transfer entropy, single cells collectives, data-driven approach

◀ Significance ▶

This paper highlights the challenges of inferring causal relationships in systems of many agents. We summarize recent developments in applying information theory and image analysis, which are able to extract information about individual cell-to-cell relationships such as their network structure or the domain of interaction. The developed method is expected to enable us to unveil them for systems of millions or more interacting cells.

Introduction

Groups can achieve complex behaviors that transcend the ability and complexity of the individuals that constitute the group. This phenomenon is evident in various biological systems, including the aggregation and collective migration of cells [1], the flocking of birds [2], and the schooling of fish [3]. For example, colonies of the amoeba *Dictyostelium*

Corresponding author: Tamiki Komatsuzaki, Research Institute for Electronic Science, Hokkaido University, Kita 20 Nishi 10, Kita ku, Sapporo, Hokkaido 001-0020, Japan. ORCID iD: <https://orcid.org/0000-0001-7175-8474>, e-mail: tamiki@es.hokudai.ac.jp

discoideum (Dd) can aggregate into mounds consisting of thousands of cells upon starvation [4]. Understanding such “emergent behaviors” is crucial for many biological applications such as wound healing, cancer growth, and embryo development. The root cause of emergent behavior can only be understood by investigating the interactions between individuals, since individual cells cannot perform such complex actions by themselves. Furthermore, it is expected that all individuals do not contribute equally in developing, especially in triggering, emergent behaviour of the collective. Indeed, there is(are) at least one or a few individuals which has(have) the ability to influence and guide other group members in collectives [5]. When an aggregate of the individuals move, sometimes an individual may trigger the neighboring others to follow itself, which may be interpreted as a leader and the others nearby the leader as the followers. Thus, understanding the collective behavior of systems relies on understanding the influence of individuals and their interactions with each other, as well as on understanding the structure of leader-follower relationships in the system.

To examine the emergence of collective systems and the individuals that they are composed of, experimental data such as sequences of fluorescence images of single cells can be used, from which one can construct time series data that is associated with properties of individuals. Understanding the influence of individuals in the colony then can be achieved through reconstructing the causal relationships between the measured time series variables of interest representing different individuals. One way to examine the existence of such causal relationships is using measures such as cross-correlation [6] or Granger Causality [7]. Alternatively, one can use information-theoretic schemes such as time-delayed mutual information (TDMI) [8], transfer entropy (TE) [9], or intrinsic mutual information (IMI) [10].

In inferring causal relationships between two stochastic variables, one underlying assumption is that all other confounding variables, or indirect influences, can be accounted for. Systems of cells, however, such as developing organs, amoeba colonies, or cancer tumors, typically consist of millions of cells, which make it computationally infeasible to condition on all other possible relevant variables. Even for a single pair of cells, there are other assumptions that complicate inferring causal inference from time series. For example, consider a time series trajectory of the positions of two cells, cell A and cell B. We would like to infer the influence of the position of cell A on that of cell B. Naturally, cell A’s position cannot instantaneously affect that of cell B, so we must consider some time lag between them, that is, we are inferring the influence of the present value of cell A on some future time point of cell B. However, cell B’s future position is also affected by its own present. Thus when inferring the relationship between cell A and cell B with some time lag, the time-lagged measurements of B are themselves confounding factors. The past of A also has an effect on the present value of A, and when there is some bi-directional influence, the past of B would have had an effect on the present or further in past values of A. Thus the past values of each cell, or the time-lagged influences, must be considered as confounding factors in inference using time series data of cells.

In addition to challenges that arise due to confounding factors, collective systems can also be difficult to measure and interpret in their data in sequences of images of cells. For example, the tracking of individual cells can be almost infeasible by human eye for many cells, and automated tracking does not always produce reliable results. When trying to compute a measure of influence between cells in a large cell colony, it may be more prudent to analyze the system using the Eulerian reference frame. In this setting, one may obtain the velocity vector field as it evolves over time in the cell colony, and then compute the measure of influence (e.g., TE) between different regions of the colony. One technique to compute the Eulerian velocity vector fields from sequences of images is called particle image velocimetry (PIV) [11]. Later, we discuss how PIV has been validated to produce reliable velocity vector fields in cell colonies that vary in their level of coherent motion.

This article reviews three central issues in inferring the influence of individuals in collective systems:

1. What challenges arise due to large numbers of interactions and the effect of time-lagged influence in collective systems?
2. Given the above challenges, what measures are promising for understanding the individuals in collective systems?
3. In sequences of images of cell colonies where individual cell tracking is not feasible, how can we obtain reliable velocity data for inferring influence?

The rest of the paper is structured as follows: In *Measuring influence with information theory* we review information-theoretic measures that are used to understand pairwise interactions in collective systems. In the *Models* section, we introduce graphic models that represent different types of time-lagged influence between a leader and a follower, as well as modifications to the Vicsek model that follow the same patterns of influence as the graphs. We then introduce a modified version of the Vicsek model which can simulate larger systems of leaders and followers. In the *Review of results* section, we review three key findings: 1) the number of agents that need to be conditioned on for inferring influence can be reduced greatly by utilizing the transfer entropy with cutoff distance. 2) Modes of information flow can help infer about the time-lagged influence in collective systems, as well as improve estimations of influence given the challenges associated

with indirect influence from the past of a variable. 3) When tracking individual cells becomes impractical, utilizing PIV to measure cell motion can determine the optimal parameters for understanding the velocity dynamics of the system. In the *Conclusion and perspectives* section we summarize our paper and discuss some future directions of this area of research.

Measuring Influence with Information Theory

Information theory plays a pivotal role in advancing our understanding of causality within the context of time series data. Utilizing fundamental concepts from information theory, such as Shannon entropy and mutual information, researchers are empowered to quantitatively assess the extent of uncertainty and information within time series data sets. This methodological approach is instrumental in elucidating causal relationships by facilitating the identification of patterns, dependencies, and the transfer of information between variables over temporal dimensions.

The concept of information theory begins with Shannon entropy or information entropy [12], simply referred to be entropy in this paper. Let X signify the time series characterizing an agent's behavior. Within this context, X encapsulates the dynamic features of an agent, for example, changes in speed, velocity, or the release of chemical signals essential modes of communication between collectively moving agents. Entropy, denoted as $H(X)$, quantifies the degree of uncertainty or amount of information carried in X . The equation for entropy is expressed as:

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i).$$

Here, $P(x_i)$ represents the probability of observing a particular value x_i in the time series X , and n is the total number of discrete values x_i can take. This mathematical formulation encapsulates the fundamental principle of entropy, allowing for a quantitative assessment of the inherent uncertainty and information content embedded in the agent's behavior. In our efforts to infer causality within the interactions of two agents, another relevant metric is conditional entropy. Consider two agents characterized by time series X and Y . The conditional entropy of X given Y is denoted by $H(X|Y)$ and is defined as follows:

$$H(X|Y) = - \sum_{i=1}^n \sum_{j=1}^m P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(y_j)}.$$

Here, $P(x_i, y_j)$ represents the joint probability of observing values x_i and y_j , while $P(y_j)$ is the probability of observing y_j , and m is the total number of discrete values y_j can take. Conditional entropy $H(X|Y)$ measures the uncertainty in the stochastic variable X that persists even after considering the information provided by another stochastic variable Y ; when Y imparts no information about X , then $H(X|Y) = H(X)$.

Stochastic variables X and Y can either be independent or dependent, with dependency implying that one variable holds information about the other. To measure the extent of information one variable provides about the other, mutual information, $I(X; Y)$, can be computed using the formula:

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

While mutual information provides a measure of the statistical dependence between two stochastic variables, its inherent symmetry makes it inadequate for inferring causal direction. However, an alternative is time-delayed mutual information, which introduces a temporal aspect to capture potential causal relationships. Time-delayed mutual information considers the influence of one variable on another with a time lag, offering insights into the directionality of causal connections. The time-delayed mutual information, denoted by $\mathcal{M}_{X \rightarrow Y}$, incorporates a time-lag parameter τ in the variable X (or Y) as follows [13]:

$$\mathcal{M}_{X \rightarrow Y}(\tau) = \sum_{x_t} \sum_{y_{t+\tau}} P(x_t, y_{t+\tau}) \log_2 \frac{P(x_t, y_{t+\tau})}{P(x_t)P(y_{t+\tau})}. \quad (1)$$

From Eq. 1, the asymmetricity of time-delayed mutual information is apparent so that one can evaluate how much the current information of X at time t shares the information with the future information of Y at time $t + \tau$, enabling it to assign the causal direction. It, however, has limitations in that it overlooks shared history and common external driving effects

between the processes [8, 14]. Consequently, this oversight can potentially result in misleading conclusions regarding directed information transfer. To unveil the authentic influence of X 's past on the present of Y , Schreiber [9] highlighted the importance of conditioning on Y 's past state(s) within Equation 1 using conditional mutual information. This leads to the introduction of transfer entropy denoted by $\mathcal{T}_{X \rightarrow Y}$ below, a concept designed to overcome the constraints of time-delayed mutual information. $\mathcal{T}_{X \rightarrow Y}$ for a time lag τ has the following form:

$$\mathcal{T}_{X \rightarrow Y}(\tau) = I(y_{t+\tau}; x_t | y_t) = \sum_{y_{t+\tau}} \sum_{y_t} \sum_{x_t} P(y_{t+\tau}, y_t, x_t) \log_2 \frac{P(y_{t+\tau} | y_t, x_t)}{P(y_{t+\tau} | y_t)}, \quad (2)$$

where $I(\cdot; \cdot | \cdot)$ denotes conditional mutual information, while $P(\cdot | \cdot)$ signifies conditional probability. Transfer Entropy (TE) is one of the most-well used measures in information theory to quantify the information flow from X to Y by measuring the reduction in uncertainty about Y 's future based on knowing X 's present, given Y 's own present, which has overcome the drawback of time-delayed mutual information. However, recently, it was pointed out that TE is not necessarily *intrinsic* to information flow solely from the past X to the future Y because of accounting for Y 's own present in its definition [10, 15]. This additional information, as highlighted by James et al. [10], includes the reduction in uncertainty about Y gained from knowing both X and Y 's present states simultaneously. This extraneous component inflates the desired measure of intrinsic information flow from X to Y , which is solely concerned with the influence X has on Y 's future independent of Y 's own state. Recognizing this limitation, James et al. [10] proposed decomposing TE into two distinct modes: intrinsic and synergistic. The intrinsic mode captures the direct influence of X on Y , while the synergistic mode accounts for the additional information gained through the simultaneous knowledge of both X and Y . This decomposition provides a more nuanced understanding of the information flow between variables (see [10], Figure 1 for a schematic representation of intrinsic, synergistic, and shared information flows). If X and Y are two stochastic processes then the amount of \mathcal{I} from process X to Y is the infimum of $I(X_t; Y_{t+\tau} | \bar{Y}_t)$ taken over all possible conditional distributions $p(\bar{Y}_t | Y_t)$ [10]:

$$\mathcal{I}_{X \rightarrow Y}(\tau) := \inf \left\{ I(X_t; Y_{t+\tau} | \bar{Y}_t) : \sum_{y \in \bar{Y}_t} p(X_t, Y_{t+\tau}, Y_t = y) p(\bar{Y}_t | Y_t = y) \right\} \quad (3)$$

and \mathcal{S} from X to Y is determined as follows:

$$\mathcal{S}_{X \rightarrow Y} = \mathcal{T}_{X \rightarrow Y} - \mathcal{I}_{X \rightarrow Y} \quad (4)$$

where \bar{Y}_t is an auxiliary random variable (that satisfies a Markov chain $Y_t \rightarrow \bar{Y}_t$) to determine the upper bound of $S(X_t; Y_{t+\tau} | Y_t)$, which is a secret-key agreement protocol between X and Y , quantifying those information passes from present X to future Y and this amount of information is equivalent to the information intrinsically coming from X to Y .

Similarly, James et al. [10] also proposed to decompose the time-delayed mutual information from X to Y , $\mathcal{M}_{X \rightarrow Y}$, into the intrinsic information from X to Y ($\mathcal{I}_{X \rightarrow Y}$) and shared information from X to Y ($\sigma_{X \rightarrow Y}$) as follows:

$$\sigma_{X \rightarrow Y} = \mathcal{M}_{X \rightarrow Y} - \mathcal{I}_{X \rightarrow Y} \quad (5)$$

where $\sigma_{X \rightarrow Y}$ estimates the shared information which appears due to the dependency of both X 's and Y 's future dynamics on the same history (current or past states) themselves.

Models

To examine the pitfalls of using TE for inferring influence, models with agents that have different dependencies on their past have been examined. Figure 1 represents four different types of two-agent models, labeled Type A, Type B, Type C, and Type D. A node defined as L or F in Figure 1 represents either a leader or follower agent, respectively. A directed edge from one agent to another signifies that the first agent exerts direct influence on the second agent. An edge from an agent to itself signifies that an agent's future state depends on its present. In all Types, L exerts direct influence on F and F does not exert direct influence on L. In Type A, the future of L and F both do not depend on their present states. In Type B, the future state of F depends on its present state, and the future state of L does not depend on its present state. In Type C, the future state of L depends on its present state, and the future state of F does not depend on its present state. In Type D, the future states of both L and F depend on their current states. In *Review of Results*, we will discuss further how measuring influence from L to F and F to L in these different models can lead to counter-intuitive results using TE or TDMI.

Our Modified Vicsek Model

Over the last few decades, many researchers have made considerable efforts to model the dynamic characteristics of individuals in the collective behavior of multi-agent systems [16–20]. One of the widely studied models in this context is the Vicsek model (VM) due to its simple mathematical design and its effectiveness in explaining the fundamental properties of collective motion [17]. The basic principle of VM is that each individual moves and updates its direction of motion based on the average direction of motion of its neighbors at each time step in presence of some noise. However, the VM does not distinguish between individuals' roles as leaders and followers, as all individuals have the same influence on their neighbors. Therefore, Basak et al. [14, 21] modified the original Vicsek model by integrating asymmetric interactions among individuals, designating individuals with greater influence as leaders and those with lower strengths as followers.

They observe the trajectories of N self-propelled individuals within a square box of length L subject to periodic boundary conditions, where all individuals are randomly positioned and oriented at the initial time i.e., $t = 0$. For simplicity, it is assumed that each individual moves uniformly at a constant speed v_0 and update their positions at each time step $\Delta t = 1$ using the equation:

$$\vec{\mathbf{r}}_i^{t+1} = \vec{\mathbf{r}}_i^t + \vec{\mathbf{v}}_i^t \Delta t, \quad (6)$$

where $\vec{\mathbf{r}}_i^t$ and $\vec{\mathbf{v}}_i^t$ ($i = 1, 2, 3, \dots, N$) denote the position of individual i and its velocity at time t , respectively. The orientation of individual at time t , denoted by $\theta_i(t)$, is updated by

$$\theta_i(t+1) = \langle \theta(t) \rangle_{R, w, \vec{\mathbf{r}}_i^t} + \Delta \theta_i. \quad (7)$$

The term $\langle \theta(t) \rangle_{R, w, \vec{\mathbf{r}}_i^t}$ represents weighted average of neighbors orientations of i , including itself is computed by the formula:

$$\langle \theta(t) \rangle_{R, w, \vec{\mathbf{r}}_i^t} = \arctan \left[\frac{\sum_{j \in \mathcal{N}_i(t)} w_{ji} \sin \theta_j(t)}{\sum_{j \in \mathcal{N}_i(t)} w_{ji} \cos \theta_j(t)} \right]$$

where the sum counts the individuals j which are within the interaction circle of radius R centered at $\vec{\mathbf{r}}_i^t$ at time t ($\mathcal{N}_i(t)$ denotes such set of the neighboring agents for agent i at time t). θ represents an angle between the agent and a reference axis (the horizontal axis) via the origin in a 2D Euclidean space. When the numerator $\sum_{j \in \mathcal{N}_i(t)} w_{ji} \sin \theta_j(t) \leq 0$, π is added to the arctan value ($\theta \in [\pi, 2\pi)$), otherwise not ($\theta \in [0, \pi)$).

The parameter w represents a non-negative asymmetric matrix, where w_{ij} denotes the interaction strength exerted by the individual i on the individual j , and $w_{ij} > w_{ji}$ when i and j are a leader and a follower, respectively. The term $\Delta \theta_i$ corresponds to random noise following a uniform distribution in the interval $[-\eta_0/2, \eta_0/2]$, where the parameter η_0 is a temperature-like parameter. The dynamics of individuals in leader-follower interactions are governed by the following principle: the leader exerts direct influence on the follower, while the follower has no influence on the leader, i.e., $w_{LF} > 0$ and $w_{FL} = 0$, where L and F denote the leader and follower, respectively (in the following sections we will also discuss cases where $w_{FL} > 0$.) Recently, Sattari et al. [15] implemented the Vicsek model including two supplementary features: unequal influence weights to address the asymmetric effect of one individual on another, and individual memory. They have elucidated the influence of individual and group memory in shaping collective behavior based on pairwise information flows estimated by transfer entropy and its decomposed informational modes such as intrinsic information (\mathcal{I}) and synergistic information (\mathcal{S}). A detailed study is carried out for a two-agent system considering four possible interactions with or without memory, as illustrated in Figure 1, where the straight arrow indicates the direct influence of leader (L) on follower (F) (i.e. $w_{ij} > 0$ and $w_{ji} = 0$, where i is the leader and j is the follower) and the curved arrow implies the dependency of the agent on itself. For each type, the influence of the current configuration of an individual on its future is modified. For the cases where the future configuration of an individual L or F does not rely on its current configuration $\theta_L(t)$ or $\theta_F(t)$, respectively, the future configurations are set randomly from the interval $[0, 2\pi]$ in computing the weighted average of the neighbors of individuals, i.e., $\langle \theta(t) \rangle_{R, w, \vec{\mathbf{r}}_i^t}$, to erase any dependency on $\theta_L(t)$ s or $\theta_F(t)$ s current state. In Type A (Figure 1A), neither the leaders nor the followers future dynamics (i.e., $\theta_L(t+1)$ and $\theta_F(t+1)$) depends on its present state $\theta_L(t)$ and $\theta_F(t)$, respectively. Only the follower's future dynamics, i.e., $\theta_F(t+1)$ depends on the present configuration of leader ($\theta_L(t)$). In Type B (Figure 1B), the followers future dynamics, i.e., $\theta_F(t+1)$ has dependency on the present states of both leader and itself, i.e., $\theta_L(t)$ and $\theta_F(t)$. The follower's self-dependency is incorporated in the dynamics by setting $w_{FF} = 1$. Similarly, in Type C (Figure 1C), the future dynamics of both leader and follower, i.e., $\theta_{L(F)}(t+1)$ depend on the current state of leader $\theta_L(t)$ and the leader's self-dependency is included in the dynamics by

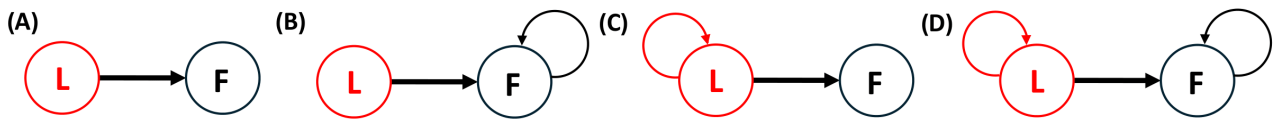


Figure 1 Graph representations for models for different interaction types. (A) Type A. (B) Type B. (C) Type C. (D) Type D. [Reprinted with permission from Ref. [15]. Copyright ©2023, American Association for the Advancement of Science.]

setting $w_{LL} = 1$. In Type D (Figure 1D), the future dynamics of both, i.e., $\theta_L(t+1)$ and $\theta_F(t+1)$ depends on their own present states $\theta_L(t)$ and $\theta_F(t)$, respectively.

Review of Results

Transfer Entropy Dependent on Distance

We introduced information-theoretic methodologies aimed at deducing the domain of interaction by analyzing the trajectories of agents. These methodologies hinge on a parameter referred to as the ‘cutoff distance λ ,’ representing the maximum distance within which interactions between agents are taken into account for Transfer Entropy (TE) estimation [21, 22]. Specifically, for a predetermined cutoff distance λ , the TE between two agents is calculated based on the following criteria: only if the distance separating two agents at time t is less than or equal to λ , their respective time series are employed to estimate probability distributions at that particular time instance [21, 22]. The process involves systematically varying the value of λ , and the TE between agents is subsequently computed as a function of λ [22].

In the context of the problem, we conceptualized the domain of interaction as a circular area with a radius denoted as R , a parameter that is often unknown. This approach allows for a comprehensive exploration of the influence of different cutoff distances on the estimation of Transfer Entropy, providing insights into the dynamics of agent interactions within the specified circular domain. With the modified VM (Eq. 7), it was revealed that the derivative of the average TE with respect to λ , $\frac{d\langle TE \rangle}{d\lambda}$ reaches a minimum in the vicinity of $\lambda = R$. An intuitive explanation is as follows: the most common computation of TE is to build up a joint probability distribution among x_t , y_t and y_{t+1} over all pairs of random variables in a give data. One expects for interacting agents, say, X and Y , that the closer the distance between a pair of the agents is, the more the two agents interact and influence to each other. Suppose that there exists some typical interaction length R beyond which two agents less interact, and below which the two interacts similarly. Then, if one computes TE as a function of λ , TE is expected to stay large until $\lambda \leq R$, but beyond this point, it gradually diminishes (because non-interacting pairs will be taken into account in computing the joint probability distribution) and ultimately approaches 0 as λ increases towards infinity. Consequently, the rate of change in TE with respect to λ tends to be nearly 0 for $\lambda \leq R$, becomes negative for $\lambda > R$, and gradually approaches 0 again as λ tends to infinity. Therefore, we anticipate a kink or minimum in the curve of $\frac{d\langle TE \rangle}{d\lambda}$ around $\lambda = R$ (See our discussion with a mathematical simple model and its derivation [14]). Nevertheless, it was observed that this approach may encounter challenges, particularly in cases where the time series data has a limited length [21]. Due to the practical constraints of obtaining extended trajectories in real experiments, we devised an alternative method centered around the ‘convexity score’ of points at a coarse-grained level. This convexity score approach demonstrated resilience against the fluctuations of $\frac{d\langle TE \rangle}{d\lambda}$ [14, 21].

Take note that the previously mentioned papers [14, 21, 22] presumed uniformity in the interaction domains among all agents in the system, conforming to the interaction rules of the VM. However, this assumption may not always hold true, as the interaction domain can vary among individuals based on factors such as chemical signals, phenotypic variation, and other characteristics. For instance, in Dd colonies, cells exhibit varying levels of excitability. The communication among Dd cells involves a diffused chemical signal, and the radius of influence of a cell is contingent upon the excitability of the cell it is influencing. This steers our efforts towards the objective of employing the TE versus cut-off technique in real-world systems that deviate from ideal VM rules. A modified VM, proposed in [23], deviates from the assumption of agents sharing uniform interaction domains. In this model, an agent influences the motion of another agent only when the former is located inside the interaction domain of the latter, which is similar to models of auditory sensing in bats [24]. Hence, an agent with a larger interaction radius is more susceptible to external influences from other agents in the system than an agent with a shorter interaction radius. Hence, in a system involving two agents with varying interaction domains, the agent with the shorter interaction domain is regarded as the leader, while the second agent is recognized as the follower

(Figure 2).

How can we identify the number of interaction radii in a system? Consider a two-agent system with interaction radii $r_1 = 2$ and $r_2 = 4$ arbitrary units. We examine how the average TE, $\langle \text{TE} \rangle$ changes with the cut-off distance (λ) under moderate noise level ($\eta_0 = \pi$): $\eta_0 = 0$ represents no noise, and $\eta_0 = 2\pi$ corresponds to high noise level so that noise spoils any prediction of interaction regime. Kinks appearing near $\lambda = 2$ and $\lambda = 4$ in the average TE as a function of λ (inset of Figure 3) correspond to the interaction radii of agents 1 and 2 in the simulation model, respectively. The derivative of the average TE with respect to λ exhibits two minima near $\lambda = 2$ and $\lambda = 4$. This indicates that, even under noisy environment, the underlying interaction radii can be estimated by the cut-off distances (λ) associated with the local minima of the TE derivative when angular dependence of agents interaction is not significant and radial dependence is a more dominant factor.

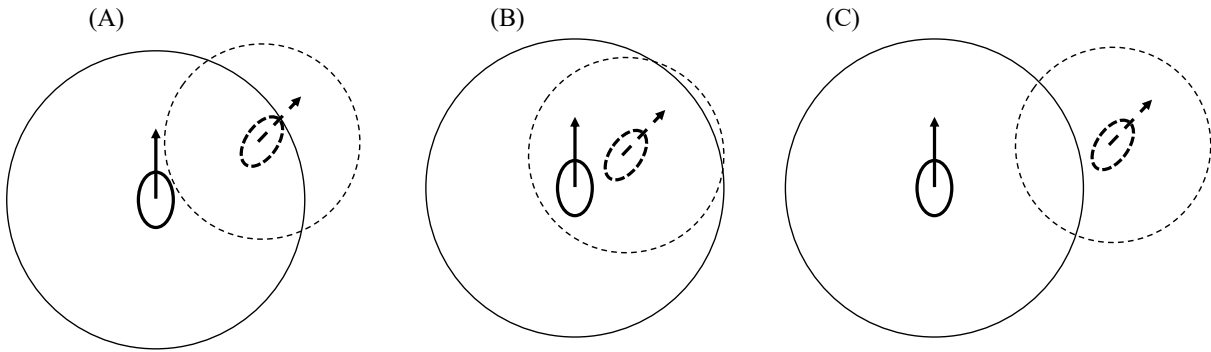


Figure 2 Particle interaction depends on position. (A) the solid particle's motion is influenced by the dotted particle within its interaction domain, but not vice versa. (B) both particles influence each other as they are within each other's interaction domains. (C) particles move independently since they are outside each other's interaction domains. [Reprinted with permission from Ref. [23]. Copyright ©2023 AIP Publishing LLC.]

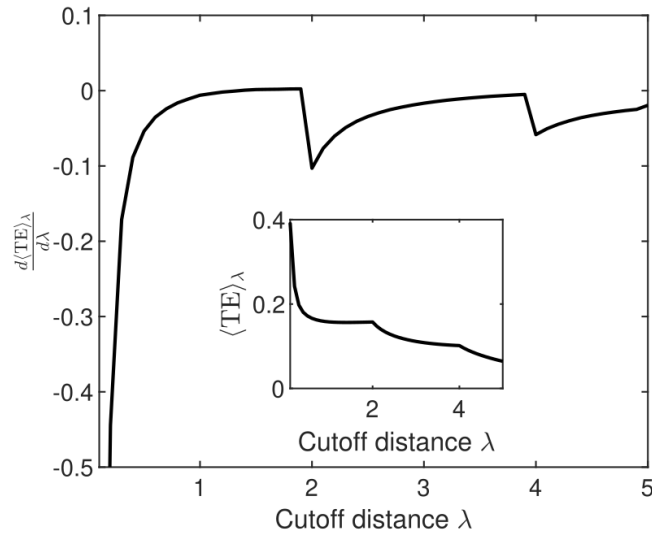


Figure 3 Derivative of average TE with respect to the cutoff distance λ , $\frac{d\langle \text{TE} \rangle_{\lambda}}{d\lambda}$ as a function of λ . Inset: Average TE, $\langle \text{TE} \rangle_{\lambda}$ as a function of cutoff distance λ . [Reprinted with permission from Ref. [23]. Copyright ©2023 AIP Publishing LLC.]

How can one determine the interaction radii of specific agents? Understanding the interaction characteristics between agents is crucial for establishing each agent's interaction domain. In the modified VM described in [23], agent j influences

agent i 's motion only when j is located inside i 's interaction domain. In simpler terms, if the distance between agents i and j is less than the interaction radius R_i of agent i , agent j affects the motion of agent i . Therefore, the average inward TE ($\langle \text{TE}_{j \rightarrow i} \rangle_\lambda$) of agent i provides insights into its interaction domain. Figure 4 illustrates the derivative of the average inward TE of agent 1 concerning the cut-off distance λ , $\frac{d\langle \text{TE}_{2 \rightarrow 1} \rangle_\lambda}{d\lambda}$. The inset in Fig. 4 displays the average inward TE of agent 1. It is evident from Fig. 4 that $\frac{d\langle \text{TE}_{2 \rightarrow 1} \rangle_\lambda}{d\lambda}$ exhibits a minimum at $\lambda = 2$, indicating the interaction radius of agent 1. Similarly, a similar analysis revealed that the derivative of the average inward TE of agent 2, $\frac{d\langle \text{TE}_{1 \rightarrow 2} \rangle_\lambda}{d\lambda}$, also has a minimum. Additionally, it was discovered that within a system marked by a diverse range of agents, the TE versus cut-off technique proves to be a robust method for forecasting the average interaction domain of all agents. This is particularly relevant as the interaction domain of each agent is randomly selected from a Gaussian distribution. This approach not only showcases the versatility of the technique but also underscores its effectiveness in capturing the dynamic interplay and collective behavior within complex systems [23].

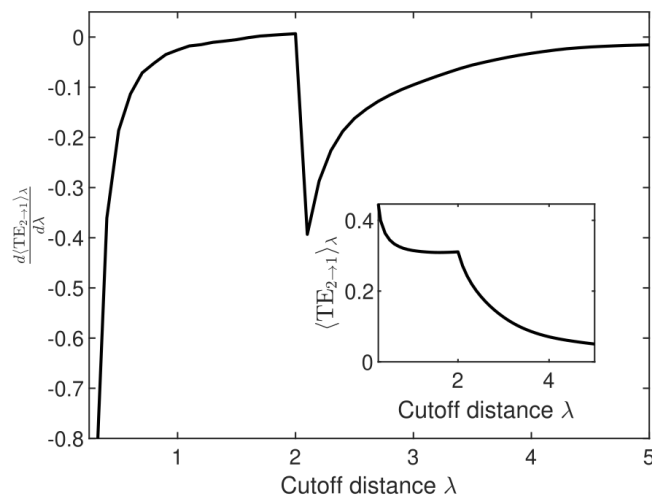


Figure 4 The derivative of the average inward TE for particle 1 with respect to the cutoff distance λ is depicted as a function of λ . In the inset, the average inward TE of particle 1 is illustrated as a function of the cutoff distance λ . [Reprinted with permission from Ref. [23]. Copyright ©2023 AIP Publishing LLC.]

Modes of Information Flow

In developing the theory of TE, Schreiber originally noted that TDMI from the present state of X to the future of Y contains information from the present state of Y itself. Later this was termed as *shared* information (Eq. 5), which is interpreted as the shared history between the two variables. In order to interpret TDMI as information *flow*, an assumption must be made that the present state of Y does not influence its future state. TE is widely thought to relieve this assumption because TE measures the reduction in uncertainty of the future state of Y from knowing the present state of X , while considering the present state of Y as being already known. James et al. [10], however have shown that TE fails to remove the so-called *synergistic* information to predict the future of Y , which can be interpreted as the ability to predict the future of Y by knowing X and Y *simultaneously*, which cannot be attributed to knowing X or Y alone (see [10] for a binary model which exemplifies this type of information flow). Thus intrinsic, shared, and synergistic information flows provide a broader picture of the relationships between agents than TE or TDMI alone.

The models shown in Figure 1 are meant to build upon the notion of shared history in computing information flow measures. In [15] the VM corresponding to each graph in Figure 1A-D was simulated, and the three information measures TE, TDMI, and IMI were computed. It was shown that in Type A, where neither the future of L or F depend on their present, TDMI, TE and IMI all give the same result in terms of measuring how well the present of L can predict the future of F . In other words, shared and synergistic effects are zero. In type B, where the future state of F depends on its present, we found that $\text{TE} > \text{TDMI}$, and their difference is accounted for in the synergistic information. This is because F is contingent on its current state, so having simultaneous knowledge of both the present state of F and L yields greater predictive capability compared to knowing only the present state of either F or L individually. In type C, where the future

Table 1 Summary of information flow modes through interaction type A-D

Interaction Type	Intrinsic	Synergistic	Shared	Dependency on past history
Type A	> 0	0	0	Both L and F do not depend on their past
Type B	> 0	> 0	0	Only F depends on its past
Type C	> 0	0	> 0	Only L depends on its past
Type D	> 0	> 0	> 0	Both L and F depend on their past

of L depends on its present, we found that $TE < TDMI$, and their difference is accounted in the shared information. This is because the current state of L conveys information to the future dynamics of F, and simultaneously, this information is embedded in the future state of F due to the reliance of F's future state on the same historical context, namely, the present of L. Perhaps surprisingly, in type C, TE and TDMI from the *follower* to the *leader* (F to L) are both non-zero, even though no interaction link exists from F to L in the dynamics. The rationale lies in the fact that the historical context of L serves as a hidden variable, imparting information onto both L and F (see Ref. [15] for further discussion). In Type D, where the future of both L and F depend on their present, both the synergistic and shared effects appear, as well as the presence of TE and TDMI from F to L.

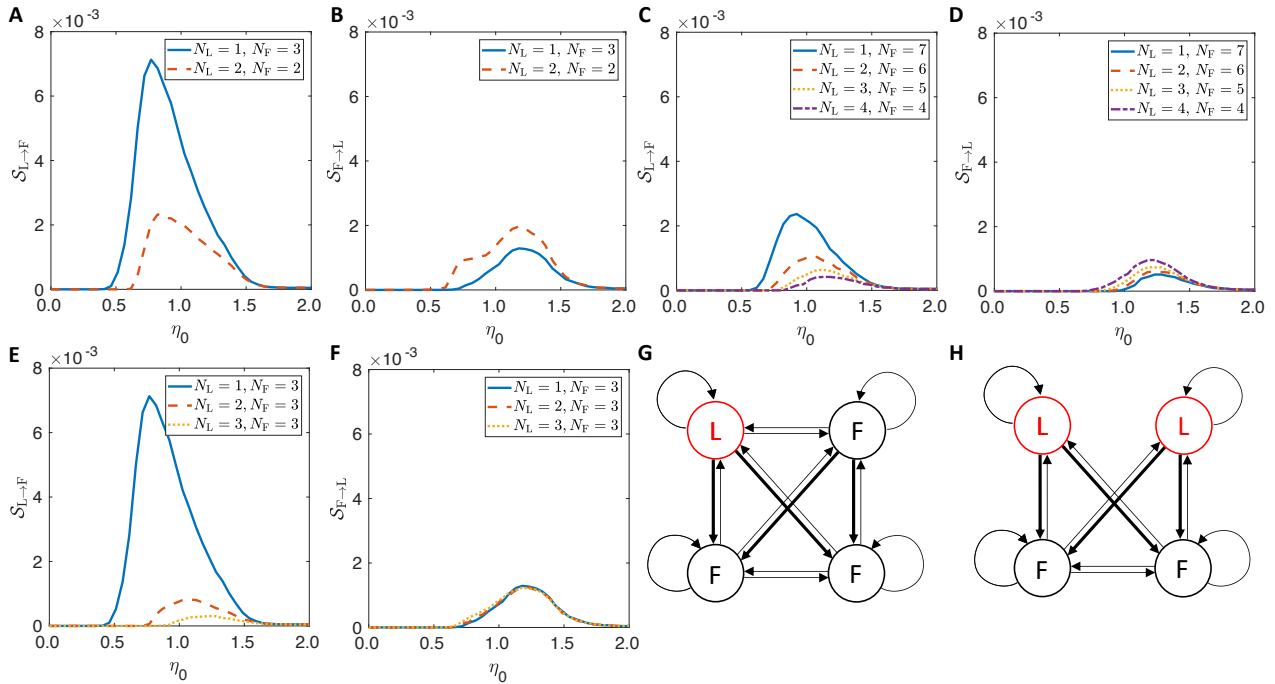


Figure 5 \mathcal{S} as a function of noise level η_0 (in units of π radians) for a model with different numbers of leaders and followers. (A) $\mathcal{S}_{L \rightarrow F}$, and (B) $\mathcal{S}_{F \rightarrow L}$ as a function of η_0 (in units of π radians) for four agents with one leader and three followers (blue) and two leaders and two followers (red). (C) $\mathcal{S}_{L \rightarrow F}$ and (D) $\mathcal{S}_{F \rightarrow L}$ for eight agents with one leader and seven followers (blue), two leaders and six followers (red), three leaders and five followers (yellow), and four leaders and four followers (purple). (E) $\mathcal{S}_{L \rightarrow F}$ and (F) $\mathcal{S}_{F \rightarrow L}$ with three followers and one leader (blue), two leaders (red), and three leaders (yellow). (G) Graph representation of model A, where there is one leader and three followers. (H) Graph representation of model A, where there are two leaders and two followers. [Reprinted with permission from Ref. [15]. Copyright ©2023, American Association for the Advancement of Science.]

It is natural to expect the presence of multiple individuals acting as leaders in the dynamics of collective behavior observed in real-world phenomenon in nature. Hence, in general, multiple leaders may contribute to the dynamics of fol-

lowers in collective behavior, and by observing their interactions, greater insights can be gleaned about collective behavior compared to scenarios where followers are influenced by a single leader. Sattari et al. [15] discussed pairwise information flows as a function of noise (η_0) between leader and follower considering a collective scenario (similar to Figure 5G) as a model in which a single leader interacts with followers, and followers also interact with both the leader and among themselves. They observed the presence of some small bumps in the information flow of well-known transfer entropy $\mathcal{T}_{L(F) \rightarrow F(L)}$ (see Figure 6). It should be noted that such bumps—even if they may be observed in the analysis—have simply been neglected because of being negligibly small contributions. It was found that as $\mathcal{T}_{L(F) \rightarrow F(L)}$ is decomposed to intrinsic information $\mathcal{I}_{L(F) \rightarrow F(L)}$ and the synergistic information $\mathcal{S}_{L(F) \rightarrow F(L)}$, no such bumps exists in the intrinsic information, and the bumps are originated from the synergistic information that diminishes with an increase in the number of follower agents (N_F). This confirms that the bumps emerge because of the simultaneous impact of the current states of both the leader and the follower. The decomposed pairwise information measures from $\mathcal{T}_{L(F) \rightarrow F(L)}$, that is, intrinsic information $\mathcal{I}_{L(F) \rightarrow F(L)}$ and the synergistic information $\mathcal{S}_{L(F) \rightarrow F(L)}$, can tell us that type of the underlying multiple interactions solely from pairwise measures (See [15]).

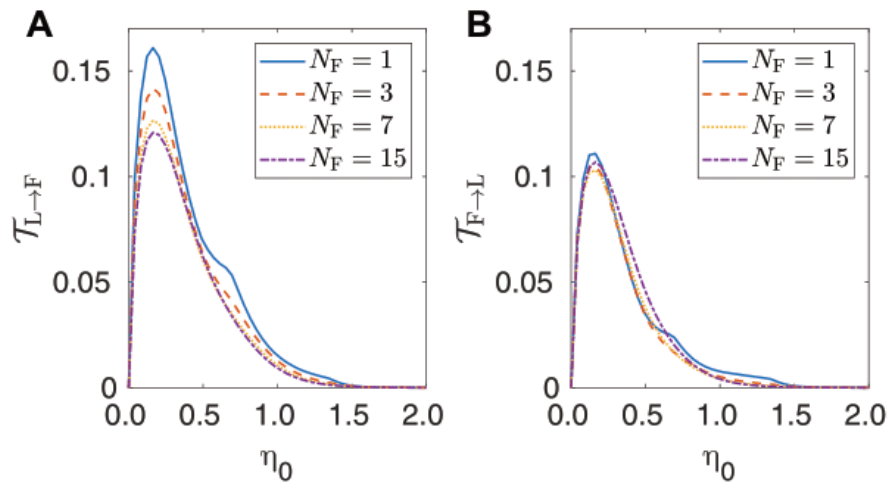


Figure 6 \mathcal{T} as a function of noise level η_0 (in units of π radians) for model A with one leader and different numbers of followers. Here, $N_F=1$ (blue), 3 (red), 7 (yellow), and 15 (purple), where the number of leader is always one. [Reprinted with permission from Ref. [15]. Copyright ©2023, American Association for the Advancement of Science.]

This confirms that the bumps emerge because of the simultaneous impact of the current states of both the leader and the follower. Then, to understand how this structure (bumps) changes as increase of number leader (N_L), the same collective scenario was also studied considering multiple leader agents wherein all agents interact with each other except leader to leader, as demonstrated in Figure- 5H, where the two leaders do not interact with each other, but the two followers interact. In Figure- 5(A-F), the results indicate that $\mathcal{S}_{L(F) \rightarrow F(L)}$ is significantly greater than zero at some noise levels (η_0). This finding confirms that the future of individuals L(F) is inferred by the simultaneous knowledge of their current configurations. The results presented in Figure- 5(A, C, E) demonstrate that $\mathcal{S}_{L \rightarrow F}$ decrease as increase both the number of leader (N_L) as well as the number of follower (N_F). Since each followers configuration is affected by the dynamics of additional individuals, so due to this reason the likelihood of the current configurations of leaders (L) and followers (F) occurring simultaneously also diminishes. The $\mathcal{S}_{L \rightarrow F}$ decrease more with the increase of follower compared to the scenario with the increase leader. Because the followers possess less weight and that's why they cannot reduce the synergistic effect as much as the multiple leaders can. The reduction in $\mathcal{S}_{L \rightarrow F}$ is more pronounced with an increase in followers compared to the scenario with an increase in leaders. This is because followers carry less weight, and consequently, they cannot mitigate the synergistic effect to the same extent as multiple leaders can (see Figure- 5(A, C)). From these results, it also can be concluded that $\mathcal{S}_{L \rightarrow F}$ decrease as the followers weighted indegree increases. But this statement, surprisingly, does not hold for $\mathcal{S}_{F \rightarrow L}$ (see Figure- 5(B, D)). In Figure- 5(E, F), where the results are demonstrated varying N_L and keeping N_F fixed. It is observed that $\mathcal{S}_{F \rightarrow L}$ remains unchanged as increase of N_L because the leaders indegree does not increase despite the increase of N_L and total number of individuals.

Obtaining Velocities for Inference using Particle Image Velocimetry

Particle tracking velocimetry (PTV) is a widely recognized method for obtaining a “Lagrangian descriptor” of velocity dynamics by tracking individual cells. However, manual implementation for numerous cells can be labor-intensive, and automated approaches using supervised or unsupervised techniques may be more problematic especially with high cell density or rapid cell movement. In contrast, an Eulerian descriptor of the velocity field can be achieved through particle image velocimetry (PIV), where flow properties are expressed as a field without identifying individual cells. We conducted a thorough exploration to identify and understand the optimal Particle Image Velocimetry (PIV) parameters specifically tailored for the examination of cell motility. To accomplish this, we employed simulation models to make it possible to compare PIV velocity vector field to the underlying individual agents’ motions [25].

PIV captures the flow patterns of a fluid as it passes a stationary observation point. It determines velocity vectors by identifying a zone in a subsequent frame that best matches a given grid unit cell in the current frame. For that, each image is initially divided into grid unit cells (simply termed as “grids” otherwise noted), with the center of each grid remaining constant across all images. For a grid at time t , the goal is to find the zone in the subsequent image at time $t + 1$ where the content of the grid at time t best aligns. Instead of scanning the entire image for the best match, an interrogation zone, typically a square of size $\mathcal{N} \times \mathcal{N}$ pixels, is defined in the succeeding image. Here, \mathcal{N} is chosen to approximate the largest possible displacement of cells between two successive frames. PIV then establishes a temporary grid (referred to as a window) within the search zone at time $(t + 1)$, matching the size of the original grid. The window is moved within the search zone to find the best match by comparing the orientation of cells within the original grid. The cross-correlation process is employed to find the best match between the displacements of grids in two subsequent images. The correlation function is maximized when the overlap between the original grid at time t and its corresponding best-matched location in the image at time $t + 1$ is largest. This peak in correlation determines the displacement of the original grid between the two images. Finally, PIV draws a vector from the center of the original grid to the center of its best-matched window. PIV vector characterizes the collective motion of cells within a grid, allowing the study of cell movement by examining all PIV vectors in the system.

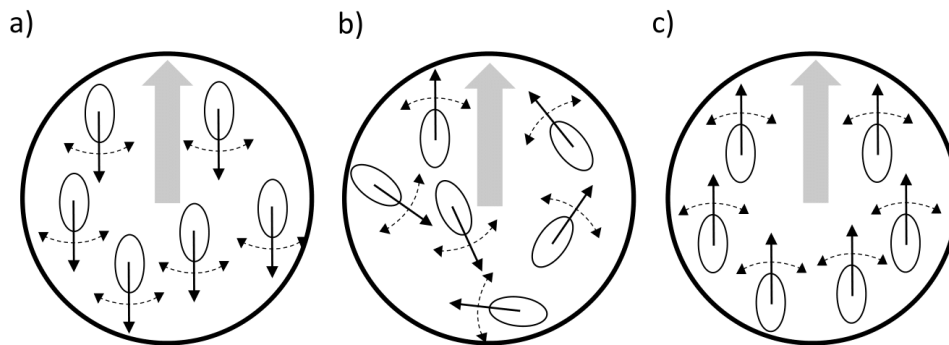


Figure 7 Schematic illustrates particle groups with diverse alignment scores (A^R), with the thick shaded arrow representing a PIV vector defined over a grid, and the ovals with arrows representing the moving particles along with their respective directions of movement. (a) A^R close to -1 due to opposing PIV vector; (b) A^R close to 0 for randomly moving particles; (c) A^R close to 1 when PIV vector aligns with particle movement. [Reprinted with permission from Ref. [25]. Copyright ©2023 Springer Nature Limited.]

In simulating the VM, we knew the positions and velocities of individual particles at each time instance. We utilized these trajectories to assess the performance of PIV in capturing the motion of single agents. The simulation involved generating images of the particles by plotting markers representing agents at each particle location, oriented in the direction of particle motion. Subsequently, the PIV technique was applied to these images, enabling a direct comparison between the velocities computed by PIV and the velocities known from the trajectories, assessed through the alignment score, $A^R(t)$ which has the following form:

$$A^R(t) = \frac{1}{M} \sum_{j=1}^M \frac{1}{N_j^R(t)} \sum_{i \text{ s.t. } |\vec{x}_j - \vec{x}_i| \leq R} \frac{\vec{v}_j^t \cdot \vec{w}_i^t}{|\vec{v}_j^t| |\vec{w}_i^t|}, \quad (8)$$

where \vec{v}_j^t and \vec{w}_i^t refer to the j th PIV vector and the velocity vector of particle i at time t , respectively, where ‘ \cdot ’ denotes the dot product. The second summation in Equation 8 is performed over all particles i situated within a distance R from the initial point of PIV vector j at time t . The alignment score, A^R , ranges from -1 to 1. A score of 1 indicates perfect alignment between particles within the circle and PIV vectors, 0 implies no tendency for alignment, and -1 signifies particles moving precisely opposite to PIV vectors (refer to Figure 7 for a visual representation). Hence, an A^R value greater than zero suggests some average alignment between PIV vectors and particle motion.

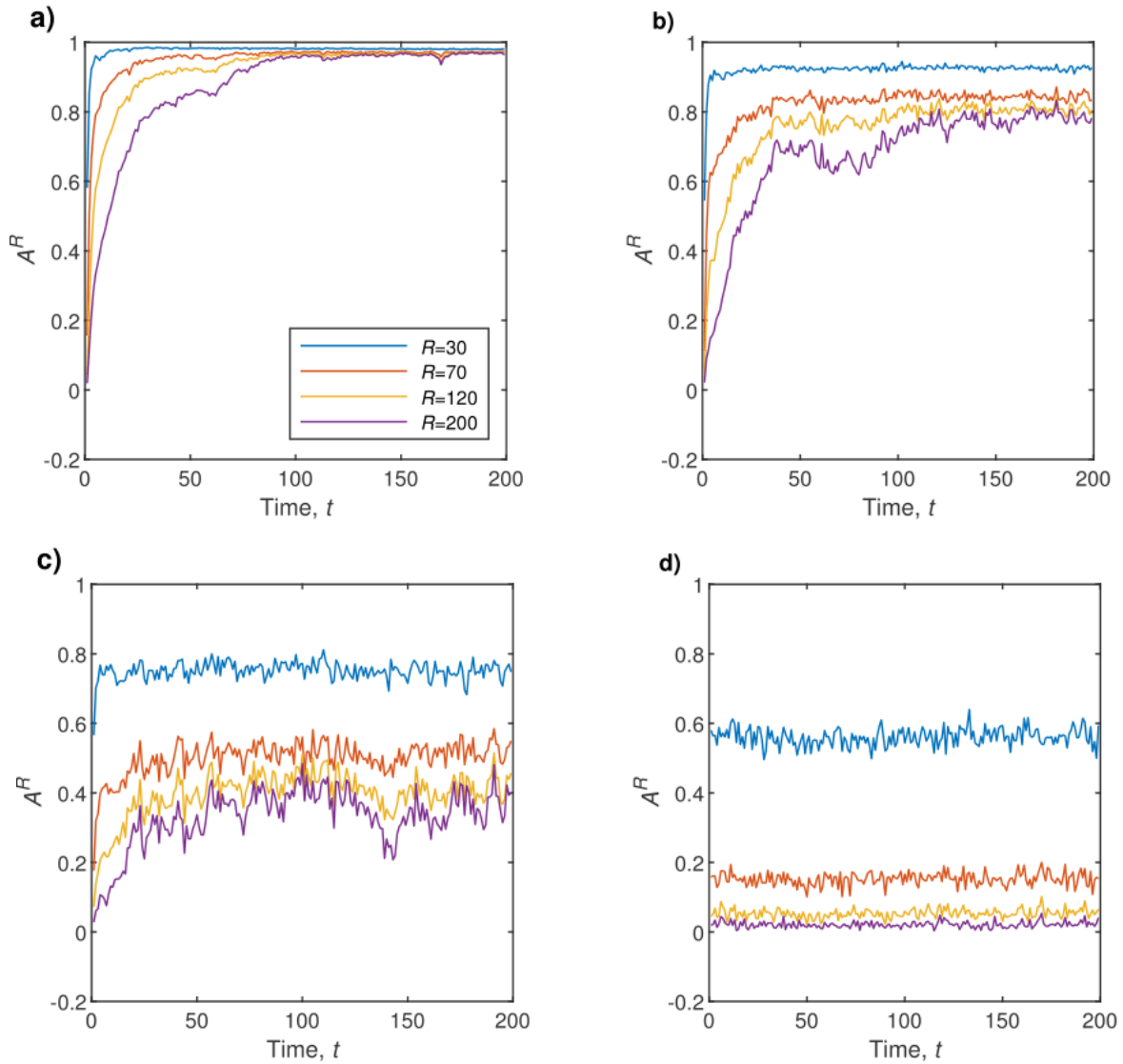


Figure 8 The alignment score A^R is plotted against time t for various radius R values under different noise conditions: (a) $\eta_0 = \frac{\pi}{6}$, (b) $\eta_0 = \frac{3\pi}{6}$, (c) $\eta_0 = \pi$, and (d) $\eta_0 = \frac{11\pi}{6}$. [Reprinted with permission from Ref. [25]. Copyright ©2023 Springer Nature Limited.]

The alignment score, defined in Eq. 8 was utilized to assess the extent to which PIV vectors reflect the motion of nearby

particles. The alignment score A^R is depicted in Figure 8 as a function of time t at different noise levels: (a) $\eta_0 = \frac{\pi}{6}$, (b) $\eta_0 = \frac{3\pi}{6}$, (c) $\eta_0 = \pi$, and (d) $\eta_0 = \frac{11\pi}{6}$. At low noise levels ($\eta_0 = \frac{\pi}{6}$), A^R rapidly converges to its maximum value of 1 for small R values. The value of A^R declines during the initial increase in radius R . This is attributed to the time required for the global coherence of particles to form, as dictated by the interaction rules of the VM, which do not allow for instantaneous global interactions among particles. This delay in coordinated agent movement is a common natural phenomenon, reflecting the time required for information cascade among group members. Once the system achieves global coherence, A^R becomes less dependent on R . As noise η_0 increases, the value of A^R decreases (Figure 8b-d). At high noise levels ($\eta_0 = \frac{11\pi}{6}$), A^R approaches zero as noise dominates the system, and particles move randomly (Figure 8d). Interestingly, even at very high noise levels, A^R remains relatively high (~ 0.5) for $R = 30$. Thus, under high noise conditions, it is crucial to select R such that the number of particles within the radius R is not excessively large, and the grid size for evaluating PIV vectors appropriately captures the underlying particle motility, ensuring accurate characterization of PIV using the quantity A^R .

The exploration of identifying the optimal values for both R and the grid size is also addressed. Determining the optimal R and grid size in PIV estimation under high-noise conditions poses a challenge. PIV measures the average particle displacement within a grid of size $\gamma \times \gamma$ ($\gamma \in \mathbb{R}$). In situations of high noise, where particles move randomly, PIV encounters difficulty in precisely measuring velocities, particularly when multiple particles inhabit a PIV grid. Theoretically, if a PIV grid contains only one particle, PIV should be effective in characterizing the motion of that particle even in high-noise scenarios. Let $\gamma_0(N)$ denote the typical size of a PIV grid for a given particle count N , ensuring an average occupancy of only one particle. It is defined as:

$$\gamma_0(N) = \sqrt{\frac{\Gamma \times \Gamma}{N}},$$

where Γ represents the length of the image data. Correspondingly, assuming a uniform distribution, $R_0(N) = \gamma_0(N)/\sqrt{\pi}$ represents the radius size of a circle centered at each pixel for a given particle count N , ensuring an average occupancy of only one particle.

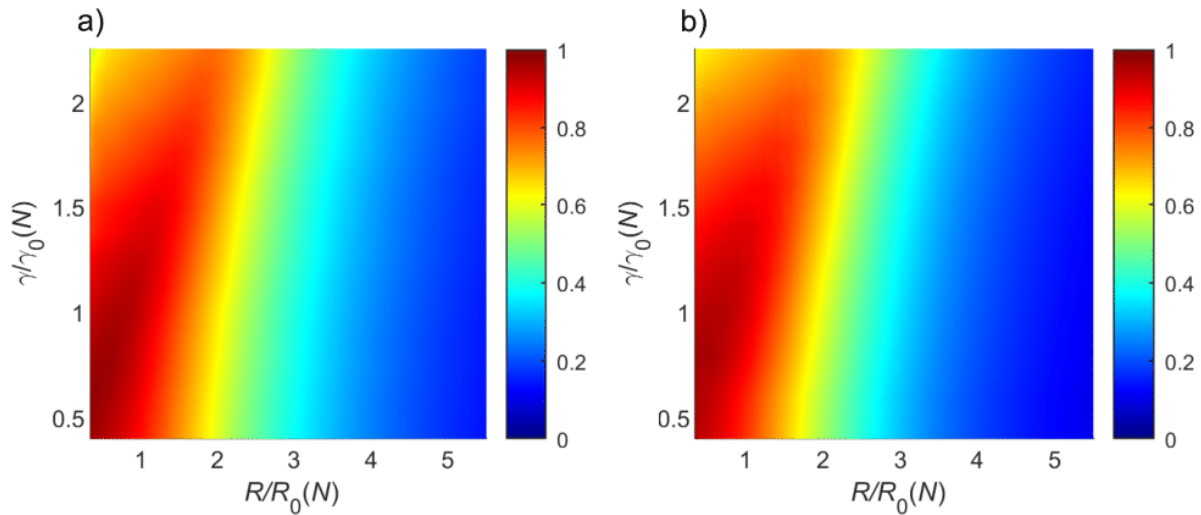


Figure 9 The landscape of A^R at $\eta_0 = \frac{11\pi}{6}$ (high noise) is depicted in relation to the normalized PIV grid size, $\frac{\gamma}{\gamma_0(N)}$, and normalized radius, $\frac{R}{R_0(N)}$, for a particle count of (a) $N = 100$ and (b) $N = 300$. [Reprinted with permission from Ref. [25]. Copyright ©2023 Springer Nature Limited.]

Figure 9 illustrates the landscape of A^R concerning the normalized PIV grid size, $\frac{\gamma}{\gamma_0(N)}$, and the normalized radius, $\frac{R}{R_0(N)}$, for (a) $N = 100$ and (b) $N = 300$ in a high-noise scenario ($\eta_0 = \frac{11\pi}{6}$). As $\frac{\gamma}{\gamma_0(N)}$ and $\frac{R}{R_0(N)}$ approach unity, A^R values tend to converge towards unity. This convergence stems from the fact that, on average, the PIV grid and A^R

estimation involve a single particle within a small circle. Conversely, $\frac{\gamma}{\gamma_0(N)}$ values exceeding 1 imply that, on average, a PIV grid may contain multiple randomly moving particles, leading to decreased A^R values for $\frac{\gamma}{\gamma_0(N)} > 1$ in noise-intensive scenarios where particles lack coherence. Similarly, an $\frac{R}{R_0(N)}$ value exceeding 1 may indicate consideration of multiple randomly moving particles, resulting in low A^R values. This trend in A^R values holds universally for different particle counts N .

Conclusion and Perspectives

The information-theoretic and image analysis techniques highlighted in this article are applicable to understanding the structure of a wide variety of many-body systems from observations alone. For example, in observing a swarm of un-manned air vehicles (UAVs) or animals, one may seek knowledge about which individuals are most influential or how their influence propagates through the system. In that case, we propose analysis based on modes of information flow or PIV can help extract important insights. More work needs to be done, however, before one can apply these techniques for systems of homogeneous quality, for example, in the human body where the interactions between parts of the system occur in different ways and in different size and time scales.

Specifically for the analysis of large colonies of cells, recent developments in live cell imaging can capture both micro- and macro-scale information simultaneously, achieving individual cell resolution while also providing a viewing window at the cell colony scale. By noting the ability to extract information about multiple scales simultaneously, these images have been termed as "trans-scale" imaging, and can provide crucial insights for addressing the question of how an individual cell influences the collective.

Inferring the relationships between individual cells from these huge data sets requires new mathematical tools due challenges specific to inference in systems of large numbers of interacting agents. We have the effects of individual and shared history on inference measurement, and propose that the decomposition of transfer entropy and time-delayed mutual information into intrinsic, shared, and synergistic information flows can help to elucidate the effects of history.

The interaction domain of individuals is one crucial parameter for inferring influence in order to reduce the number of candidate agents that can influence a given agent. We have shown that transfer entropy dependent on distance can determine the interaction domain of agents in a collective even when the sizes of domains of individuals can be heterogeneous. We also have addressed tools that can extract information about the network structure and role of cell memory in collectives using modes of information flow. It was demonstrated using simple two-agent models that shared history determines the existence of shared or synergistic information flows. In systems of more than two agents, it was shown that synergistic information flow can be used to distinguish systems that have follower-to-follower interaction in addition to leader-to-follower interaction from those which only have leader-to-follower interaction.

In addition to the information-theoretic techniques for extracting information from trajectories, we have analyzed how PIV can extract relevant velocity information from cell images even when individual cell tracking is not possible. This can be applied in conjunction with information theory in future works. For example, using PIV, one can extract velocity fields of regions, and then use transfer entropy dependent on distance to infer the domain of influence of cell motion or the network structure of interaction between different regions of cells. This can help understand the emergent properties of systems of cells that cannot be addressed by studying individual cells alone.

Understanding emergent behavior is a task whose benefit will extend beyond cell colony dynamics. Emergent behavior has puzzled scientists in many disparate fields such as artificial intelligence [26], social science [27], robotics [28], and ecology [29]. Since information-theoretic methods do not confine their applicability to a single model, the works summarized here may be applied to understanding how individual influence plays a role in emergent behavior in different fields of research.

Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

Code Availability

The codes are available from the corresponding author upon request.

Acknowledgments

We thank Prof. Kazuki Horikawa for his continuous fruitful discussions. This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas Singularity Biology (No. 8007) (Grant No. 18H05413), MEXT, the research program of Five star Alliance in NJRC Matter and Dev (Grant No. 20191062-01), the JSPS (Grant Nos. 25287105 and 25650044, T.K.), and the JST/CREST (Grant No. JPMJCR1662, T.K.). U.S.B is supported by the University Grants Commission (UGC) of Bangladesh research grant (grant no. Physical Science-72-2021) and the Pabna University of Science and Technology research fund. M.T. is supported by the Research Program of “Dynamic Alliance for Open Innovation Bridging Human, Environment and Materials” in “Network Joint Research Center for Materials and Devices” and a Grant-in-Aid for Scientific Research (C) (No. 22654047, No. 25610105, No. 19K03653, and No. 23K03265) from JSPS. The composition of our A02-02 group and our collaborators in the Singularity Biology project are displayed in Table 2.

Table 2 Our A02-2 group composition and collaborators in the Singularity Biology

Principal Investigator:	Tamiki Komatsuzaki (Hokkaido University)
Co-Investigator (CI):	Atsuyoshi Nakamura (Hokkaido University) Shunsuke Ono (Tokyo Institute of Technology)
Collaborating Researcher (CR)	Ichigaku Takigawa (Kyoto University)
Collaborators in the Singularity Biology	
A01-1 PI	Tomonobu Watanabe (RIKEN/Hiroshima University)
A01-2 PI	Takeharu Nagai (Osaka University)
A03-2 PI	Kazuki Horikawa (Tokushima University)
Publicly offered research A03	Satoshi Sawai (Tokyo University)
Publicly offered research A03	Masakazu Sonoshita (Hokkaido University)

Conflicts of Interest

All authors declare that they have no conflict of interest.

Author Contributions

T.K., S.S. designed the project over discussions. U.S.B. and S.S. conducted analyses and experiments. All authors wrote and confirmed the contents of the manuscript.

References

- [1] Weijer, C. J. Collective cell migration in development. *J. Cell Sci.* 122, 3215–3223 (2009). <https://doi.org/10.1242/jcs.036517>
- [2] Flack, A., Nagy, M., Fiedler, W., Couzin, I. D., Wikelski, M. From local collective behavior to global migratory patterns in white storks. *Science*. 360, 911–914 (2018). <https://doi.org/10.1126/science.aap7781>
- [3] Crosato, E., Jiang, L., Lecheval, V., Lizier, J. T., Wang, X. R., Tichit, P., et al. Informative and misinformative interactions in a school of fish. *Swarm Intell.* 12, 283–305 (2018). <https://doi.org/10.1007/s11721-018-0157-x>
- [4] Ichimura, T., Kakizuka, T., Horikawa, K., Seiriki, K., Kasai, A., Hashimoto, H., et al. Exploring rare cellular activity in more than one million cells by a transscale scope. *Sci. Rep.* 11, 16539 (2021). <https://doi.org/10.1038/s41598-021-95930-7>
- [5] Mwaffo, V., Butail, S., Porfiri, M. Analysis of pairwise interactions in a maximum likelihood sense to identify leaders in a group. *Front. Robot. AI* 4, 35 (2017). <https://doi.org/10.3389/frobt.2017.00035>
- [6] Butail, S., Mwaffo, V., Porfiri, M. Model-free information-theoretic approach to infer leadership in pairs of zebrafish. *Phys. Rev. E* 93, 042411 (2016). <https://doi.org/10.1103/PhysRevE.93.042411>
- [7] Seth, A. Granger causality. *Scholarpedia* 2, 1667 (2007). <https://doi.org/10.4249/scholarpedia.1667>
- [8] Bossomaier, T., Barnett, L., Harré, M., Lizier, J. T., Bossomaier, T., Barnett, L., et al. An Introduction to Transfer

- Entropy (Springer, Hoboken, New Jersey, 2016).
- [9] Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* 85, 461 (2000). <https://doi.org/10.1103/PhysRevLett.85.461>
- [10] James, R. G., Ayala, B. D. M., Zakirov, B., Crutchfield, J. P. Modes of information flow. *arXiv* (2018). <https://doi.org/10.48550/arXiv.1808.06723>
- [11] Thielicke, W. Sonntag, R. Particle Image Velocimetry for MATLAB: Accuracy and enhanced algorithms in PIVlab. *J. Open Res. Softw.* 9, 12 (2021). <https://doi.org/10.5334/jors.334>
- [12] Cover, T. M. Elements of information theory (John Wiley & Sons, Cham, Switzerland, 1999).
- [13] Jeong, J., Gore, J. C., Peterson, B. S. Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clin. Neurophysiol.* 112, 827–835 (2001). [https://doi.org/10.1016/s1388-2457\(01\)00513-2](https://doi.org/10.1016/s1388-2457(01)00513-2)
- [14] Basak, U. S., Sattari, S., Hossain, M., Horikawa, K., Komatsuzaki, T. Transfer entropy dependent on distance among agents in quantifying leader-follower relationships. *Biophys. Physicobiol.* 18, 131–144 (2021). <https://doi.org/10.2142/biophysico.bppb-v18.015>
- [15] Sattari, S., Basak, U. S., James, R. G., Perrin, L. W., Crutchfield, J. P., Komatsuzaki, T. Modes of information flow in collective cohesion. *Sci. Adv.* 8, eabj1720 (2022). <https://doi.org/10.1126/sciadv.abj1720>
- [16] Reynolds, C. W. Flocks, herds and schools: A distributed behavioral model. *Comput. Graph (ACM)*. 21, 25–34 (1987). <https://doi.org/10.1145/37402.37406>
- [17] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* 75, 1226 (1995). <https://doi.org/10.1103/PhysRevLett.75.1226>
- [18] Levine, H., Rappel, W.-J., Cohen, I. Self-organization in systems of self-propelled particles. *Phys. Rev. E.* 63, 017101 (2000). <https://doi.org/10.1103/PhysRevE.63.017101>
- [19] Chaté, H., Ginelli, F., Grégoire, G., Raynaud, F. Collective motion of self-propelled particles interacting without cohesion. *Phys. Rev. E.* 77, 046113 (2008). <https://doi.org/10.1103/PhysRevE.77.046113>
- [20] Fily, Y. Marchetti, M. C. Athermal phase separation of self-propelled particles with no alignment. *Phys. Rev. Lett.* 108, 235702 (2012). <https://doi.org/10.1103/PhysRevLett.108.235702>
- [21] Basak, U. S., Sattari, S., Hossain, M. M., Horikawa, K., Komatsuzaki, T. An information-theoretic approach to infer the underlying interaction domain among elements from finite length trajectories in a noisy environment. *J. Chem. Phys.* 154, 034901 (2021). <https://doi.org/10.1063/5.0034467>
- [22] Basak, U. S., Sattari, S., Horikawa, K., Komatsuzaki, T. Inferring domain of interactions among particles from ensemble of trajectories. *Phys. Rev. E.* 102, 012404 (2020). <https://doi.org/10.1103/PhysRevE.102.012404>
- [23] Basak, U. S., Islam, M. E., Sattari, S. Inferring interaction domains of collectively moving agents with varying radius of influence. *AIP Adv.* 13, 035312 (2023). <https://doi.org/10.1063/5.0135053>
- [24] Roy, S. Lemus, J. How Does the Fusion of Sensory Information From Audition and Vision Impact Collective Behavior? *Front. Appl. Math. Stat.* 7, 758711 (2021). <https://doi.org/10.3389/fams.2021.758711>
- [25] Basak, U. S., Sattari, S., Hossain, M. M., Horikawa, K., Toda, M., Komatsuzaki, T. Comparison of particle image velocimetry and the underlying agents dynamics in collectively moving self propelled particles. *Sci. Rep.* 13, 12566 (2023). <https://doi.org/10.1038/s41598-023-39635-z>
- [26] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., et al. Emergent abilities of large language models. *arXiv* (2022). <https://doi.org/10.48550/arXiv.2206.07682>
- [27] Huang, J. P. Experimental econophysics: Complexity, self-organization, and emergent properties. *Phys. Rep.* 564, 1–55 (2015). <https://doi.org/10.1016/j.physrep.2014.11.005>
- [28] Liu, Q., He, M., Xu, D., Ding, N., Wang, Y. A mechanism for recognizing and suppressing the emergent behavior of UAV swarm. *Math. Probl. Eng.* 2018, 6734923 (2018). <https://doi.org/10.1155/2018/6734923>
- [29] Couzin, I. D. Krause, J. Self-Organization and Collective Behavior in Vertebrates. *Adv. Study Behav.* 32, 1–75 (2003). [https://doi.org/10.1016/S0065-3454\(03\)01001-5](https://doi.org/10.1016/S0065-3454(03)01001-5)