Hokkaido University

Doctoral Dissertation

Study on Identification of Leader and Follower Agents and its Interaction Domain from Trajectories in a Collectively Moving Colony

（協同的コロニーの軌道データによる先導·従エージェントとそれらの相互作用領域の同定に関する研究）

---

by

Udoy Sankar Basak

Graduate School of Life Science
Hokkaido University
Japan

March, 2021

Declaration

I hereby declare that the matter embodied in this thesis entitled "Study on Identification of Leader and Follower Agents and its Interaction Domain from Trajectories in a Collectively Moving Colony" is the result of investigations carried out by me under the supervision of **Prof. Tamiki Komatsuzaki** at Molecule & Life Nonlinear Sciences Laboratory (MLNS), Hokkaido University, Japan, and it has not been submitted elsewhere for the award of any degree or diploma.

In keeping with the general practice of reporting scientific observations, due acknowledgment has been made whenever the work described has been based on the findings of the other investigators. Any omission that might have occurred by oversight or error of judgment is regretted.

Udoy Sankar Basak

February 16, 2021

Certificate

I hereby certify that the work described in this thesis entitled "Study on Identification of Leader and Follower Agents and its Interaction Domain from Trajectories in a Collectively Moving Colony" has been carried out by **Udoy Sankar Basak**, under my supervision at Molecule & Life Nonlinear Sciences Laboratory (MLNS), Hokkaido University, Japan.

Prof. Tamiki Komatsuzaki

February 16, 2021

This thesis is dedicated to
my beloved mother **Dipali Rani Basak** and father **Ashutosh Basak**,
my son **Dibban Basak,** and my lovely wife **Dipika Basak**.

Abstract

Systems composed of locally interacting particles or agents such as birds, fish, and cells show spontaneous, spatiotemporal, collective behaviors as a whole. Exploration of the underlying mechanisms and principles of such coordination of agents has been made in experimental, and theoretical studies. In many systems, it has been proved that the presence of influential individuals, known as 'leaders', are responsible for the collective motion of agents. These leaders control the movement of the whole community.

In some cases, e.g., fish shoal, MDCK epithelial cells, the relative position of the agents helps to identify the leader. But in many cases where agents do not move in the same direction, e.g., Dictyostelium Discoideum, the relative position does not help to identify leaders. Hence identifying leader agents in a collectively moving community is a perplexing work.

Since the follower's movement is regulated by the leader agents, hence there is a correlation between the movement of leader and follower agents at a certain time lag. Consequently, cross-correlation has been used in identifying leader and follower agents. But its performance is questionable in non-linear systems. Transfer entropy, an information-theoretic measure, is capable of capturing non-linear interaction between agents, hence it has been used to identify leader agents.

The effectiveness of TE in identifying leader agents has been tested in a Dictyostelium discoideum colony. It has been found that the result obtained using TE is almost identical to the expert's result.

The Vicsek model (VM) often studied as a metaphor for collectively moving animals is employed. A modified version of the VM has helped us to investigate the classification performance of CC and TE. It has been found that TE outperforms CC. Different model parameters have been varied to check their effect on classification scores.

An information-theoretic scheme is proposed to estimate the underlying domain of

interactions and the timescale of the interactions for many-particle systems. Based on ensemble data of trajectories of the model system, it is shown that using the interaction domain significantly improves the performance of classification of leaders and followers compared to the approach without utilizing knowledge of the domain. Given an interaction timescale estimated from an ensemble of trajectories, the first derivative of transfer entropy averaged over the ensemble with respect to the cut-off distance is presented to serve as an indicator to infer the interaction domain. It is shown transfer entropy is superior to infer the interaction radius compared to cross-correlation, hence resulting in a higher performance to infer leader-follower relationship. Effects of noise size exerted from the environment, and the ratio of the numbers of leader and follower on the classification performance is also discussed.

Later it was found that the 'minimum derivative' scheme is dependent on how transfer entropy can be estimated so that it takes into account enough statistics of interacting particles, and positions and numbers of the minimum of the derivative of average transfer entropy along with the cutoff distance λ may also be subject to the extent of external noise and time length of trajectories. The author has scrutinized how the prediction performance in capturing the underlying interaction domain depends on the size of noise and time length of the trajectory data. An alternative scheme has been proposed which is expected to be stable against noises and time length, that relies on the degree of convexity at coarse-grained scale in the derivative of average transfer entropy along with the cutoff distance, and time variance of underlying interaction radius of particles.

Key words: Collective motion, Causality, Leader-follower, Cross-correlation, Transfer entropy, Vicsek model, Interaction radius, cAMP, PIV.

Acknowledgements

This thesis would not have been accomplished without the encouragement, assistance, and continuous support of many wonderful people. My heartfelt gratitude to all of them.

First and foremost, I would like to gratefully thank my advisor Professor Tamiki Komatsuzaki, for his excellent guidance, encouragement, patience, motivation, and cooperation during my whole period of study at Hokkaido University. His wonderful mentorship provided me with a great opportunity to be introduced to the field of multi-disciplinary research. I am very much thankful for his contributions of incredible ideas and funding to make my Ph.D. research productive and interesting. Thank you, Tamiki sensei, for everything you have done for me.

I belonged to a cheerful group of professor Komatsuzaki, having a helpful and motivational attitude which is important for effective research. I would like to thank all the fellow lab members of this group who were very friendly and helpful. This group has contributed immensely to my daily life and my professional time at Hokkaido University.

Especially, I would like to express my sincere gratitude to Dr. Sulimon Sattari, for supporting me during the past three years with valuable scientific advice, many insightful discussions, and suggestions. He introduced me to the concepts of the Vicsek model on which this thesis is based. We worked together on the analysis of self-propelled particle motion data to accomplish the thesis. He was my primary resource to get the answer to my science questions regarding my Ph.D. research, and he has done a major role to develop methodologies and code to obtain the results related to this thesis. I will forever be thankful to you Sulimon, for your ideas, time, and patient that have been great contributors to accomplish the thesis.

I would like to thank Associate Professor Hiroshi Teramoto for his valuable scientific advice, fruitful discussions and useful guidance. He helped me to understand the mathematical concept of intrinsic, shared and synergistic information flow.

I would like to thank Associate Professor Hiroshi Teramoto for his valuable scientific advice, fruitful discussions, and useful guidance. He helped me to understand the mathematical concept of intrinsic, shared, and synergistic information flow.

I thank Prof. Koji Tabata, a very nice and generous person who helped me to understand some codes written in Python. I also thank Dr. Khalifa Mohammad Helal, a genuinely nice guy who was always very supportive, friendly as an elder brother in my daily life in Sapporo. I thank Dr. Jean-Emmanuel Clement, a helpful, friendly, nice, and smart French guy, who was always encouraging and cheerful about everything. I am very much happy to have worked with this interesting group in a wonderful environment. I would like to thank Asst. Prof. James N. Taylor, Asst. Prof. Yuta Mizuno, and other members to make this an interesting research group. I thank the past members of our group. My special gratitude goes to the former assistant professor of our lab Dr. Sosuke Ito, who helped me to explore the diversity of information theory. My special thanks to Hiroko Yamada, who helped me a lot when I first arrived in Sapporo.

This work is a collaborative study between our group and another group from the Department of Optical Imaging, the Institute of Biomedical Sciences, Tokushima University Graduate School. I would like to thank Professor Kazuki Horikawa for his valuable comments and discussions.

It would be inappropriate if I omit to acknowledge Maiko Ishida and Matsue Koida, two extremely helpful secretaries in our lab. I thank them for being so helpful and friendly both in my daily family life and academic time. I also thank all other staff for their helping attitude.

It is my pleasure to acknowledge my friends and colleagues who have generously offered their help and friendship and provided some much-needed humor and entertainment during the years of my Ph.D. period.

I thankfully acknowledge all the funding sources that helped to accomplish my Ph.D. research. In the framework of Japanese Government-financed scholarship under the International Graduate Program (IGP), Hokkaido University, The Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT) provided funding for this study. I wish to acknowledge MEXT for the financial support during my stay in Japan.

I would also like to express deep gratitude to Pabna University of Science and Technology, Bangladesh, for allowing me a study leave for the whole period of my study at

Hokkaido University, Japan.

Most importantly, I would like to express my deep gratitude towards my family, my biggest source of strength, for their love and support. The blessings of my father Ashutosh Basak, who is spending whole of his life to support me and my family with great inspiration in all my pursuits, have made a great contribution to reach in this stage of my life. I am also very much thankful to my mother Dipali Basak for her unconditional love and pray. I would not have been where I am today without her encouragement, caring, and pray. I thank my grandparents, my parents-in-law, my maternal uncle, and my friends and colleagues for their constant support, pray and love. I woe a great deal to them.

Finally, there are no words to express my gratitude to my loving, caring, supportive, and patient soul-mate, my wife Dipika Basak, whose true love during my Ph.D. journey made this work possible. Her inspiration and motivating speech during my hard times were always cheering me up. This work would not accomplish without her great sacrifice. I would like to express my thanks to our wonderful and precious gift, my beautiful son Dibban Basak, who is always the reason for my happiness in any situation of my life.

Lastly and above all, I thank the Almighty Lord for providing me the strength to reach this stage of my life.

Udoy Sankar Basak

February 16, 2021

Contents

List of Figures

List of Tables

List of Symbols

Information theory

X,Y	Random variables
t	time
x_t	Observation of X at time t
y_t	Observation of Y at time t
$\bar{x}$	Time average of $\{x_t\}$
$\bar{y}$	Time average of $\{y_t\}$
$p(.)$	Probability function
$p(.,.)$	Joint probability function
$p(.\|.)$	Conditional probability function
τ	Delay time
$H(.)$	Shannon entropy
$H(.,.)$	Joint entropy
$H(.\|.)$	Conditional entropy
$I(.,.)$	Mutual information
$I(.\|.)$	Conditional mutual information

Vicsek model

N	Number of particles
i	Index of a particle
Δt	Time interval
$\vec{r}_i^t$	Position of particle i at time t
$\vec{v}_i^t$	Velocity of particle i at time t
$\theta_i(t)$	Orientation of particle i at time t
κ	Interaction time-scale
R	Interaction radius
$\boldsymbol{w}$	Interaction matrix
η_0	Noise

Others

λ	Cutoff distance
χ	Classifier
γ	Threshold
$\langle . \rangle_\lambda$	Average over λ
ΔR	Relative error

List of Abbreviations

VM	Vicsek model
MDCK	Madin-Darby canine kidney
TE	Transfer entropy
CC	Cross-correlation
MI	Mutual information
TDMI	Time-delayed mutual information
cAMP	Cyclic adenosine monophosphate
L	Leader
F	Follower
ROC	Receiver operating characteristic curve
TPR	True-positive rate
FPR	False-positive rate
AUC	Area under ROC curve
a.u.	Arbitrary unit
PIV	Particle Image Velocimetry
KD	Kantorovich metric

Chapter 1

Introduction

Collectively moving animals is one of the most beautiful creations of Nature Mother. The synchronized movement of bird flocks at dusk or the beauty of schools of fish can fill our hearts with an abundance of love, joy, and endless happiness again and again. Collective migration can also be observed at the cellular level, e.g., wound healing, cancer development, and organogenesis [1,2]. The basic property of collective motion is that the movement of an individual depends on the movement of its surroundings. Researchers from different disciplines have been working together to find out how microscopic one to one interaction between agents regulates the macroscopic group behavior [3].

The general conjecture of collectively moving agents is the presence of influential individuals, known as 'leaders', who control the movement of the whole group [4]. These special agents employ asymmetric influence on the other group members, known as the 'followers'.

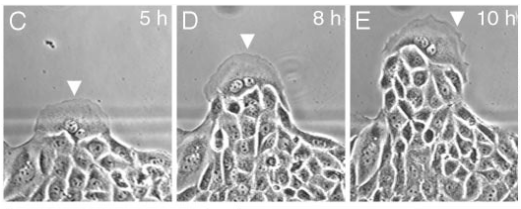

Figure 1.1: Leadership in wound healing [5].

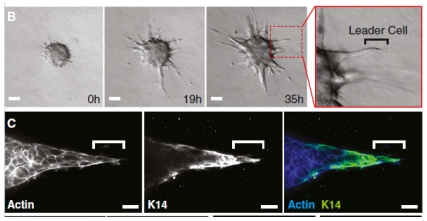

Figure 1.2: Leader cells in cancer growth [6].

In wound healing procedure, the collective behavior of cells has been observed [5]. It was found that the wound healing procedure starts with the formation of small 'mounds' and one or rarely two 'leader' cells positioned at the top of the mound pull neighboring follower cells forward into the wound. Figure 1.1 is taken from [5]. Also in cancer growth,

it has been found that some cells are 'highly protrusive and migratory' who are referred to as 'invasive leader cells' and these leader cells are readily identified at the front of the invasive strands [6]. Figure 1.2 is taken from [6].

It was found [7,8] for MDCK epithelial cells that leader cells control the collective migration of cells from the tip of finger-like structures, and removal of leader cell(s) from the tip causes interruption of the cell migration. Figures 1.3 and 1.4 are taken from [8].

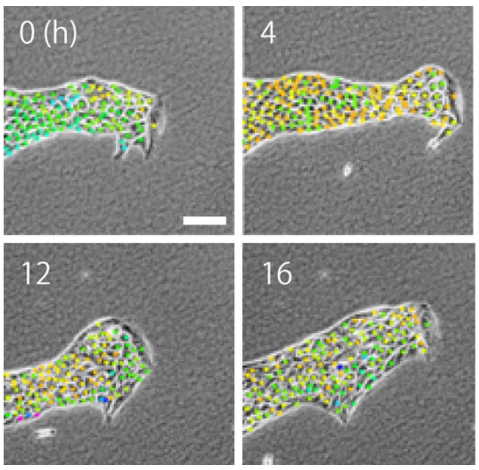

Figure 1.3: Collective motion of MDCK epithelial cells in presence of the leader cells [8].

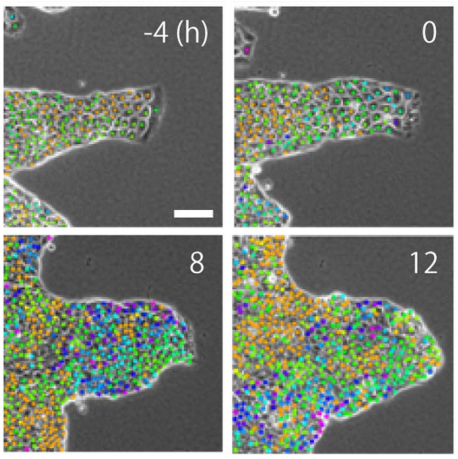

Figure 1.4: Collective motion of MDCK epithelial cells after removal of the leader cell [8].

For a fish shoal, it was found that the front fish have a powerful effect on the movement of that shoal [9]. This leader-follower relationship has also been observed in chacma baboons [10]. Studies have shown that leadership helps the group to find food, building a nest at a better place, or survive from a predator [11–13].

Identifying leader and follower particles is a puzzling endeavor. Although the relative cell position might be an indicator of potential leaders in MDCK epithelial cells, cancer growth, and fish shoal, what characterizes leaders and followers is an open question in general, for instance, for Dictyostelium discoideum [15–17] cells where there is no apparent indication of which cells might have more influence over others (Fig. 1.5, which is taken from [14]). Various types of empirical data, e.g., the ensemble of trajectories of agents, can be used to infer the differential influence in interaction and leader-follower relationship. By definition, leaders are expected to be more persuasive compared to the followers. In other words, it can be said that leader agents control the movement of the follower agents, which means that it is the leaders who 'cause' the movement of followers. Consequently, leader and follower agents can be identified by determining the direction of 'causality'.

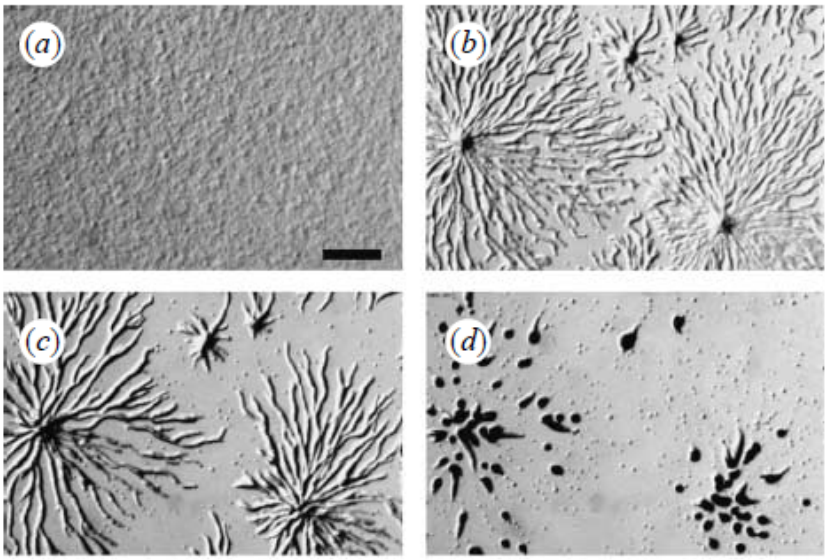

Figure 1.5: Aggregation of D.Discoideum cells [14]

Since followers follow the movement of leaders, hence some correlation should exist between a physical quantity of a leader and that of a follower. Here the words 'correlation' and 'causation' can seem misleadingly analogous. Let's first explain the difference between these two terms. Correlation is a statistical quantity that quantifies the degree of association between two random variables. There are three types of correlations that can be identified:

- **Positive correlation:** A positive correlation is a relationship between two variables in which both variables move in the same direction, i.e., when one variable increases as the other variable also increase, or one variable decreases while the other decreases as well. For example, there is a positive correlation between height and weight—taller people tend to be heavier, and vice versa.

- **Negative correlation:** A negative correlation between two variables means that one variable increases whenever the other decreases. A negative correlation can be found between illness and vaccination.

- **No correlation:** No correlation means a change in a variable has no impact on the other one. For example, there is no relationship between the amount of coffee drunk and the level of intelligence.

On the other hand, causation indicates that one variable causes another variable to

occur. Hence subtle differences exist between correlation and causation. Let's consider the following examples:

Suppose 'ice cream sales' and 'bank robbery' have a highly positive correlation. It does not mean that "ice cream encourages people to commit bank robberies". Then one may ask what is the source of this correlation. Maybe the 3rd variable, say, "weather" explains the correlation between them. Both ice cream seals and bank robberies increase during the summer. Consequently, one may conclude that bank robberies and ice cream seals are correlated but do not have any causal relationship. Hence, just because two variables are moving together (highly correlated) does not necessarily mean that one variable causes the other to occur, which leads to the famous quote "correlation does not imply causation".

Consider another example. It has been found that a high correlation exists between people who 'smoke' and people who are infected with 'cancer'. Further medical data has confirmed that smokers are much more likely to get mouth and lung cancer than non-smokers. Consequently, one can say that 'smoking causes cancer'.

Measuring causality is not a straightforward process. The first quantifiable definition of causality was propounded by mathematician Wiener [18]. It was proposed that one variable could be called 'causal' to another variable if the information about the first variable improves the predictability of the second variable [19]. The first variable is known as the 'cause' and the second one is called the 'effect'. But the practical implementation of this idea was missing. Later C. Granger [20] proposed a definition that can be applied to real data. Suppose that $X = \{..., x_{t-1}, x_t, x_{t+1,...}\}$ and $Y = \{..., y_{t-1}, y_t, y_{t+1}, ...\}$ are two random variables and we are trying to identify if there is a causal relationship between them. By definition, a causal relationship can be inferred if one variable helps us to predict the other one. Suppose we are trying to predict x_{t+1}. One can use only the past terms of X to predict x_{t+1}. Also, past terms of both X and Y can be used to predict x_{t+1}. If the second prediction is more accurate compared to the first one, then one can certainly conclude that the past of Y contains information that is useful to predict x_{t+1} that is not in the past of X. In this case, Y is said to be a G-cause (Granger cause) of X [19].

When there is a causal relationship between two variables, there must exist a certain time lag (delay) however small between the cause and the effect [19,21]. One could also

say the Wiener-Granger framework of prediction based 'causality' is equivalent to looking for dependencies between the variables at a certain time delay [22].

Granger causality is being widely used in different dynamical systems from economics and finance [23,24], physics [25], bioinformatics [26], social science [27], and many other fields. Granger's analysis is based on the linear relationship between the two variables. But dynamical systems could be highly nonlinear, which calls into question the application of Granger analysis in the non-linear systems [28].

To overcome the non-linearity challenge, T. Schreiber [29] proposed a probability distribution based measure called transfer entropy (TE) which is capable of capturing the direction of causality in non-linear systems using the Markov property [22]. TE measures the reduction of uncertainty of a variable. The underlying concept of TE is that if a random variable Y influences the outcome of another variable X, then the outcome of the variable X can be predicted more accurately by incorporating the history of both X and Y compared with that only incorporating X's own history [28].

TE has been applied in many areas of science and engineering, such as neuroscience [30], complex dynamical systems[31,32], environmental engineering [33], and many other fields. It has also been used in identifying leader and follower agents. Butail et al. [34] found by using the zebrafish interaction model that the net transfer entropy is the most accurate classifier compared to extreme-event synchronization and cross-correlation for classifying leader-follower relationship. Orange et al. [35] also found that there exists leader-follower interaction between the front bat and the rear bat based on the transfer entropy. Using the trajectories of handball players it has been proved that transfer entropy is capable of capturing the causal relationships between players [36].

In real applications, it is near impossible to know the true identity of the agents *a priori*. Hence different simulation models can be used to check the classification performance of different methods. To study the collective behavior, a variety of simulation models have been introduced, such as the Reynolds' flocking model [37,38], the Vicsek model (VM) [39], and the Couzin model [40,41], etc,. These models are easy to simulate on computers which have facilitated the analysis of collective behavior [3]. Because of its minimal character, the VM is one of the widely used models to study various dynamics of collectively moving self-propelled particles (SPPs) [42–44]. In the VM, the direction of movement of each particle is determined by the average direction of its neighboring

particles which are located within a circle of an interaction radius R centered on the particle. For instance, the Vicsek model (VM) has been studied as the minimalist model for shedding light on unveiling the mechanisms of collective motion of a group of self-propelled particles, such as symmetry breaking [45,46], phase transition [39,47,48], and classification of leader and follower agents in a community [34,49]. In this thesis, a modified version of the VM has been proposed to check the performance of our classifier in identifying leader and follower particles and also in identifying interaction radius R from the motion data of the moving agents.

As stated above, TE has been used in identifying leader and follower agents. But in these studies, all pairs of agents are taken into account to evaluate transfer entropy irrespective of the distance between the agents. This thesis demonstrates a protocol to identify the interaction domain. It has been found that the classification of leader and follower particles is significantly improved by using the identified interaction domain.

This thesis will be structured as follows:

- Chapter 2 will provide some measures that can be used to identify the direction of causality. Section 2.1 will discuss cross-correlation, which is a method that has been used to identify the linear relationship between two variables at certain time-lag. Section 2.2 will provide the information-theoretic measures starting from entropy, mutual information (MI), time-delayed mutual information (TDMI), and transfer entropy.

- Chapter 3 will discuss the application of TE in identifying leader cells in a colony of Dictyostelium Discoideum amoeba.

- Chapter 4 will represent a modified form of the Vicsek model (VM). In this modified VM, an extra term, in the form of 'interaction strength' will be used based on which leaders and followers will be determined.

- Chapter 5 will show the performance of cross-correlation and transfer entropy in identifying leader and follower particles. In this chapter, different parameters will be varied to check the effect of these parameters on the classification score.

- Chapter 6 will be dedicated to determining the interaction domain. It will also confirm that the classification performance can be improved by utilizing the pairwise

distance information of two agents. Also, the performance of cross-correlation and transfer entropy in determining the interaction domain will be examined.

- Chapter 7 will discuss the effect of finite length trajectories and external noise on the identification of the interaction domain. Also, a different scheme to identify the interaction domain will be discussed here.

- Chapter 8 will discuss some directions for this research that could be addressed in the future.

Chapter 2

Quantifying influence between particles

Introduction

Identifying leader and follower particles is a puzzling endeavor. Various types of empirical data, e.g., the ensemble of trajectories of agents, can be used to infer the differential influence in interaction and leader-follower relationship. By definition, leaders are expected to be more persuasive compared to the followers. Since followers follow the movement of leaders, hence some correlation should exist between a same physical quantity of a leader and a follower with a certain time delay. Different types of correlation, e.g., cross-correlation [9], extreme-event synchronization [34], can be used to quantify the interaction between two agents. But in nature, correlations are highly likely to be non-linear. Therefore the methods based on correlation are too restrictive to investigate highly non-linear systems [50]. To measure non-linear causal influence among multivariate time series, information theory (IT) provides a variety of approaches [51]. Some of them are mutual information [52], time-delayed mutual information [29], transfer entropy [29], causation entropy [53] etc,. These methods have been used on the time-lapse motion data of moving individuals to determine the direction of influences. This chapter provides some of the measures that can be used to infer leader and follower particles.

Consider $X = (..., x_{t-1}, x_t, x_{t+1}, ...)$ and $Y = (..., y_{t-1}, y_t, y_{t+1}, ...)$ are two Markov processes with probability mass functions $p(x_t) = Pr\{X = x_t\}$ and $p(y_t) = Pr\{Y = y_t\}$ respectively.

2.1 Cross-correlation

Cross-correlation (CC) has successfully been used to identify whether some individuals have a strong influence on other particles [9,54,55]. CC between two random variables X and Y as a function of time-delay τ has the following form

$$C_{XY}(\tau) = \frac{\sum_t [(x_t - \bar{x})(y_{t+\tau} - \bar{y})]}{\sqrt{\sum_t (x_t - \bar{x})^2} \sqrt{\sum_t (y_{t+\tau} - \bar{y})^2}}, \tag{2.1}$$

where $\bar{x}$ and $\bar{y}$ denote the time averages of $\{x_t\}$ and $\{y_t\}$, respectively. For instance, if there exists a positive delay time τ that maximizes the CC $C_{XY}(\tau)$, one might interpret that X has some influence over Y at the level of linearity [9].

CC can, however, take into account the solely linear correlation between the variables [50]. But in nature correlations are highly likely to be non-linear. As a result, CC is too restrictive to investigate highly non-linear systems.

2.2 Information theory

Information theory is a branch of mathematical theory within probability and statistics that can be used to quantify information. This section discusses some of the information-theoretic measures.

2.2.1 Shannon Entropy

In information theory, the fundamental measure of information is *Shannon entropy* or simply *entropy*. It is defined as the average minimum number of bits needed to encode a random variable based on the frequency of its value. The entropy of X is defined as follows [52,56,57]:

$$H(X) = -\sum_{x_t} p(x_t) \log_2 p(x_t), \tag{2.2}$$

The unit of entropy depends on the base of the *logarithm*. In this paper, we use 2 as the base of *logarithm* so that entropy is measured in *bits*. Since $0 \leq p(x_t) \leq 1$, hence it is clear that $H(X)$ is a non-negative quantity.

Entropy is a measure of average uncertainty in a random variable [52]. Consider the

following two examples.

Example 2.2.1 *Suppose we have six-sided fair dice. Fair dice means that each of the faces has a probability of* $\frac{1}{6}$*. Let the random variable* X *represents the outcome of the toss of that fair dice. The probability distribution of the random variable* X *is as follows:*

Table 2.1: Distribution of a fair dice

X	$p(X)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Thus, a fair six-sided dice should have the entropy,

$$\begin{aligned} H(X) &= -\sum_{i=1}^{6} H(X=i)\log_2 p(X=i), \\ &= -6 \times \frac{1}{6} \times \log_2 \frac{1}{6}, \\ &= 2.5850 \; bits, \end{aligned}$$

Example 2.2.2 *Let's discuss the entropy of biased dice. A biased dice is a dice where some numbers are more likely to be rolled than others. In this example, we assume that the probability of getting an even number in a dice roll is three times higher than an odd number. Suppose the random variable* Y *represents the outcome of the toss of that biased dice whose probability distribution is as follows:*

Table 2.2: Distribution of a biased dice

X	$p(X)$
1	1/12
2	1/4
3	1/12
4	1/4
5	1/12
6	1/4

The entropy of Y is,

$$\begin{aligned} H(Y) &= -3 \times \frac{1}{12} \times \log_2 \frac{1}{12} - 3 \times \frac{1}{6} \times \log_2 \frac{1}{6}, \\ &= 2.3962 \; bits, \end{aligned}$$

Since X is the distribution of fair dice, hence it is difficult of predicting the outcome of this dice compared to the biased dice Y (as an even number is more likely to appear). In other words, X has more uncertainty compared to Y. As the entropy is the measure of uncertainty, consequently we have found that $H(X) > H(Y)$. Thus we can conclude the followings:

1. A random variable has a maximum entropy when all of its outcomes are of equal probable.

2. Entropy would be minimum when one value of the random variable is more likely to happen and others are very rare.

3. In the rest of the cases, entropy would be a positive quantity.

2.2.2 Joint Entropy

The *joint entropy* of X and Y is denoted by $H(X, Y)$ and is defined as follows [52,58]:

$$H(X, Y) = -\sum_{x_t} \sum_{y_t} p(x_t, y_t) \log_2 p(x_t, y_t), \tag{2.3}$$

where $p(x_t, y_t)$ represents the joint probability. Joint entropy $H(X, Y)$ represents the amount of uncertainty contained in the pair of random variables X and Y. From Eq. (2.3)it is obvious that joint entropy is symmetric i.e., $H(X, Y) = H(Y, X)$ [59]. It is a non-negative quantity. If X and Y are statistically independent then $p(x_t, y_t) = p(x_t)p(y_y)$. Then Eq. (2.3) implies that $H(X, Y) = H(X) + H(Y)$. In rest of the cases, $H(X, Y) < H(X) + H(Y)$ [52].

Example 2.2.3 *Suppose X and Y represent the outcomes of rolling two fair dice. Then the joint probability distribution has the following form:*

Table 2.3: Joint distribution of two dice

Y \ X	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

The joint entropy of the random variables X and Y is:

$$\begin{aligned} H(X,Y) &= -\sum_{x_t=1}^{6}\sum_{y_t=1}^{6} p(X=x_t, Y=y_y)\log_2 p(X=x_t, Y=y_t), \\ &= -36 \times \frac{1}{36} \times \log_2 \frac{1}{36}, \\ &= 5.17\ bits, \end{aligned}$$

Since both X and Y are fair dice, hence from example (2.2.1) we have $H(X) = 2.5850\ bits$ and $H(Y) = 2.5850\ bits$. Thus, the sum of their entropies is

$$\begin{aligned} H(X) + H(Y) &= 2.5850 + 2.5850, \\ &= 5.17, \\ &= H(X,Y), \end{aligned}$$

which is reasonable as the two events are independent.

2.2.3 Conditional entropy

The conditional entropy of Y given X has the following form [52,60]:

$$\begin{aligned} H(Y|X) &= -\sum_{x_t} p(x_t)\log_2 p(y_t|X=x_t), \\ &= -\sum_{x_t} p(x_t)\sum_{y_t} p(y_t|x_t)\log_2 p(y_t|x_t), \\ &= -\sum_{x_t}\sum_{y_t} p(x_t,y_t)\log_2 p(y_t|x_t), \end{aligned} \tag{2.4}$$

where $p(x_t, y_t)$ and $p(y_t|x_t)$ are the joint probability and conditional probability of Y given X respectively. Conditional entropy $H(Y|X)$ represents the remaining uncertainty about Y after knowing X and it is not symmetric i.e., $H(Y|X) \neq H(X|Y)$.

Relationship between joint entropy and conditional entropy: From Eq. 2.3 we have [52],

$$\begin{aligned}
H(X,Y) &= -\sum_{x_t}\sum_{y_t} p(x_t, y_t)\log_2 p(x_t, y_t),\\
&= -\sum_{x_t}\sum_{y_t} p(x_t, y_t)\log_2 p(y_t)p(x_t|y_t),\\
&= -\sum_{x_t}\sum_{y_t} p(x_t, y_t)\log_2 p(y_t) - \sum_{x_t}\sum_{y_t} p(x_t, y_t)\log_2 p(x_t|y_t),\\
&= -\sum_{p(y_t)} p(y_t)\log_2 p(y_t) - \sum_{x_t}\sum_{y_t} p(x_t, y_t)\log_2 p(x_t|y_t),\\
&= H(Y) + H(X|Y),\\
&= H(X) + H(Y|X),
\end{aligned} \tag{2.5}$$

If Y is independent of X then $p(y_t|x_t) = p(y_t)$, which yields

$$H(Y|X) = H(Y), \tag{2.6}$$

which means that the knowledge about X does not help to reduce the uncertainty about Y at all.

Example 2.2.4 *Let us consider the following table of height (X) (in cm.) and weight (Y) (in kg.) of five (05) students:*

Table 2.4: Height and weight table

Height(X)	Weight(Y)
152.4	57
167.0	75
154.5	67
177.3	82
169.0	82

From the 1st column of the table (2.4) it is clear that all the values of height (X) are

unique, whereas in the weight (Y) column 82 appears twice. Equation (2.2) implies that $H(X)$ should be greater than $H(Y)$. After some calculations, we get

$$H(X) = 2.3219 \; bits$$
$$H(Y) = 1.9219 \; bits$$
$$H(X,Y) = 2.3219 \; bits$$

From Eq. (2.4) we get $H(Y|X) = 0 \; bits$ and $H(X|Y) = 0.4 \; bits$. From the table (2.4) it is obvious that if anyone knows the height (X) of a student, he/she will be able to determine the weight (Y) of that student. Hence no uncertainty remains in Y once we know X, which implies that $H(Y|X) = 0$. On the other hand, suppose $Y = 57 \; kg$, then certainly one can identify the height (X) of that student which is 152.4 cm. But if $Y = 82 \; kg$, then one cannot determine the height of that student precisely. Because both two students are having the same weight $Y = 82 \; kg$ but different height 177.3 cm and 169.0 cm. Consequently, some uncertainty remains about X once Y is known.

2.2.4 Mutual information

The random variables X and Y might be dependent or independent with respect to each other. To quantify the degree of dependence one may compute mutual information (MI) as follows [52,61,62]:

$$I(X;Y) = \sum_{x_t, y_t} p(x_t, y_t) \log_2 \frac{p(x_t, y_t)}{p(x_t)p(y_t)}, \tag{2.7}$$

Mutual information represents the amount of information that one variable contains about the other one. In particular, if $I(X;Y) = 0$ then Eq. (2.7) implies that $p(x_t, y_t) = p(x_t)p(y_t)$, which indicated that the random variables X and Y are independent [28]. Easily one can also find that mutual information is symmetric, that is, $I(X;Y) = I(Y;X)$ [63–65]. So it has no directional sense.

Relationship between MI and entropy, joint entropy and conditional en-

tropy: From Eq. (2.7), we have [52]

$$\begin{aligned}
I(X;Y) &= \sum_{x_t,y_t} p(x_t,y_t)\log_2\frac{p(x_t,y_t)}{p(x_t)p(y_t)},\\
&= \sum_{x_t,y_t} p(x_t,y_t)\log_2\frac{p(x_t|y_t)p(y_t)}{p(x_t)p(y_t)},\\
&= \sum_{x_t,y_t} p(x_t,y_t)\log_2\frac{p(x_t|y_t)}{p(x_t)},\\
&= \sum_{x_t,y_t} p(x_t,y_t)\log_2 p(x_t|y_t) - \sum_{x_t,y_t} p(x_t,y_t)\log_2 p(x_t),\\
&= \sum_{x_t,y_t} p(x_t,y_t)\log_2 p(x_t|y_t) - \sum_{x_t} p(x_t)\log_2 p(x_t),\\
&= -H(X|Y) + H(X),\\
&= H(X) - H(X|Y),\\
&= H(Y) - H(Y|X),
\end{aligned} \tag{2.8}$$

Using Eq. (2.5) in Eq. (2.8) we get

$$\begin{aligned}
I(X;Y) &= H(X) - H(X|Y),\\
&= H(X) + H(Y) - H(X,Y),
\end{aligned} \tag{2.9}$$

Using Venn diagram Eqs.(2.8) and (2.9) can be represented as follows [52]:

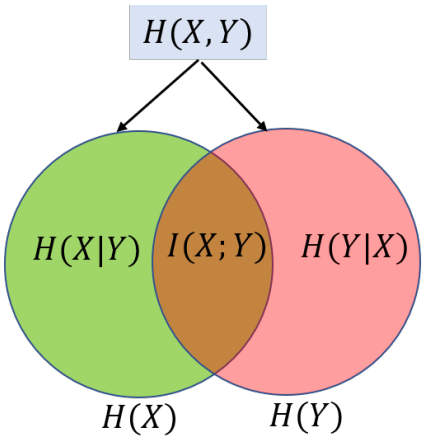

Figure 2.1: Relationship between MI and entropy.

2.2.5 Time-delayed mutual information

Since MI is symmetric, so it cannot be used to infer the direction of interaction between two signals [28]. But to infer causality, the direction of information flow is necessary. However, to capture the delay of information transfer between the two processes one can introduce a time-lag parameter τ in any variable to compute MI which is known as time-delayed mutual information (TDMI). TDMI has the following form [66,67]:

$$I(X(t);Y(t+\tau)) = \sum_{x_t,y_{t+\tau}} p(x_t,y_{t+\tau}) \log_2 \frac{p(x_t,y_{t+\tau})}{p(x_t)p(y_{t+\tau})}, \tag{2.10}$$

A nonzero amplitude of the mutual information as a function of delay-time τ indicates the presence of interaction between the processes X and Y. The sign of τ can be used to deduce the causal direction of influence [28]. If there exists a positive τ for which TDMI has a peak as a function of τ (Eq. (2.10)), then one can deduce that X shares a maximum amount of information with the future state of Y. In other words, one can say that X drives Y. Whereas a negative τ indicates that X shares the maximum amount of information with the past of Y, hence X is driven by Y [28].

TDMI is capable of capturing both linear and nonlinear correlation between two time-series and hence can be used as a measure for mutual coupling or information transfer [68]. However, TDMI fails to consider shared history (also common external driving effects) between two processes and lead to spurious inferences of directed information transfer [69]

2.2.6 Transfer entropy

To resolve the above-mentioned problems Schreiber[29] proposed time-delayed mutual information via conditioning $I(X(t);Y(t+\tau)|Y(t))$, a quantity known as transfer entropy (TE)

$$\begin{aligned} \mathrm{TE}_{X\to Y} &= I(y_{t+\tau};x_t|y_t), \\ &= \sum_{y_{t+1},y_t,x_t} p(y_{t+1},y_t,x_t) \log_2 (\frac{p(y_{t+1}|y_t,x_t)}{p(y_{t+1}|y_t)}), \\ &= H(y_{t+1}|y_t) - H(y_{t+1}|y_t,x_t), \end{aligned} \tag{2.11}$$

However, conditioning does not imply purely subtraction, which means that the following operation may not always be true:

$$I(X;Y|Z) \leq I(X;Y)$$

Rather, conditioning may increase the shared information between two variables. Let us consider the following example. Suppose X and Y are two mutually independent fair Bernouli variables, i.e., can take values $\{0,1\}$ with probability $1/2$. Another random variable Z is defined as follows:

$$Z = X \oplus Y,$$

where $\oplus$ represents XOR operation. Then the probability distribution would be as follows:

Table 2.5: Probability distribution

X	Y	Z	Pr
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	0	1/4

Easily one can find that all these variables are pairwise independent, which produces zero mutual information as follows:

$$I(X;Y) = I(X;Z) = I(Y;Z) = 0.$$

But by knowing X (or Y) certainly, Y (or X) can predict the outcomes of Z, that yields $I(X;Z|Y) = I(Y;Z|X) = 1$ bit. Similarly, we can find that $I(X;Y|Z) = 1$ bit. Thus, even though the variables X and Y do not share any information (mutually independent), but if we are given Z (condition on Z), they are perfectly correlated.

The idea of TE is that if a process X influences another process Y, then the prediction of Y becomes easier after knowing the past of both X and Y, compared to only knowing the past of Y. TE from X to Y is defined as follows [70–72]:

It is a non-negative quantity. If Y is independent of the past of X, that is, in the absence of causality (X to Y), Eq. (2.6) is fully satisfied, then $\mathrm{TE}_{X\to Y} = 0$. Thus, a positive value of $\mathrm{TE}_{X\to Y}$ is the indication of causal influence of X on Y [73,74]. For a pair of agents X and Y, the net transfer entropy from X to Y, defined as $\mathrm{NTE}_{X\to Y} =$

$\mathrm{TE}_{X\to Y} - \mathrm{TE}_{Y\to X}$ can be used to infer the direction of causal influence. A positive $\mathrm{NTE}_{X\to Y}$ indicates that Y follows X. In that case, we may infer that X is a leader and Y is a follower.

2.3 Concluding Remarks

To detect causality different types of linear measures have been used. But later, information theoretic measures have been found productive as these measures are capable of capturing non-linear information flow, which enables us to detect the direction of causality more precisely. In this thesis we compare the performance of both linear (cross-correlation) and information-theoretic measure (transfer entropy) in identifying leader and follower particles.

Chapter 3

Leaders in D. Discoideum Colony

3.1 Introduction

Dictyostelium discoideum is a soil-living amoeba [75,76]. The life cycle of D. Discoideum has two distinct phases: vegetative phase in which single-cell amoebae feed on bacteria by phagocytosis and divide by fission every few hours, and a social phase initiated by a period of starvation, in which the single-cell amoebae aggregate to form a multicellular structure [77,78].

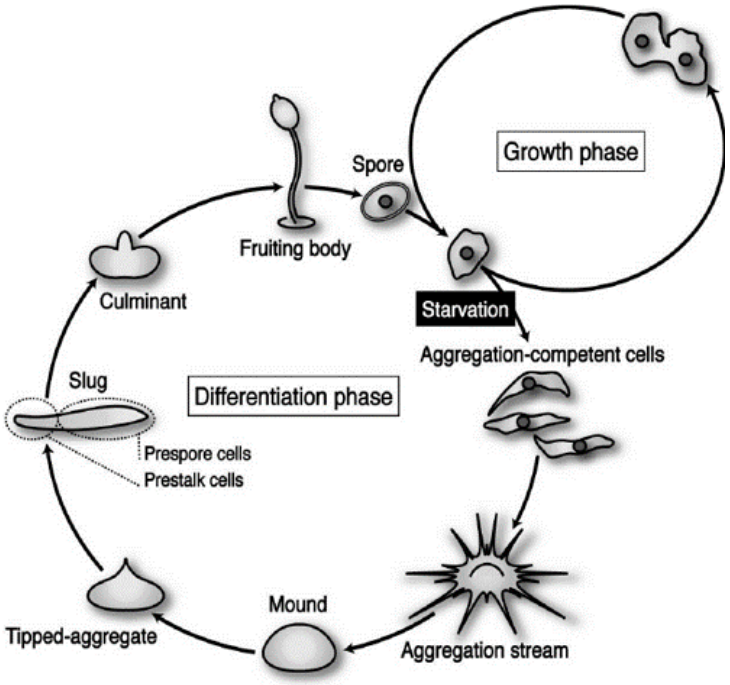

Figure 3.1: Life cycle of D. Discoideum cells [79].

Since D. Discoideum exhibits both single-cell and multi-cellular characteristics, hence

it can be used as a model for the study of single-cell functions (e.g., cell division, differentiation, chemotaxis, and death), as well as multi-cellular activities such as organ development, etc, [80–82]. Because of its short life cycle, easy to cultivate, and can be harvested in large amount, D. Discoideum is one of the widely used species [83].

In the social phase, cells release cAMP pulse to communicate with each other. Finally, these cells aggregate to form a multi-cellular structure. It has been found that some cells first start to emit rhythmic pulses of cAMP [84]. These cells can be termed as the 'leader cells' as they start the aggregation procedure and the other cells can be termed as the 'follower cells'. Our main focus is to identify the leader and follower cells based on the cAMP secretion. We will use an information-theoretic scheme called transfer entropy to identify leader and follower cells and compare our results with the expert's results.

*I express my gratitude to Prof. Kazuki Horikawa of Tokushima University for providing us the imaging data as well as single-cell tracking data.

3.2 cAMP secretion and cell aggregation

During its vegetative stage, the D. Discoideum cells move independently, replicate by binary fission and extract nutrients from bacteria in the soil [85]. Upon nutrient depletion, starvation induces amoebae to aggregate in a place to form mounds (multi-cellular stage) that organize into a migrating slug [85]. Figure 3.2 shows the aggregation steps of D. Discoideum cells at three different periods. This aggregation process raises some questions:

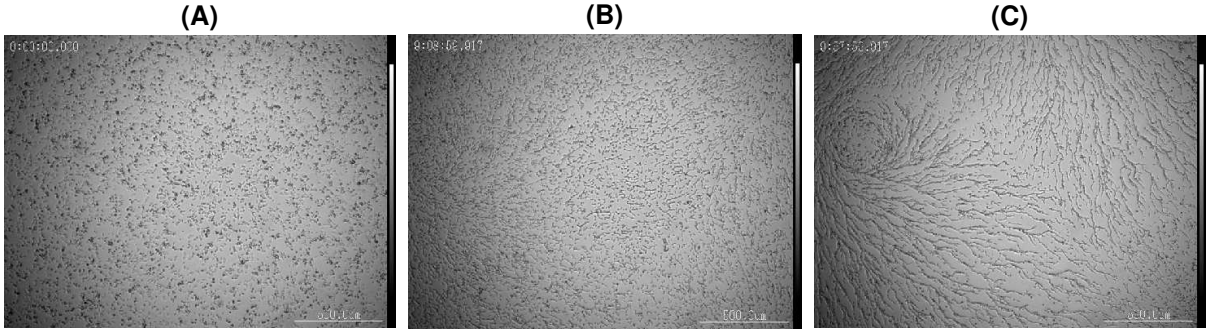

Figure 3.2: Aggregation of D. Discoideum cells at three different time periods. (A) Cells are moving independently. (B) Cells are trying to aggregate in one place. (C) Final aggregation.

- How do these cells communicate with each other?

- How do they know when to aggregate?
- How do they know where to aggregate?

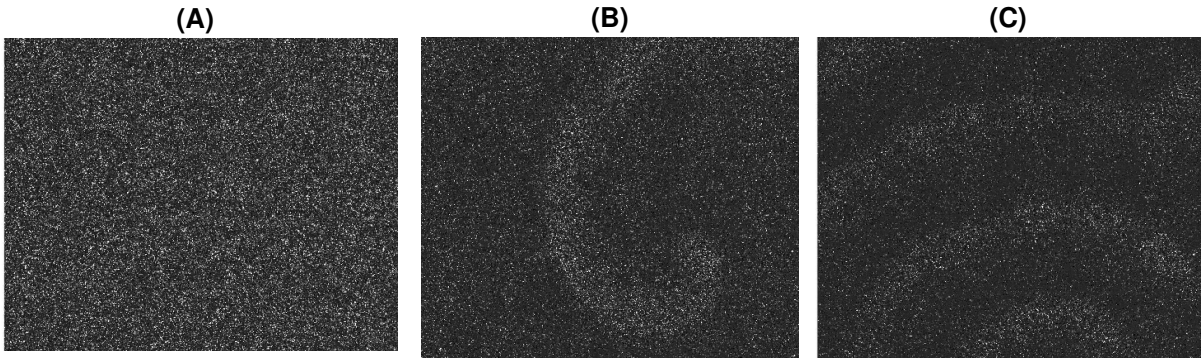

Figure 3.3: cAMP secretion of D. Discoideum cells at three different time periods.

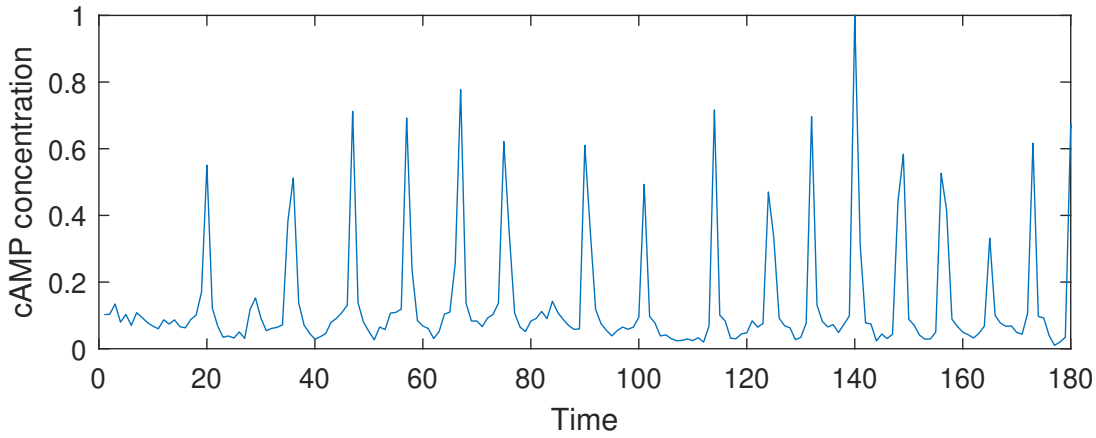

Figure 3.4: cAMP secretion of D. Discoideum cell

In response to the above questions, scientists have found that the aggregation takes place employing 'chemotaxis' which is the movement of an organism accompanied by the relay of the chemoattractant [86,87]. The attractant is a chemical called cyclic adenosine monophosphate (cAMP), which is produced and periodically released by the cells when the starvation takes place [87]. This cAMP is the primary signaling molecule that drives the aggregation process. A cell that senses external cAMP moves towards the source of cAMP and in turn, releases a burst of cAMP. Consequently, local variations in cell density get amplified which finally results in an aggregation [87]. Figure 3.3 shows the cAMP wave of a D. Discoideum colony at three different periods. In Fig. 3.3(A) no cAMP is visible and the cells are moving randomly. Figure 3.3(B) shows that the spiral wave is forming and Fig. 3.3(C) shows the cAMP during the aggregation.

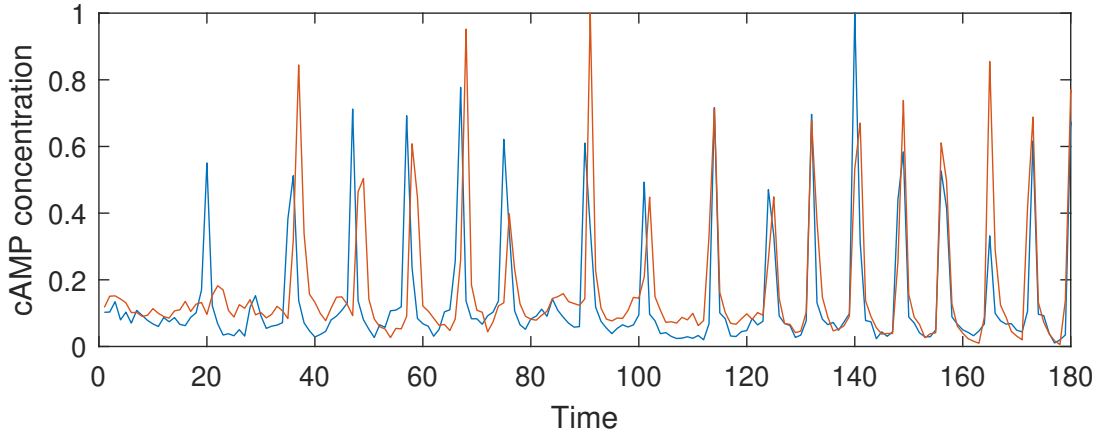

Figure 3.5: Firing comparison of two D. Dscoideum cells

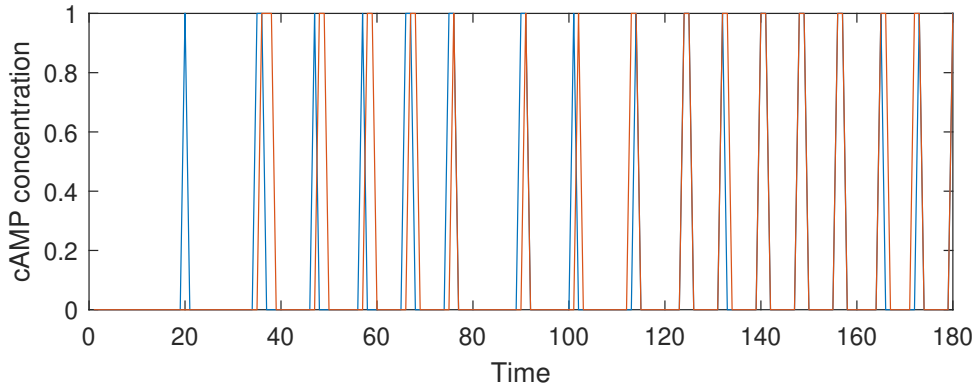

Figure 3.6: Binarization of cAMP concentration.

3.3 Identifying Leaders cells in Dicty colony

3.3.1 Measuring cAMP

The aggregation of D. Discoideum cells begins with the secretion of cAMP.

Our working hypothesis is proposed by the expert that leader cells first release cAMP. Later followers also emit cAMP and move towards the origin of cAMP secretion i.e., to leader cells. Hence time-lapse image data can be used for inferring leader and follower cells. Follower cells get the information of releasing cAMP from the leader cells, which means that the information of cAMP secretion flows from leader to follower, not from follower to leader. Consequently, based on the direction of information transfer between cells, leader and follower cells can be inferred.

Different types of the fluorescent indicator have been developed to measure cAMP concentration level [88,89]. Figure 3.4 shows the cAMP concentration of a single D. Discoideum cell over a period of time which is obtained by the 'single-cell tracking' technique

and measuring cAMP concentration. A flash-type red fluorescent cAMP indicator called R-FlincA has been used for this purpose. The images were captured every 90 seconds. The peaks are associated with the cAMP wave which means that when a cell release cAMP the intensity of the cells increases. In between two secretions of pulse, there is a silent region which is represented by the empty region in Fig. 3.4.

3.3.2 Comparison of cAMP secretion between cells

Figure. 3.5 shows the cAMP secretion of two D. Discoideum cells over a period of time. From the figure it is clear that the cells do not emit cAMP at the same time. Some cells first start to release cAMP and the others cells try to follow them. In this section our focus would be to compute TE from one cell to another based on the cAMP concentration and then classify leader and follower cells based on the TE.

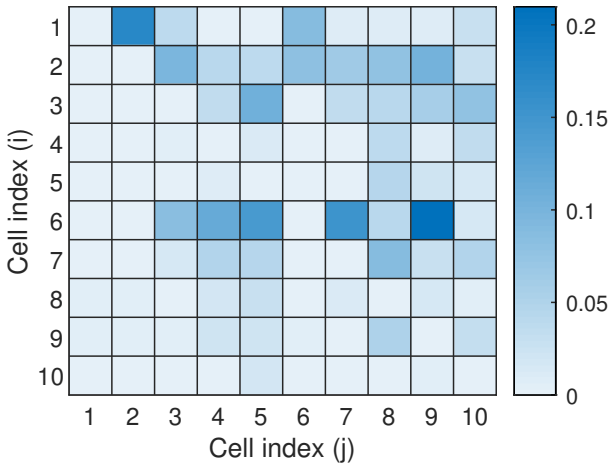

Figure 3.7: TE between cells.

3.3.3 Transfer entropy to identify leader cells

From Eq. (2.11) it is clear that in order to compute TE from one cell to another, we need to discretize cAMP pulse data. For simplicity we use simple binarization technique for discretization. Figure 3.6 shows the binarize representation of Fig. 3.5, where the peaks in the actual cAMP data have been represented by 1 and the others are by 0.

Here we analyze the binarized cAMP data of 10 single-tracked cells. Figure 3.7 shows the TE from one cell to another and the numbers 1-10 represent the index of the cells. In particular, the box whose position is (i, j) represents the TE from the cell i to cell j.

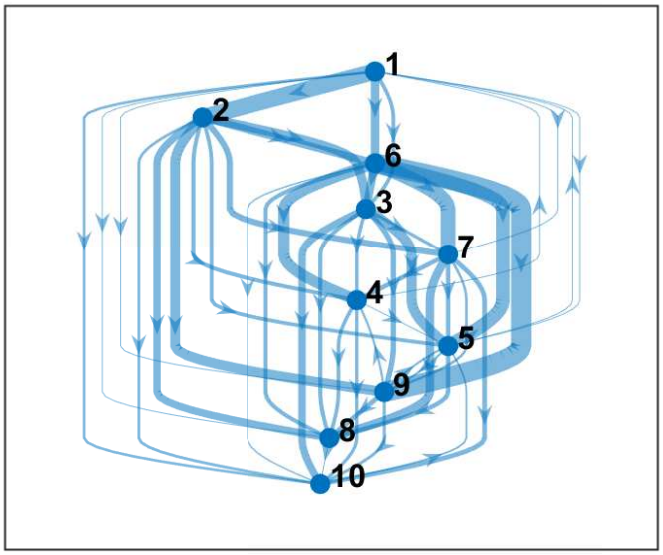

Figure 3.8: Hierarchical representation of cells. The position of the cells is based on the influence of the respective cell on other cells.

From the TE chart 3.7, it is clear that cell 1 and cell 2 have higher outward TE (TE from cell 1 and 2 to others) compared to inward TE (from other cells to cell 1 and 2). Also, cell 6 has higher outward TE. Based on the TE values we may conclude that cell 1, cell 2, and somehow cell 6 are the possible leader cells.

Figure 3.8 is the hierarchical representation of the cells based on the TE values. Edges represent the strength of interaction and the arrows are the directions of the interaction.

3.4 Concluding Remarks

In this chapter, the author has analyzed the single-cell tracking cAMP data. The single-cell tracking data of 10 cells have been analyzed and TE has been used to identify the leader and follower cells. Based on the TE between cells we find that cell 1 and cell 2 are the leaders of the colony which matches the classification results obtained by the expert who generated this data set. But cell 6 also appears to be a leader in our protocol, so it leaves some ambiguity whether our method is correct or expert. To check the reliability of the classification process, the well-known Vicsek model has been used (in the later chapters) where identities of cells (leaders or followers) are known.

Chapter 4

Modified Vicsek Model

4.1 Introduction

In the original VM [39], the influences of all the particles were the same. Since our main focus is to classify leader and follower particles, hence we let some particles having a higher influence on other particles in the system. These influential particles are known as the leader particles and the others are the followers. In the simulation model, the identity of leader and follower particles are known to us, which enables us to check the performance of our scheme in identifying leader and follower particles.

4.2 The Modified Vicsek Model

To scrutinize how our scheme works in estimating the interaction domain and leader-follower relationship from an ensemble of trajectories of many-particle systems, we employ a model similar to the original VM [39] (Fig. 4.1). Consider a two-dimensional square box of length L with periodic boundary conditions and N particles which are randomly distributed and oriented at the initial state.

The position of particle i $(i = 1, 2, ..., N)$ at time $t + 1$ is denoted by $\vec{r}_i^{\,t+1}$ and is updated at each time step Δt as:

$$\vec{r}_i^{\,t+1} = \vec{r}_i^{\,t} + \vec{v}_i^{\,t}\Delta t, \tag{4.1}$$

where $\vec{r}_i^{\,t}$ is the position of particle i at time t, and $\vec{v}_i^{\,t}$ is its velocity. For the sake

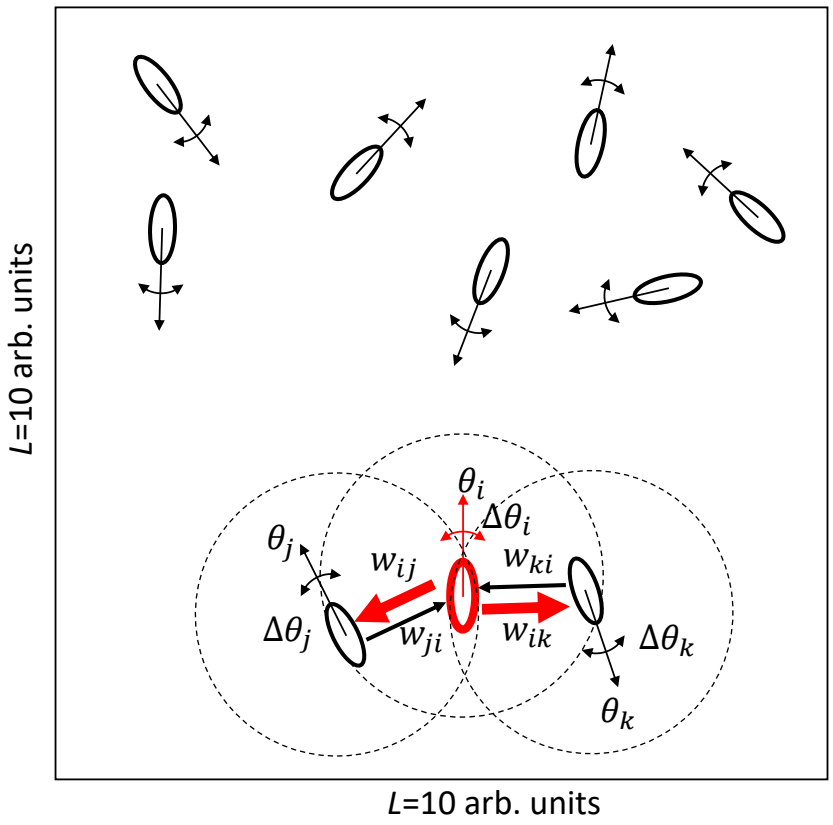

Figure 4.1: A Vicsek model with different weights between particles. The thick and thin ovals denote leader and followers, respectively, and the dotted circles represent the interaction domain for each particle. Only when the particles are within the interaction radius R, the interactions among the particles are set.

of simplicity, we assume that each particle moves with a constant speed v_0 so that its velocity is characterized by its orientation angle $\theta_i(t)$. This orientation angle is updated at each time step using the following equation:

$$\theta_i(t+1) = \langle \boldsymbol{\theta}(t-\kappa+1) \rangle_{R,\boldsymbol{w},\vec{r}_i^{\,t}} + \Delta\theta_i, \tag{4.2}$$

where $\langle \boldsymbol{\theta}(t-\kappa+1) \rangle_{R,\boldsymbol{w},\vec{r}_i^{\,t}}$ is the weighted orientation averaged over other particles which are located within a circle of radius R at time $(t-\kappa+1)$ centered on the position $\vec{r}_i^{\,t}$ of particle i at time t:

$$\arctan\Bigg[\Bigg[w_{ii}\sin\theta_i(t) + \sum_j{}' w_{ij}\sin\theta_j(t-\kappa+1)\Bigg]\Big/ \Bigg[w_{ii}\cos\theta_i(t) + \sum_j{}' w_{ij}\cos\theta_j(t-\kappa+1)\Bigg]\Bigg],$$

where $\sum'$ takes over all j satisfying $\mid \vec{r}_i^{\,t} - \vec{r}_j^{\,t-\kappa+1} \mid \leq R$ excluding particle i itself (for the derivation see Appendix A). The variable κ represents the time delay in interactions between particles. In nature, birds, fishes, cells, and so forth interact with each other in some finite time different from the timescale of the movement, e.g., interactions between Dictyostelium discoideum cells are mediated by cAMP in which the timescale of diffusion of the chemicals determines the time lag κ [15]. cAMP requires some finite time to reach other cells from the cell that emits the cAMP, and during this time lag, the cells change their positions.

In Eq. (4.2), $\boldsymbol{w}$ is a matrix whose element w_{ij} corresponds to the interaction strength that particle i exhibits on its neighboring particle j. The matrix $\boldsymbol{w}$ is introduced to emulate asymmetric interactions between particles. If particle i is a leader and j is a follower, then $w_{ij} > w_{ji}$. We set the values of leaders' and followers' interaction strengths to 1.05 and 1.00, respectively (i.e., $w_{\mathrm{LL}} = 1$, $w_{\mathrm{LF}} = 1.05$ and $w_{\mathrm{F}*} = 1$ ($*$ means either leader (L) or follower (F))). Note that, if all w_{ij}'s are unity and κ is equal to 1, this model coincides with the original VM. $\Delta\theta_i$ is a random number uniformly distributed in the range $[-\eta_0/2, \eta_0/2]$, where η_0 may correspond to a temperature-like parameter [39].

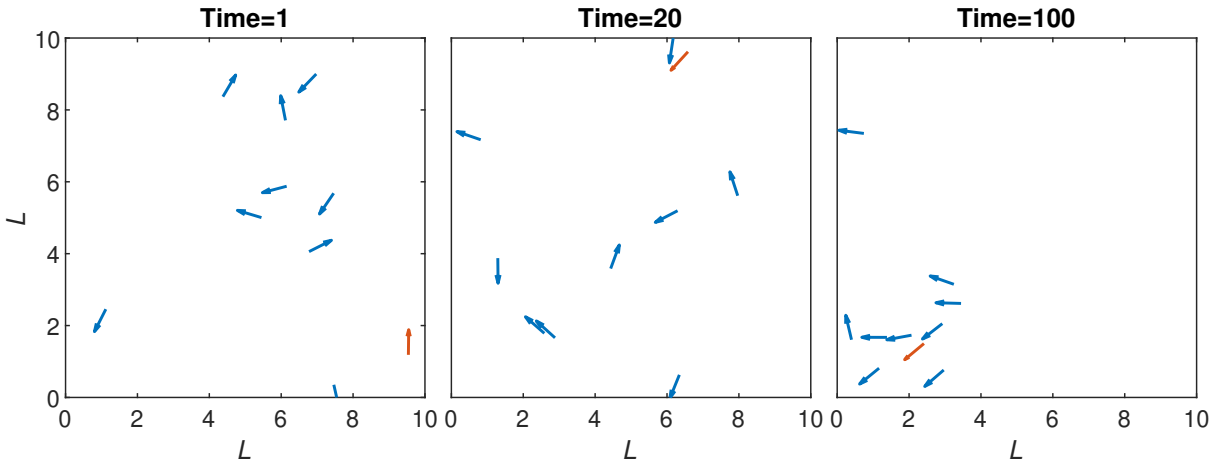

Figure 4.2: Snapshots of some moving particles at different times. The red arrow represents a 'leader' and the blue arrows are the 'followers'.

4.3 Concluding Remarks

The Vicsek model displays a transition to collective motion, where all particles have same interaction strength. But in our model (Eq.4.2), the interaction strength of all particles are not the same. Higher interaction strength has been used to some particles (we call it the leaders) compared to the other particles (known as followers), which enables us to study classification ability of some of the measures. As the motion of the cells depend on its neighboring, so there should be some transfer of information between particles. Based on the amount of transferred information, the identity of the particles can be inferred.

Chapter 5

Identifying influential particles

5.1 Introduction

This chapter describes the procedure that we use to identify leader and follower particles. We use the motion data of the particles (obtained from chapter 4) to classify the particles. Our main focus is to check the performance of transfer entropy. We also compare its performance with the linear measure cross-correlation.

At the beginning we define a classifier. Based on the classifier we compute true-positives and false-positives for each threshold values. Later, an receiver-operating characteristic (ROC) curve is drawn whose X-axis is the false-positives and Y-axis is the true-negatives. Finally, thew area under ROC curve (AUC) is obtained which quantifies the performance of classification.

5.2 Classifier

For inferring leader and follower particles we first define classifier. Based on this classifier we identify the leader and follower particles. For particle i, the classifier χ has the following form:

$$\chi^i = \frac{1}{N-1} \sum_{j(\neq i)} (I_{i \to j} - I_{j \to i}),$$

where $I_{i \to j}$ represents the influence that the particle i exhibits on particle j and N is the total number of particles in the system. Here I can be measured using different measures discussed in Chapter 2.

5.2.1 Identification

The value of χ is then used to identify leader and follower particles. First a threshold ϵ is selected. A particle for which classifier χ^i is higher than the threshold ϵ, we identify the particle i as a leader, otherwise a follower. Then this classification result is compared to the ground truth to obtain the how many particles have been identified correctly.

5.2.2 ROC curve and AUC

To show the classification performance of a classifier receiver-operating characteristic (ROC) curve is used [90,91]. An receiver operating characteristic (ROC) curve is a graph that shows the performance of a classifier at all classification threshold ϵ. It is a 2-dimensional plot whose x-axis represents false positive rate (FPR) and y-axis stands for true-positive rate (TPR) [92,93].

$$\mathbf{TPR} = \frac{TP}{(TP + FN)},$$

$$\mathbf{FPR} = \frac{FP}{(FP + TN)},$$

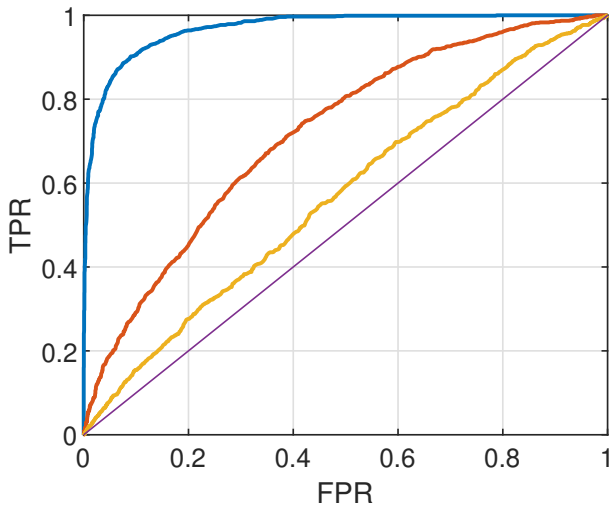

Figure 5.1: ROC curve. As the blue curve lies above all the curves, hence the classifier corresponds to the blue curve produces better classification score.

A good classifier has few false positives and a large number of true positives for a range of thresholds. Hence the top left corner of the ROC box is the point where TPR= 1.0 and FPR=0.0. This represents a perfect test. The closer the ROC curve gets to the top left corner, the better the test is overall. The closer the curve comes to the center diagonal line ($y = x$ *line*), the worse the test.

As we know, the closer the ROC curve reaches to the top left corner, the better the test. The closer a curve lies to the top left corner, the greater the area underneath it. Consequently, a larger area under the ROC curve implies a better test [94,95]. To quantify the accuracy of the classifier's performance and to compare the performance of different classifiers, area under ROC curve (AUC) has been used [94]. An AUC score of 1.0 means that that classifier accurately predicts the identities of the particles whereas a value of 0.5 means that the classifier has no class separation capacity whatsoever [96,97].

5.3 Some model parameters

To compare with TE, we also employ cross-correlation (CC) to elucidate between time series of the physical quantity associated with the particles' motion. We examine the performance of our scheme mainly for a 10-particle system of the modified VM in which one serves as a leader and the other nine particles as followers. For the computation of CC we use the values of $\boldsymbol{\theta}$ explicitly in Eq. (2.1). For TE, a probability distribution of discrete events is required, and in order to discretize θ we divide the orientation space $[0 : 2\pi]$ into n different discrete domains that are represented by unique symbols. Over time, each particle changes its orientation, which may result in a new symbol on its time series. We varied n from 2, 4, 6, 8, 10, and 12 and found that except $n = 2$ and 4 which are too small to capture the interaction domain and leader-follower identities, similar classification results are obtained for $n = 6 - 12$, therefore we chose $n = 6$ hereinafter.

We set the box size to L=10 a.u. with v_0 being 0.3 a.u., which means that each particle moves 3/100 of the box's length in a one-time step. It was found that in the range of $0.05 \leq v_0 \leq 0.9$ the classification of leader and follower particles are not affected by v_0 in our simulation running over a long enough number of time steps to yield convergence of TE and CC. At each noise level η_0, all analyses were performed with 1000 realizations. For each realization, the values of $\vec{r}_i^t$ and θ_i^t at $t = 0$ are chosen randomly in Eq. (4.1) and propagated for 100,000 time steps. The value of θ_i depends on the reference axis to which the angle is evaluated. It was confirmed with 100,000-time steps that the value of TE for $n = 6$ is the same within $\pm 6 \times 10^{-4}$/bits, irrespective of the choice of reference axis in our model simulation.

5.4 Results and discussion

In this section, we vary some parameters in the VM to check how these parameters affect the identification of leader and follower particles.

5.4.1 Interaction time-scale and delay-time

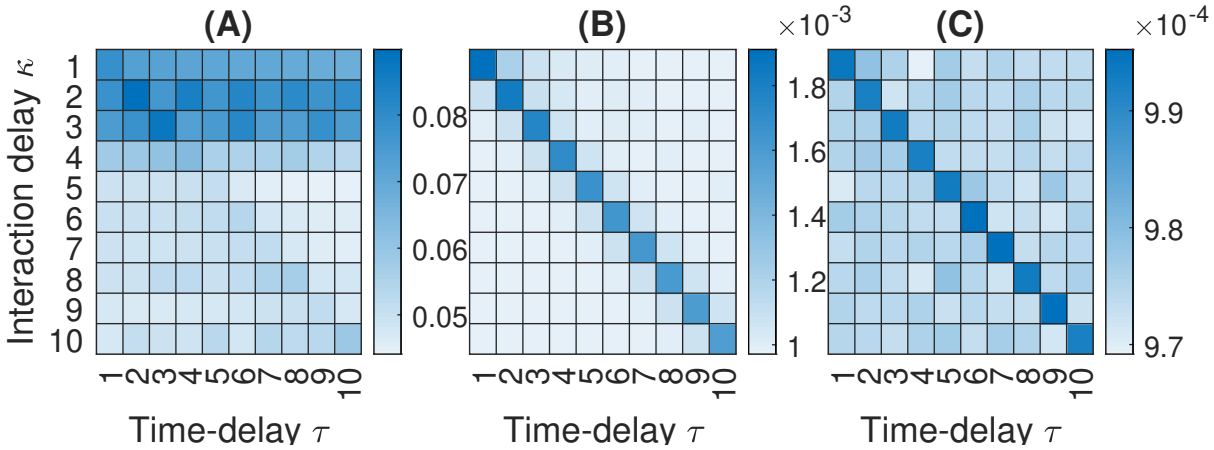

Figure 5.2: Average TE landscape for different interaction timescale κ and delay time τ at noise level (A) $\eta_0 = 0.1\pi$, (B) $\eta_0 = 1.2\pi$, and (C) $\eta_0 = 1.8\pi$. Note that each color grade has different scale.

Figure (5.2) shows the average TE landscape as a function of interaction timescale κ (intrinsic to the system) and delay time τ (to compute TE) at (A) $\eta_0 = 0.1\pi$, (B) $\eta_0 = 1.2\pi$, and (C) $\eta_0 = 1.8\pi$. The average TE, averaged over all pairs of particles, has a peak when interaction timescale κ in Eq. (4.2) and delay time τ in Eqs. (2.11) and (2.1) are the same. This suggests that the maximum information flow between particles can be obtained if the appropriate delay time τ is used while computing TE. Hence a delay time τ that makes the average TE maximized with respect to τ can be used to infer the actual interaction timescale between particles. Note at the low noise level in the average TE landscape that periodicity is perceived in τ, as shown in Fig. 5.2 (A), especially at $\kappa = 1 - 4$. This is due to the fact that a particle i at time t can interact with another particle j at time $(t + \kappa)$ directly. But particle i can interact indirectly with the particle j (through a third particle, say k) at time $(t + 2\kappa)$. As a result the influence of particle i on particle j can be observed at $t = \kappa$, $t = 2\kappa$, $t = 3\kappa$, and so on. Since the average TE exhibits a peak when τ is equal to the delay in the interaction between particle i and particle j, consequently TE peaks at $\tau = \kappa$, $\tau = 2\kappa$, $\tau = 3\kappa$, and so on. As the

interaction timescale κ increases, say more than 4, or as the noise level increases, such periodicity becomes smeared out due to thermal fluctuation. In the following analysis, we set $\kappa = 1$ and $\tau = 1$ unless stated otherwise.

5.4.2 Distributions of classifiers

Figure 5.3 shows the distribution of the classifier χ using TE, χ_{TE}, at three different noise levels. As an overall trend, the classifier $\chi^{(i)}$ when i is a leader is greater than that when i is a follower over different noise levels. At low noise ($\eta_0 = 0.1\pi$) and very high noise (1.8π), the distributions of χ for leaders and followers significantly overlap each other, which makes classification difficult. However, the distributions are well-separated and classifying leader or follower is much easier from the orientation dynamics at moderate noise level ($\eta_0 = 1.2\pi$). At very low noise levels, the amount of TE depends

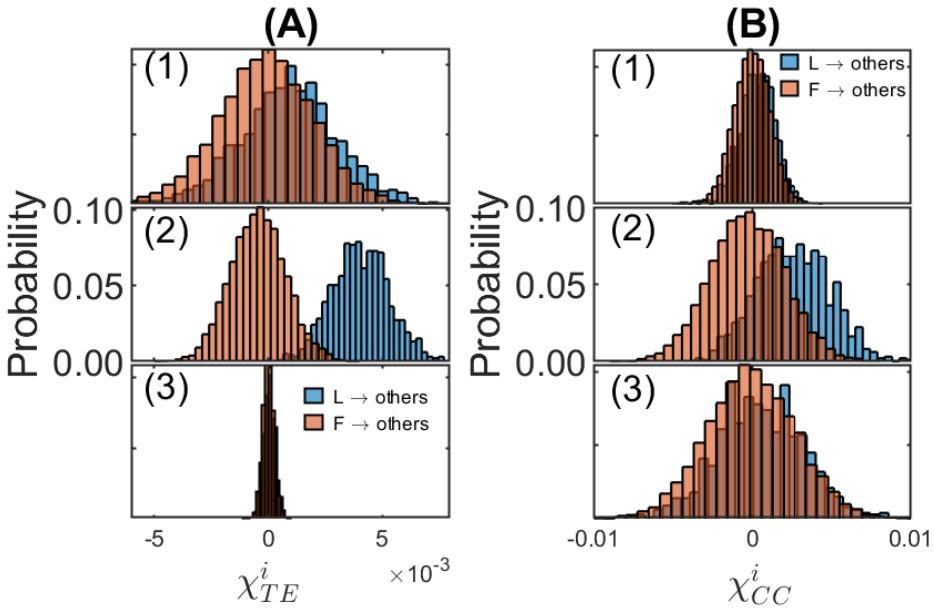

Figure 5.3: Distributions of the classifiers (A)χ_{TE}(/bits), (B) χ_{CC}(/bits) from the leader to the others, and that from a follower to the others at different noise levels η_0, (1) 0.1π, (2) 1.2π, and (3) 1.8π, respectively.

heavily on initial configurations $\{\vec{r}_i^0\}$. Namely, the probability of encountering within the interaction domain largely depends on $\{\vec{r}_i^0\}$ at very low noise and, moreover, follower and leader particles are promptly aligned to the same orientation as soon as they encounter one another. This results in the same symbols in their time series perpetually, making $p(y_{t+1}|y_t, x_t)$ quickly converged, while its values are largely dependent on $\{\vec{r}_i^0\}$. Due to

this $\{\vec{r}_i^0\}$ dependence, distributions of the classifier $\chi_{\rm TE}$ have larger variances at very low noise (Fig. 5.3 A(1)), making it difficult to distinguish followers and leaders. As the noise increases, $p(y_{t+1}|y_t, x_t)$ becomes less dependent on the $\{\vec{r}_i^0\}$ and the two distributions start to deviate from each other (Fig. 5.3 A(2)). However, at very high noise, the two distributions become indistinguishable with much smaller variances (Fig. 5.3 A(3)).

On the other hand, at a very low level of noise, the motion of both leader and follower has a strong correlation, irrespective of whether two particles interact or not. Hence the distribution of the classifier $\chi_{\rm CC}$ has lower variance at very low noise (Fig. 5.3 B(1)). As the noise increases, the motion of the follower becomes more dependent on the leader that produces higher CC from leader to follower compared to that from follower to leader. Hence two distributions start to drift from each other (Fig. 5.3 B(2)). But at a very high noise level, the distributions have higher variance which makes the classification very difficult (Fig. 5.3 B(3)).

The efficiency of χ to classify the leader-follower relationship is estimated as follows: by choosing a threshold γ along the χ axis (e.g., Fig. 5.3), the particle whose $\chi^{(i)}$ is larger (smaller) than the threshold γ is classified as leader (follower), and the classification is compared to the ground truth (defined by different w_{ij}) to obtain the true-positive rate (TPR) and the false-positive rate (FPR) for the γ. We then plot TPR versus FPR at different γ's to obtain a receiver-operating characteristic (ROC) curve [34]. We use the area under the ROC curve (AUC) score for the quantification of the performance of χ. Here an AUC score of 1.0 (0.5) means that the classifier accurately predicts the identities of particles (predicts them as randomly as a coin toss).

5.4.3 AUC for different interaction strength of leader

Figure (5.4) represents the effect of interaction strength of leader on followers w_{FL} (from now on we will use w to represent w_{FL}) in identifying leader and follower particles as a function of noise (η_0). Figure (5.4)(A) shows the performance of TE whereas Fig. (5.4)(B) is for CC. As an over all trend, TE outperforms CC in identifying leader and follower. But at very high noise level the performance of CC is slightly better than TE. Both TE and CC perform better at mid level of noise. At both very low and very high level of noise, it is very difficult to distinguish leader and follower particles. At very low noise, the amount of transferred information between leader and follower is very low

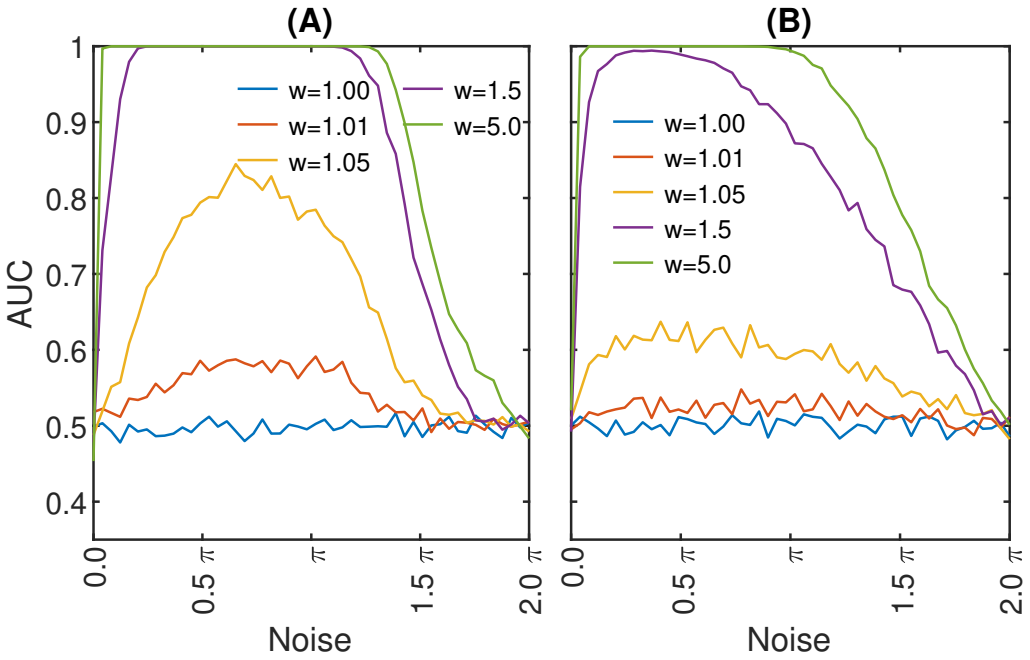

Figure 5.4: (A) AUC using TE. (B) AUC using CC. As an overall trend both TE and CC perform better at moderate noise levels in identifying leader and follower particles and TE outperforms CC.

and because of that insufficient information transfer, classification is very hard. Here we varied w.

In our model, leader is based on the interaction strength $w = w_{FL}$ and the interaction strength of followers w_{FL} and w_{FF} is set equal to 1. Hence $w = 1$ means that there is no leader in the system. Since there is no leader in the system and our classifier looks for leader, hence its performance is very poor which results in AUC value close to 0.5. As w increases, leader particle becomes more influential. Consequently the performance of both classifiers improve which results in higher AUC values.

5.4.4 Classification score for different system size

Figure (5.5) represents the classification results for the different number of birds at different noise levels keeping box size L fixed, which means that density increases as the number of birds increases. In Fig. 5.5(A) we use TE and in Fig. 5.5(B) CC has been used. In both the figures, the general propensity is that as the number of birds increases classification score decreases. Because of the information flow between two followers and also the indirect information flow from a follower to another follower through the leader, the influence of a leader on particular follower decreases. Since CC cannot capture non-

linear relationship, hence the performance of CC is very poor for the higher number of birds.

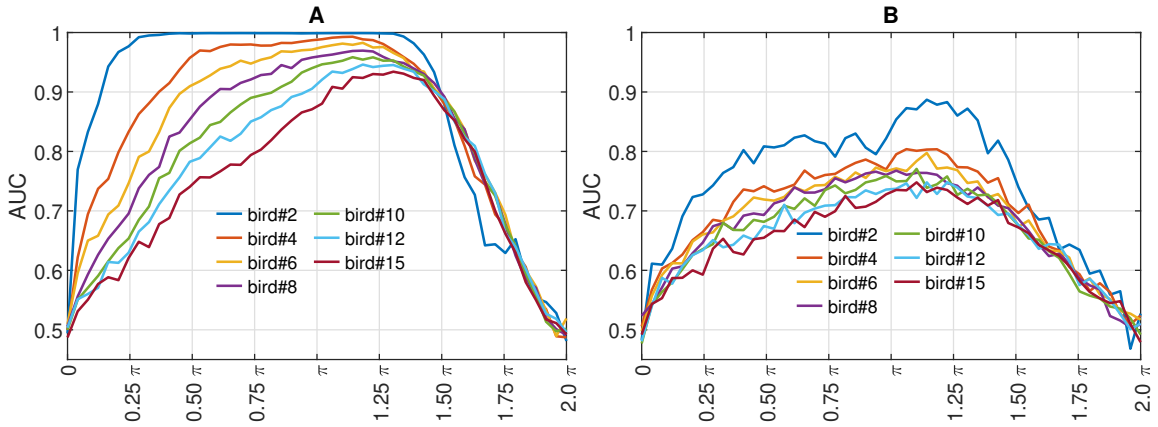

Figure 5.5: Classification score of different system size at different noise levels using (A) TE, (B) CC keeping fixed box size. Classification is random at both low and high level of noise. In general AUC decreases with number of birds.

Figure (5.6) represents the classification results for a different number of birds at different noise levels keeping density fixed, which means that L increases as the number of birds increases. Similar to Fig. (5.5), the classification score decreases as the number of birds increases. When the density of particles is much higher, there is more chance of indirect interaction between leader and followers. A dissimilarity between Fig. (5.5) and Fig. (5.6) can be observed at a higher level of noise. At the high level of noise, AUC is independent of the number of birds for the fixed box system whereas AUC decreases gradually for the fixed density system.

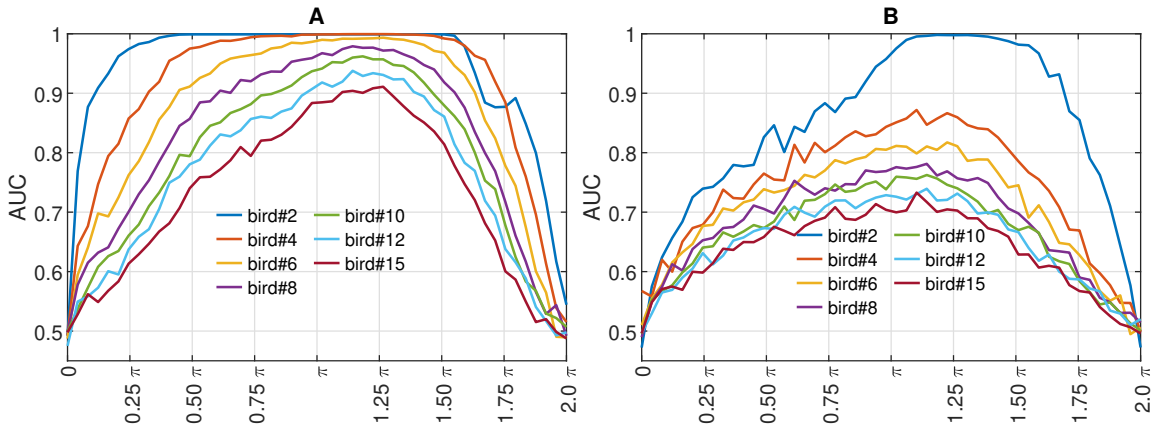

Figure 5.6: Classification score of different system size keeping density fixed. (A) Classification performance of TE. (B) Classification performance of CC. AUC decreases with number of birds.

5.4.5 Effect of data length on classification score

Now let us check the effect of interaction time (data length) and some other physical quantities on the classification performance. Figure 5.7 shows the AUC landscape as a function of data length and leaders' interaction strength (w_{LF}). It has been found that the AUC score increases with T. For shorter T, due to insufficient sampling in characterizing leader-follower interaction relationship, the distributions of χ_{TE} from leader to the others and from follower to the others have higher variance as shown in Fig. 5.8. As the data length increases, particles get more time to interact with each other which results in higher information flow between particles. As a result classification score also increases.

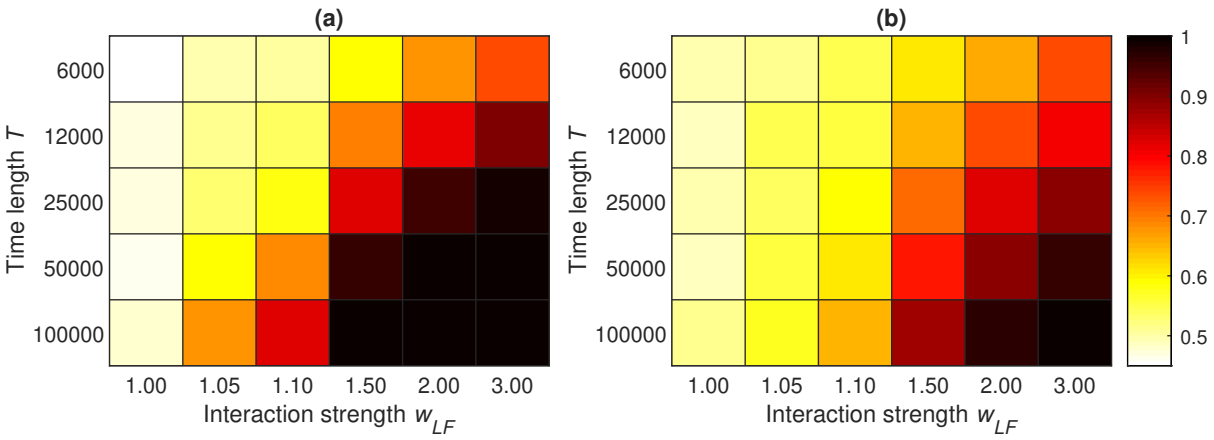

Figure 5.7: AUC for different data length and interaction strength w. (A) Using TE. (B) Using CC. These figures are for noise level $\eta_0 = 1.2\pi$. It has been found that as the data length and interaction strength w increase, classification score also increases.

Figure 5.9 shows the AUC landscape for different data length and interaction radius R. Interaction radius represents the maximum distance up to which a particle can interact with another particle. Hence smaller interaction radius R of a particle means that the particle has a smaller area of influence. Consequently, the particle shares information only with the particles which are within that small region around it. So it cannot share information with the particles which are outside the region. As the interaction radius R increases, the area of influence also increases. At the same time, information flow also increases. Hence classification score increases as the interaction radius R increases.

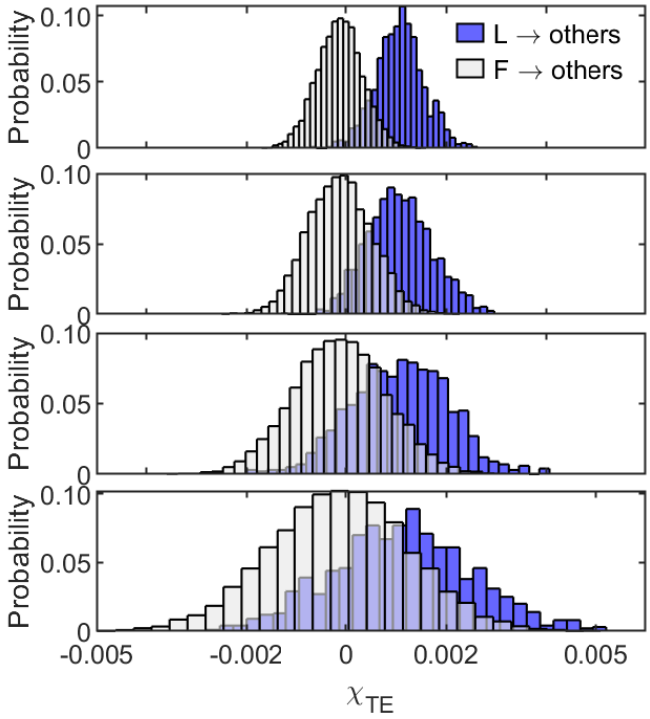

Figure 5.8: Distributions of the classifiers χ_{TE} from the follower to the others, and those from a follower to the others for $R = 3, \eta_0 = 1.2\pi$, and (a)$T = 100,000$ arb. units, (b) $T = 50,000$, (c)T =25,000, and (d) $T = 12,000$.

5.5 Concluding Remarks

In this chapter, we varied different physical quantities from the modified VM and examined the effect of these parameters on the classification score. We use two measures in the form of TE and CC to identify leader and follower particles. It has been found that because of the initial position dependency, the classification score is very low at a very low noise level. At a moderate level of noise, classification is relatively easier. But at the very high level of noise, it is very difficult to identify leader and follower particles. We also found that as the number of particles increases classification becomes difficult because of the indirect interaction (through another particle) between particles. Also, the classification score increases as the value of interaction strength w of the leader increases. Data length also plays a vital role in the classification score. AUC value increases along with the data length and interaction radius R.

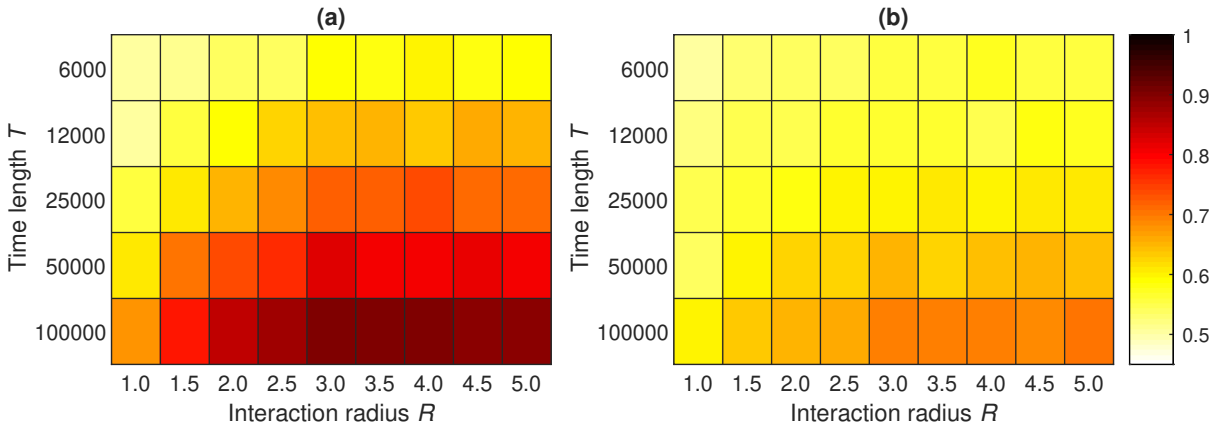

Figure 5.9: AUC for different data length and interaction radius R. (A) Using TE. (B) Using CC. It has been found that as the data length and interaction radius R increase, classification score also increases.

Chapter 6

Identification of Interaction radius

6.1 Introduction

Systems composed of locally interacting particles or agents such as birds, fish, and cells show spontaneous, spatiotemporal, collective behaviors as a whole [98–100]. Exploration of the underlying mechanisms and principles of such coordination of agents has been made in experimental [7,8,101], and theoretical [34,102–104] studies.

Interactions among agents are an important feature of spatiotemporal coordination of the agents. In most cases two agents cannot communicate over infinite distance, and thus it is natural to suppose the interactions are inherently local. Indirect influence, however, can permeate globally even when individual interactions are local [105–107]. The cascading of local influence to a global level is highly dependent on the radius of interaction between individual pairs of agents. Thus the interaction radius is an important quantity in the field of collective behavior analysis as it defines the interaction network between agents [108]. In the VM, the direction of movement of each particle is determined by the average direction of its neighboring particles which are located within a circle of an interaction radius R centered on the particle. The interaction radius R plays a vital role for the synchronization of the particles. It was found that the smallest interaction radius for synchronization of the ensemble is $\sqrt{\log(m)/\pi m}$, where m represents the population size of the VM [109]. Depending on the law of interaction in a general system, the distance at which influence is most likely to occur is critical for understanding the outcome of the system.

Inferring differential influence in interactions and leader-follower relationships from

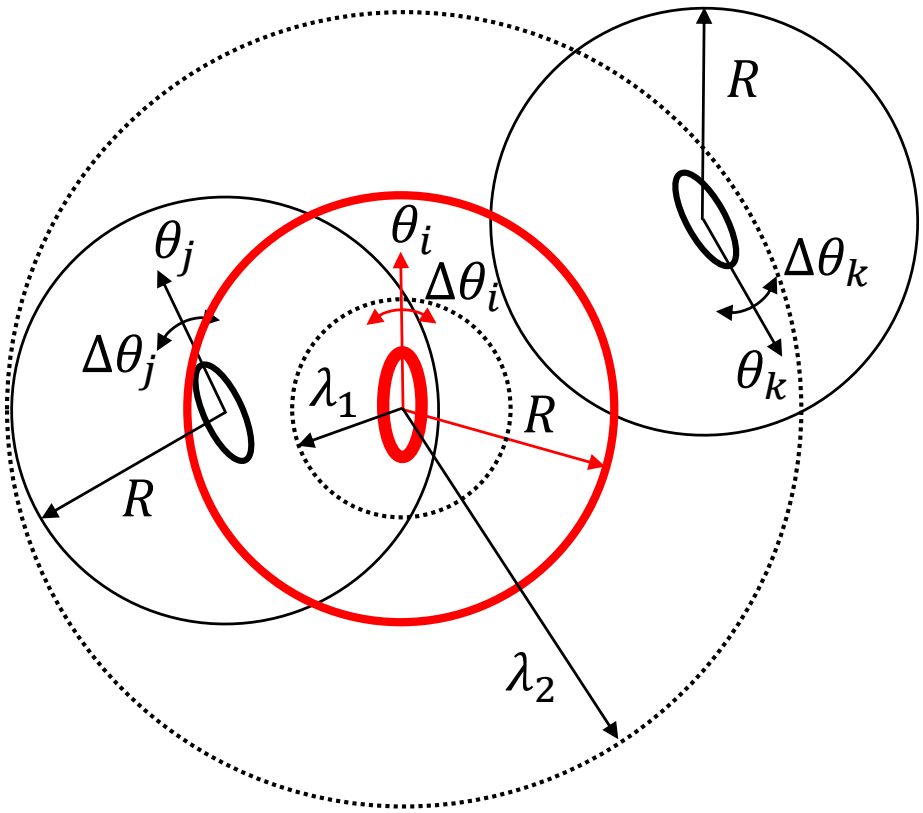

Figure 6.1: Schematic diagram of cutoff distance.

empirical data such as an ensemble of trajectories of particles is thus an intriguing subject. Various measures such as cross-correlation and extreme-event synchronization [4], as well as information-theoretic measures, such as time-delayed mutual information [28,110,111], transfer entropy [29,112,113], and causation entropy [53,114] have been used on time-lapse motion data to quantify the direction of influence between pairs of particles.

In this chapter we present an information-theoretic scheme to estimate the underlying interaction domain or distance between composite particles from an ensemble of trajectories in many particle systems. In collective behavior analysis, transfer entropy (TE) was used to study predator-prey interactions [115] and leader-follower identification in both simulation data [4,34] and experimental data [35]. We show for a modified VM that the classification of leaders and followers is significantly improved by using the identified optimal interaction domain compared to the case in which transfer entropy is computed over all pairs irrespective of distance.

6.2 Cutoff distance

Knowledge of the interaction domain greatly improves the classification of leaders and followers[116,117]. In practice, however, one does not know the interaction domain for a group of animals, cells, or birds a priori. To deduce it from an ensemble of trajectories, the 'cutoff distance variable' λ was introduced [117]. In this problem setting, the interaction domain is considered as a circle of radius R, which is unknown, however, in general the same technique can be applied to infer an interaction domain of any shape. Then for the estimation of TE, the cutoff distance λ is defined as a distance up to which the interactions between particles are taken into consideration [Fig. 6.1]. In other words, for the estimation of TE from the 'symbolic time series' of a particle to another, the probability distributions are estimated only at the time instance t when the distance between those two particles is less than the cutoff distance λ. For a fixed cutoff distance λ, TE from a particle X to another particle Y has the following form:

$$\begin{aligned} \mathrm{TE}_{X\to Y}(\lambda) = \sum_{y_{t+\tau},y_t,x_t} & p(y_{t+\tau},y_t,x_t|d\le\lambda)\times \\ & \log_2\left(\frac{p(y_{t+\tau}|y_t,x_t,d\le\lambda)}{p(y_{t+\tau}|y_t,d\le\lambda)}\right) \end{aligned} \tag{6.1}$$

where x_t, y_t, and $d = |\vec{r}_X^{\,t} - \vec{r}_Y^{\,t}|$ represent the orientation of the particles X and Y at time t, and the distance between the particles X and Y at time t, respectively.

Finally the value of λ is varied and TE between particles is computed as a function of λ. Whenever there is no mention of a cutoff distance λ, it means that the distance information between particles is not considered which is the conventional way of TE computation [35].

In a group of particles, the motion of a particle is influenced by other particles lying within its interaction domain. When two particles enter into their interaction domain, they share or transfer information, resulting in some change in their movements. Such information flow should decrease as the distance between the two becomes greater than the interaction domain, and goes to zero at the limit of the distance goes to infinity because of loss of interactions. For the transfer entropy (Eq. 6.1) where all pairs satisfying $d \le \lambda$ are taken into account, when the cutoff distance λ exceeds the underlying interaction domain

R, $\mathrm{TE}_{X\to Y}(\lambda)$ should decrease because it takes into account not only interacting pairs of particles but also non-interacting, independently-moving pairs, which should "dilute" the information flow among interacting pairs in the system. The magnitude of negative gradient of the transfer entropy with respect to λ after exceeding R is expected not to be a constant along λ but the magnitude becomes get smaller as λ gets larger. When the two particles are within interaction domain, more or less their motility is influenced to each other so that the transfer entropy is expected to have some finite value with some fluctuation for $\lambda \leq R$.

6.3 Effect of cutoff distance and interaction radius in classification score

Figures 6.2 (A) and 6.2 (B) show the AUC landscape as a function of noise levels and cutoff distance for TE for interaction radius $R = 2$ and $R = 4$, respectively. Figures 6.2 (C) and 6.2 (D) represent the same landscape but for CC. One can see that χ_{TE} outperforms χ_{CC} and many AUC scores are found to be close to unity for χ_{TE} while the maximum value of χ_{CC} is only 0.817. As shown in Fig. 6.2 (A) and 6.2 (B), both at very low noise and at very high noise, the AUC score of χ gets smaller. One striking consequence is the existence of maximum peak near the cutoff distances $\lambda = 2$ and $\lambda = 4$ in the AUC landscape of both χ_{TE} and χ_{CC} (note that the actual interaction radius R used for the trajectory calculations in Figs. 6.2 (A) and 6.2 (C) is 2, whereas in Figs. 6.2 (B) and 6.2 (D) it is 4. When $\lambda < R$, it implies that, even though two particles interact, useful information is not taken into account for computing the classifier χ, making AUC lower. Conversely, when $\lambda > R$, it implies that unnecessary information concerning non-interacting particles is taken into account, yielding lower AUC than the maximum. This demonstration manifests that the best classification occurs when cutoff distance λ and the underlying interaction radius R are the same.

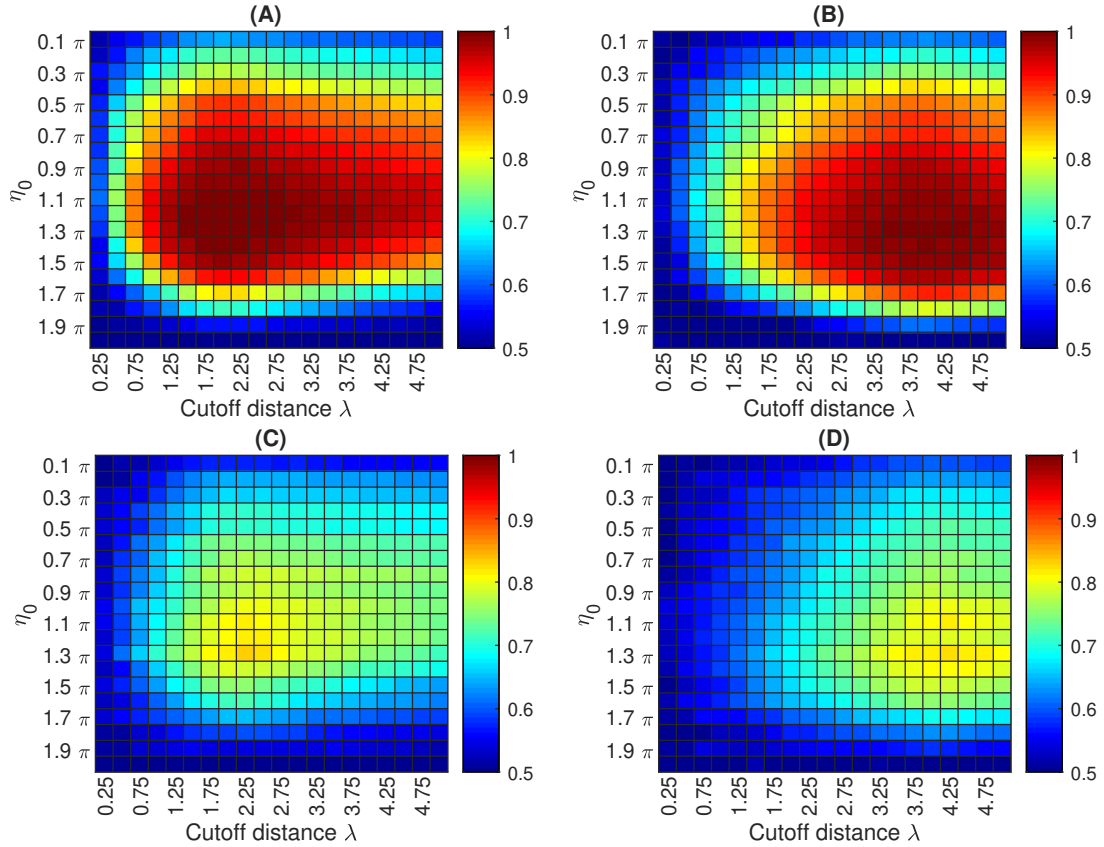

Figure 6.2: (A) and (B) AUC landscape of the classifier χ_{TE} at different noise levels and cutoff distance λ. (C) and (D) AUC landscape of the classifier χ_{CC} at the same noise levels used in the above two figures. At very short λ there exist situations where two particles rarely enter each other's interaction domains, and to avoid the unclassified cases $\lambda < 0.25$ a.u. are excluded. Also at very low η_0 there were cases where the symbols of a particle seldom change in time, which makes CC undefined, and, hence, $\eta_0 < 0.1\pi$ are excluded. The actual interaction radius R used for ensemble of trajectories in (A) and (C) is 2 a.u., and 4 a.u. in (B) and (D).

6.4 Comparison between cutoff and without cutoff schemes

Figure 6.3 compares the classification performance of cutoff, and without-cutoff techniques in identifying leader and follower particles based on the AUC values. Here TE is computed using two different λ values. First, we set $\lambda = R$ (denoted as cutoff distance scheme) and then $\lambda = 5\sqrt{2}$ (denoted as without-cutoff). The cutoff technique clearly possess higher AUC scores compared to the without-cutoff technique for both TE and CC. It manifests that the cutoff distance scheme outperforms the conventional scheme without cutoff distance used in previous studies [4,34]. It was found that the classification performance is optimized in the AUC landscape once the underlying interaction radius R

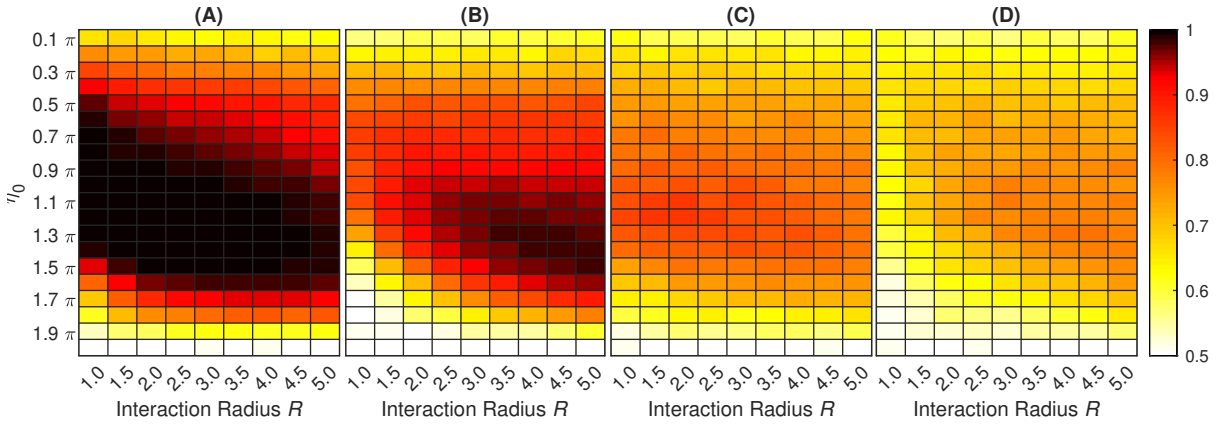

Figure 6.3: AUC landscapes as a function of noise levels η_0 and the interaction radius R obtained by our cutoff distance scheme compared to those by the conventional scheme without cutoff distance. (A) The classifier χ_{TE} with cutoff distance ($\lambda = R$), (B) χ_{TE} without cutoff distance ($\lambda = 5\sqrt{2}$), (C) χ_{CC} with the cutoff distance, and (D) χ_{CC} without the cutoff distance. Since $\lambda = 5\sqrt{2}$ corresponds to the without-cutoff distance scheme, and the AUC values with and without cutoff distance schemes converge to the same values as λ approaches to $5\sqrt{2}$.

is identified. The AUC estimate, however, requires the knowledge of ground truth and, in real applications, one cannot know the ground truth *a priori*.

6.5 Interaction radius identification

6.5.1 A simple binary model

We now derive a simplest analytic model to manifest the relationship between the interaction domain R and transfer entropy $\mathrm{TE}_{X\to Y}(\lambda)$ as follows: Consider an agent X represented as a binary random process, i.e. $X_t = 0$ with probability $\frac{1}{2}$ and $X_t = 1$ with probability $\frac{1}{2}$ (similar to "fair coin toss"), whose value influences to another agent's future value Y_{t+1} depending on an auxiliary variable Z_t. Z_t is a continuous, uniformly distributed variable on the interval $[0, L]$, and $Y_{t+1} = X_t$ whenever $Z_t \leq R$, otherwise Y_{t+1} is a binary random process as X, where $Y(0) = 0$. Namely, $R(\leq L)$ serves a similar function as the interaction radius in the VM model, and Z does as the distance between a pair of particles. In a more extreme sense of the VM, Y_{t+1} depends deterministically on X_t whenever Z_t below R and Y_{t+1} is completely random otherwise. Therefore intuitively, the function $\mathrm{TE}_{X\to Y}(\lambda)$ should have a similar shape to that of the VM for the binary system, especially near the value $\lambda = R$. The benefit of the binary system is that we can compute $\mathrm{TE}_{X\to Y}(\lambda)$ analytically.

To analytically derive $\mathrm{TE}_{X\to Y}(\lambda)$ for the binary system described above we must compute Eq. 6.1 for the binary system. Using the chain rule for conditional probability and replacing x_t, y_t, and d by X_t, Y_t, and Z_t, respectively, from our binary system, Eq. 6.1 can be written as

$$\mathrm{TE}_{\mathrm{X}\to\mathrm{Y}}(\lambda) = \sum_{Y_{t+1}}\sum_{X_t}\sum_{Y_t} P(Y_{t+1}, Y_t, X_t|Z_t \le \lambda)\times \log_2 \frac{P(Y_{t+1}, Y_t, X_t|Z_t \le \lambda)P(Y_t|Z_t \le \lambda)}{P(Y_{t+1}, Y_t|Z_t \le \lambda)P(Y_t, X_t|Z_t \le \lambda)}. \tag{6.2}$$

$P(Y_t|Z_t \le \lambda) = P(Y_t) = \frac{1}{2}$ since $Y_t \perp Z_t$ (where $A \perp B$ denotes A and B are independent), and $P(Y_t) = \frac{1}{2}$ by definition. Since $X_t \perp Y_t$ and $Y_t \perp Y_{t+1}$, $P(Y_t, X_t|Z_t \le \lambda) = P(Y_t, X_t) = P(Y_t, Y_{t+1}|Z_t \le \lambda) = P(Y_t, Y_{t+1}) = \frac{1}{4}$. $P(Y_{t+1}, Y_t, X_t|Z_t \le \lambda) = P(Y_{t+1}, X_t|Z_t \le \lambda)P(Y_t) = \frac{1}{2}P(Y_{t+1}, X_t|Z_t \le \lambda)$ since $Y_t \perp (Y_{t+1}, X_t, Z_t)$ and $P(Y_t) = \frac{1}{2}$. Thus, λ dependency of $\mathrm{TE}_{\mathrm{X}\to\mathrm{Y}}(\lambda)$ arises from $P(Y_{t+1}, X_t|Z_t \le \lambda)$. Now we first compute a general form of $P(Y_{t+1}, X_t)$ which holds irrespective of λ and then re-write it in terms of λ shortly thereafter. Here $P(Y_{t+1}, X_t)$ can be written as

$$P(Y_{t+1}, X_t) = P(Y_{t+1}, X_t, Z_t \le R) + P(Y_{t+1}, X_t, Z_t > R),$$

and by chain rule of probability distributions,

$$P(Y_{t+1}, X_t) = \frac{1}{2}[P(Y_{t+1}|X_t, Z_t \le R)P(Z_t \le R)+ P(Y_{t+1}|X_t, Z_t > R)P(Z_t > R)],$$

since $X_t \perp Z_t$ and $P(X_t) = 1/2$. When $\lambda \le R$, $Y_{t+1} = X_t$ by the model setting so that $\mathrm{TE}_{X\to Y}(\lambda)$ is always unity. Thus in the following we focus on the case of $\lambda > R$. $P(Y_{t+1}, X_t|Z_t \le \lambda)$ can be written as

$$\begin{aligned} P(Y_{t+1}, X_t|Z_t \le \lambda) = \frac{1}{2}[&P(Y_{t+1}|X_t, Z_t \le R, Z_t \le \lambda)\times \\ &P(Z_t \le R|Z_t \le \lambda)+ \\ &P(Y_{t+1}|X_t, Z_t > R, Z_t \le \lambda)\times \\ &P(Z_t > R|Z_t \le \lambda)]. \end{aligned}$$

Here the first and second terms in the right hand side of above equation, respectively, correspond to the case of the "distance" Z_t being within the interaction "radius" R ($Z_t \leq R$) and that of Z_t being larger than R where there exists no interaction ($R < Z_t \leq \lambda$). $P(Z_t \leq R | Z_t \leq \lambda)$ and $P(Z_t > R | Z_t \leq \lambda)$ are equal to $\frac{R}{\lambda}$ and $1 - \frac{R}{\lambda}$, respectively, since we have chosen a uniform distribution of Z_t. In the first term of the interaction regime, $P(Y_{t+1} | X_t, Z_t \leq R, Z_t \leq \lambda)$ is divided into two cases, one case where Y_{t+1}, X_t are the same and another case where Y_{t+1}, X_t are different. In the case that they are the same, $P(Y_{t+1} | X_t, Z_t \leq R, Z_t \leq \lambda) = P(Y_{t+1} | X_t, Z_t \leq R) = 1$, and in the case where Y_{t+1}, X_t are different, $P(Y_{t+1} | X_t, Z_t \leq R, Z_t \leq \lambda) = 0$ by definition of the model. In the second term of the no-interaction regime, $P(Y_{t+1} | X_t, Z_t > R, Z_t \leq \lambda) = P(Y_{t+1}) = \frac{1}{2}$ since $Y_{t+1} \perp X_t$ for $Z_t > R$.

The above equation implies, in the computation of $P(Y_{t+1}, X_t | Z_t \leq \lambda)$, that the contribution of independently-moving pairs of Y_{t+1} and X_t that do not "interact" to each other becomes dominated, as λ gets larger than the interaction domain R.

Finally, in order to compute $\mathrm{TE}_{X \to Y}(\lambda)$, we plug in all the values back into Eq. 6.2 and obtain the equation for TE as follows:

$$\mathrm{TE}_{X \to Y}(\lambda) = \frac{\lambda + R}{2\lambda} \log_2 \frac{\lambda + R}{\lambda} + \frac{\lambda - R}{2\lambda} \log_2 \frac{\lambda - R}{\lambda} \tag{6.3}$$

whenever $\lambda > R$. When $\lambda \leq R$, $\mathrm{TE}_{X \to Y}(\lambda)$ is equal to 1 since the dynamics for Y_{t+1} $Z_t \leq R$ is always deterministically dependent on X_t.

The function is then differentiated with respect to λ to obtain $\frac{\mathrm{dTE}}{d\lambda}$. The derivative $\frac{d\mathrm{TE}_{X \to Y}}{d\lambda}$ when $\lambda > R$ is given by

$$\frac{d\mathrm{TE}_{X \to Y}}{d\lambda} = \frac{R}{2\lambda^2} \log_2 \frac{\lambda - \mathrm{R}}{\lambda + \mathrm{R}}. \tag{6.4}$$

Eq. 6.3 suggests that, as the cutoff distance λ increases longer than the underlying interaction domain, $\mathrm{TE}_{X \to Y}(\lambda)$ monotonically decreases (i.e., downward-convex) and converges to some nonzero finite value when the space size L is finite (because $\lambda \leq L$) otherwise zero. This further implies that conventional computation procedure of transfer entropy taking into account all pairs of agents, including non-interacting pairs, should "dilute" the contribution of interacting pairs in the system, which can yield some misleading interpretation for the relationship among the agents.

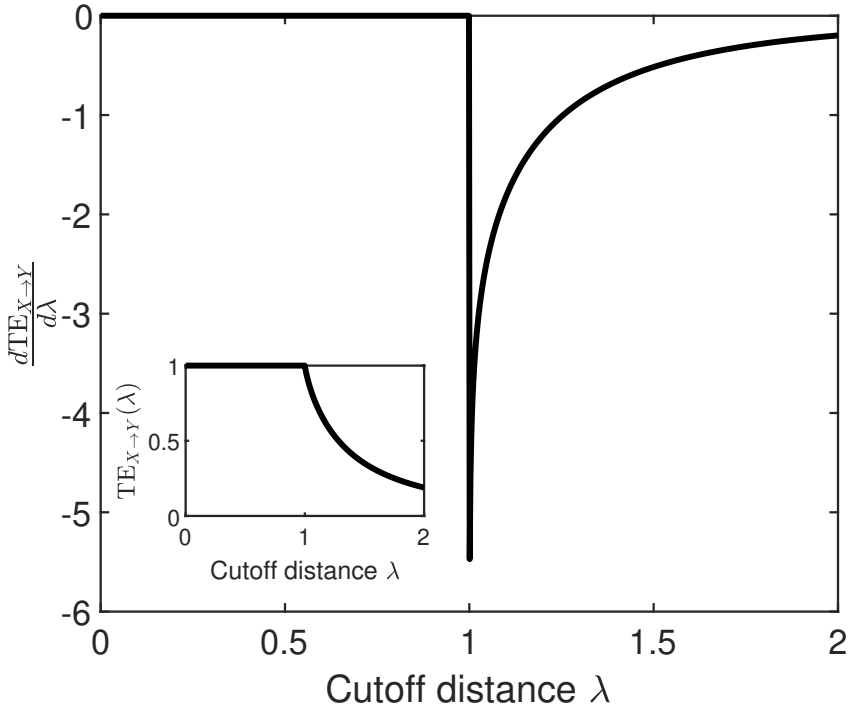

Figure 6.4: $\mathrm{TE}_{X\to Y}(\lambda)$ and $\frac{d\mathrm{TE}_{X\to Y}}{d\lambda}$ of the simplest binary model with $L = 2$, $R = 1$.

The values of $\mathrm{TE}_{X\to Y}(\lambda)$ and $\frac{d\mathrm{TE}_{X\to Y}}{d\lambda}$ for the parameters $R = 1$ and $L = 2R$ are shown in Fig. 6.4. The function $\mathrm{TE}_{X\to Y}(\lambda)$ has a clear visible kink at $\lambda = R$, which is made more apparent by looking at the derivative function $\frac{d\mathrm{TE}_{X\to Y}(\lambda)}{d\lambda}$ of the inset in Fig. 6.4. By computing $\mathrm{TE}_{X\to Y}(\lambda)$ from trajectory data and adequately detecting this kink, one may readily infer the interaction domain without any prior knowledge about the system, as we will demonstrate in the following section.

6.5.2 Identification of interaction domain from trajectories

How can one infer the underlying interaction domain or radius solely from ensembles of trajectories? Figure 6.5(A) exemplifies how the TE averaged over the ensemble of trajectories (denoted by $\langle\mathrm{TE}\rangle_\lambda$) behaves with respect to cutoff distance λ at $\eta_0 = 1.20\pi$ for different R, associated with their derivatives in Fig. 6.5(B). The corresponding averaged CCs ($\langle\mathrm{CC}\rangle_\lambda$) and their derivatives are also shown in Fig. 6.5(C) and (D). The value of $\langle\mathrm{CC}\rangle_\lambda$ as a function of λ is found to be almost flat up to $\lambda \simeq R$ and then it decreases gradually beyond the underlying interaction radius R. While $\langle\mathrm{TE}\rangle_\lambda$ drops for a region of very short cutoff distance $\lambda < 1$, $\langle\mathrm{TE}\rangle_\lambda$ is almost flat up to the interaction distance R, and then $\langle\mathrm{TE}\rangle_\lambda$ again decreases when $\lambda > R$. One can see the local minimum in the derivative (i.e., inflection point in $\langle\mathrm{TE}\rangle_\lambda$ and $\langle\mathrm{CC}\rangle_\lambda$) located near the underlying

interaction radius R.

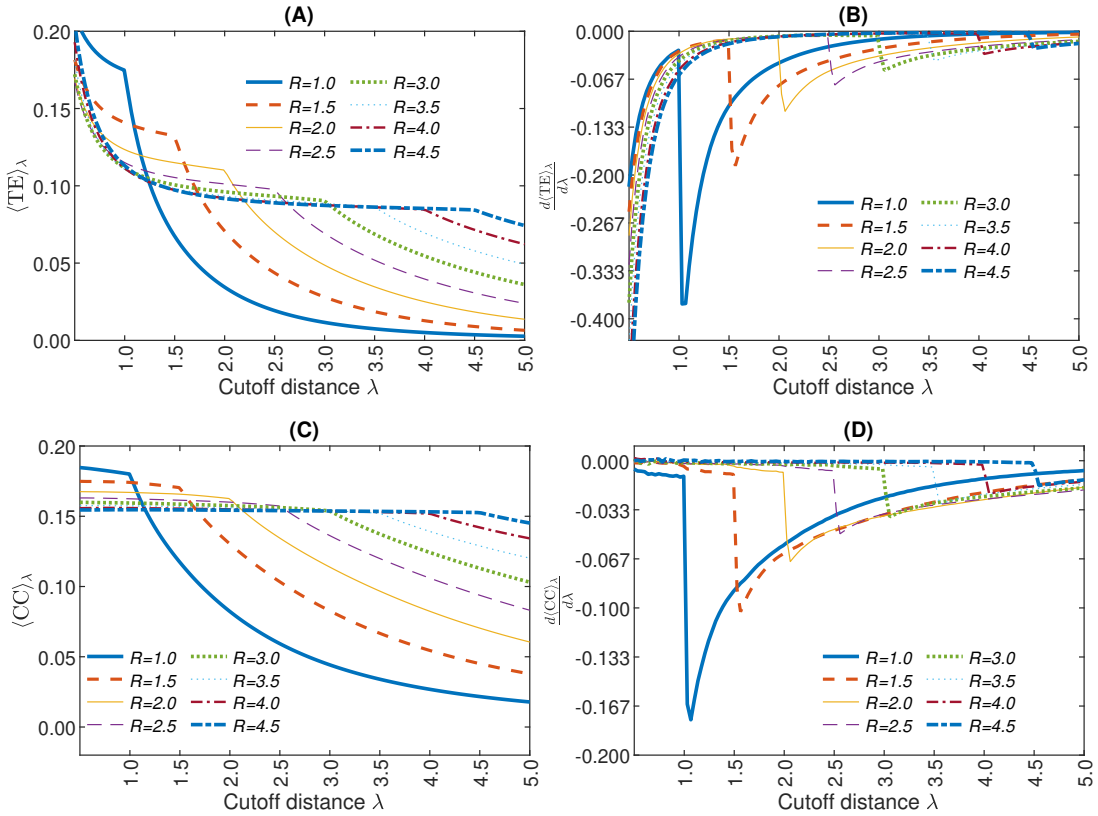

Figure 6.5: Dependence of averaged TE and CC over the ensembles on cutoff distance λ for different interaction radius R at noise level $\eta_0 = 1.2\pi$. (A) the averaged TE $\langle \mathrm{TE} \rangle_\lambda$, (B) the derivatives of $\langle \mathrm{TE} \rangle_\lambda$ with respect to λ, (C) the averaged CC $\langle \mathrm{CC} \rangle_\lambda$, (D) the derivatives of $\langle \mathrm{CC} \rangle_\lambda$ with respect to λ.

Based on this observation, we propose a simple scheme to infer the underlying interaction distance from ensembles of trajectories without referring to the ground truth: the interaction distance is defined as that where the local minimum of the derivatives exists along the cutoff distance λ:

$$\hat{R} \equiv \underset{\lambda}{\operatorname{argmin}} \frac{d\langle \mathrm{XX} \rangle_\lambda}{d\lambda}$$

where XX is either TE or CC and $\frac{d^2\langle \mathrm{XX} \rangle_\lambda}{d\lambda^2} = 0$.

Figure 6.6 shows the performance of TE and CC in identifying the interaction domain. Our scheme based on transfer entropy infers the underlying interaction distance satisfactorily at different noise levels ranging from 0.1π to 2π while as noise gets smaller, e.g., $\eta_0 = 0.1\pi$, cross correlation (CC) cannot estimate the underlying correct R at shorter R (≤ 2.0). At low noise, the $\langle \mathrm{CC} \rangle_\lambda$ vs λ curve has a local minimum at high λ where the $\frac{d\langle \mathrm{CC} \rangle_\lambda}{d\lambda}$ drops below the local minimum near $\lambda = R$.

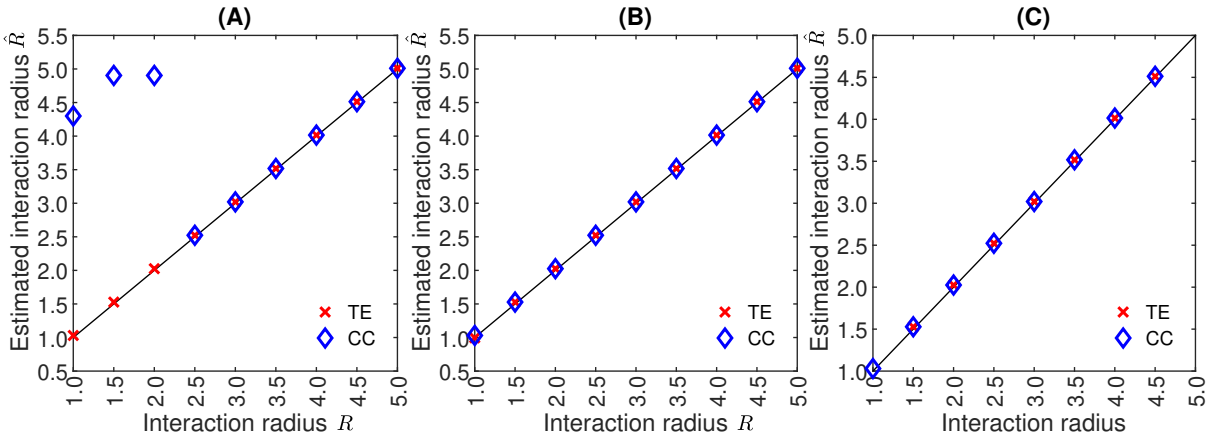

Figure 6.6: Estimation of interaction radius at (A) $\eta_0 = 0.1\pi$, (B) $\eta_0 = 1.2\pi$, and (C) $\eta_0 = 1.8\pi$ based on $\langle \text{TE} \rangle_\lambda$ and $\langle \text{CC} \rangle_\lambda$.

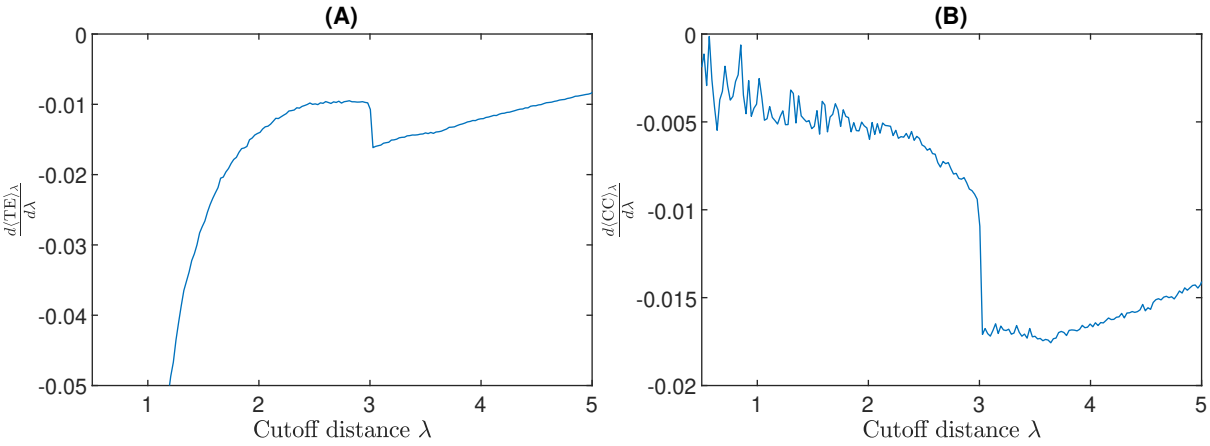

Figure 6.7: Estimation of interaction radius at $\eta_0 = 1.2\pi$ for $R = 3$ and $\tau_{obs} = 2$ using (A) TE, and (B) CC. It was found that the minimum of $\frac{d\langle \text{TE} \rangle_\lambda}{d\lambda}$ occurs at $\lambda = 3.02$, but the minimum of $\frac{d\langle \text{CC} \rangle_\lambda}{d\lambda}$ was found at $\lambda = 3.62$.

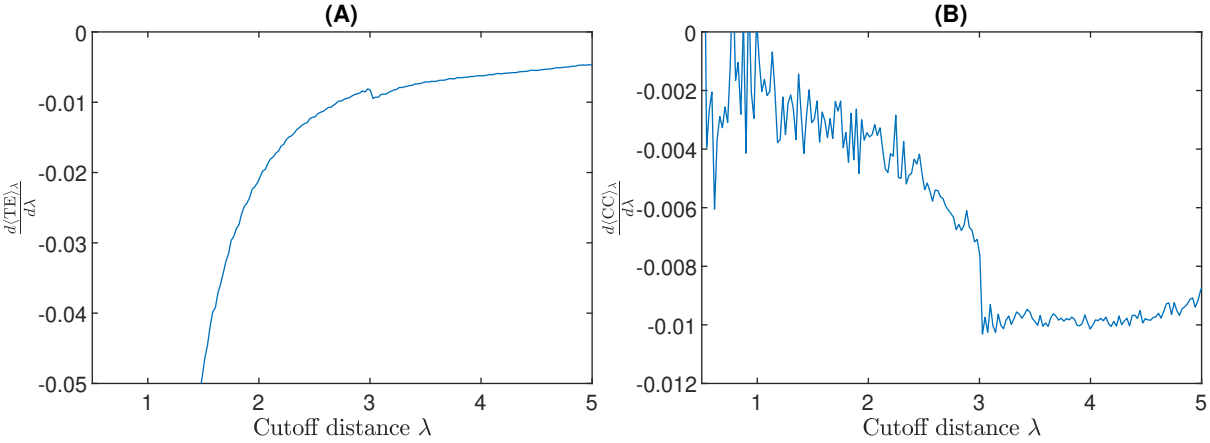

Figure 6.8: Estimation of interaction radius at $\eta_0 = 1.2\pi$ for $R = 3$ and $\tau_{obs} = 4$ using (A) TE, and (B) CC. It was found that the minimum of $\frac{d\langle \text{TE} \rangle_\lambda}{d\lambda}$ occurs at $\lambda = 3.02$, but the minimum of $\frac{d\langle \text{CC} \rangle_\lambda}{d\lambda}$ was found at $\lambda = 3.97$.

Even while the AUC scores are manifestly smaller in CC than those in transfer entropy (TE), CC can also estimate the interaction radius, as TE does, except for short R when the timescale of the system dynamics in Eq. 4.2 (i.e., 1 arb. unit), and the "observation" timescale, say, $\tau_{\rm obs}$, are the same. In practice, the timescale of measurement does not necessarily coincide with the timescale of the system. Consider the case where one can only sample the data once every $\tau_{\rm obs}$ time steps. In that case, the relationship between the system dynamics at the current time t and that at the next available time step $t + \tau_{\rm obs}$ is more nonlinear. For example, $\hat{R}$ estimated by TE and CC at $\eta_0 = 1.2\pi$ with the underlying actual interaction radius $R = 3.0$ for the observation timescale $\tau_{\rm obs} = 1 - 4$ with the total time length being 100 000 are as follows: $\hat{R}_{\rm TE} = 3.02$ and $\hat{R}_{\rm CC} = 3.02$ with $\tau_0 = 1$ as shown in Fig. 6.6, but $\hat{R}_{\rm TE} = 3.02$ and $\hat{R}_{\rm CC} = 3.62$ with $\tau_{\rm obs} = 2$ [Fig.6.7], and $\hat{R}_{\rm TE} = 3.02$ and $\hat{R}_{\rm CC} = 3.97$ with $\tau_{\rm obs} = 4$ [Fig. 6.8]. At the noise level gets much larger, e.g., 1.8π, it was found that both CC and TE could not assign the interaction radius due to high noise when $\tau_{\rm obs} > 1$. This implies that TE is superior to CC in predicting leader-follower relationship (Fig 6.3) and in inferring the underlying interaction domain and when the noise level gets much larger, both TE and CC fail to capture the underlying interaction distance for the case of the observation time $\tau_{\rm obs} > 1$.

6.6 Identification of interaction radius for different interaction time-scale and delay-time

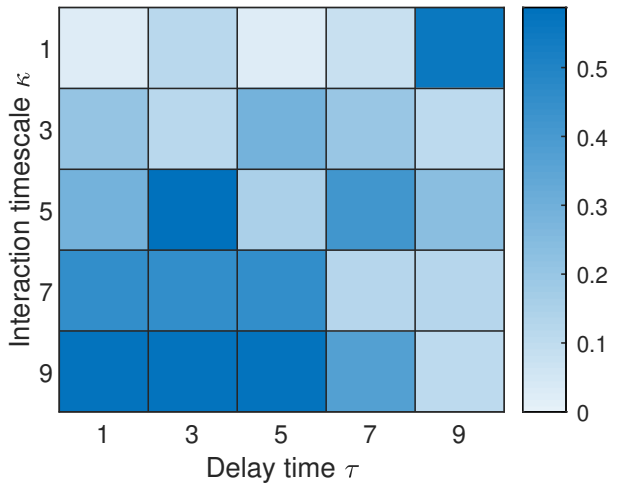

Figure 6.9: Relative error in identification of interaction radius for different κ and τ.

Figure 6.9 demonstrates the relative error (ΔR) landscape, in identifying interaction radius, as a function of interaction timescale κ and delay time τ using TE values at the noise level $\eta_0 = 1.2\pi$ and the interaction radius $R = 3$. Relative error (ΔR) is defined by $\Delta R = \frac{|R-\hat{R}|}{R}$. It is clear that the relative error is minimum when the interaction timescale κ and the delay time τ are the same (see the diagonal elements). ΔR tends to get larger while $\kappa > \tau$ more than $\kappa < \tau$ (c.f. off diagonal elements). It is natural because $\kappa > \tau$ implies that the observation τ is shorter than the underlying timescale κ intrinsic to the system at which particles interact with others.

6.7 Concluding Remarks

In this chapter we present a protocol to identify the interaction radius. For that purpose we used TE and CC, and found that TE outperforms CC. But at very high level of noise and shorter R, TE fails to identify the interaction but CC could estimate them very crudely. We found that our protocol is independent of leader and follower ratio and can be used even if no leader exists in the system (see Appendix B). Even for a large number of particles, our method is applicable (see Appendix C). We also found that by incorporating the identified actual interaction radius, the classification performance can be improved significantly.

Chapter 7

Effects of finite length trajectories and external noise on inferring interaction domain

7.1 Introduction

The scheme discussed in the previous chapter in inferring interaction domain is dependent on how the transfer entropy can be estimated so that it takes into account enough statistics of interacting particles, and positions and numbers of the minimum of the derivative of average transfer entropy along the cutoff distance λ may also be subject to the extent of external noise and time length of trajectories. How the prediction performance in capturing the underlying interaction domain depends on the size of noise and time length of the trajectory data have been discussed in this chapter. An alternative scheme has been proposed which is expected to be stable against noises and time length, that relies on the degree of convexity at coarse-grained scale in the derivative of average transfer entropy along with the cutoff distance, and time variance of underlying interaction radius of particles.

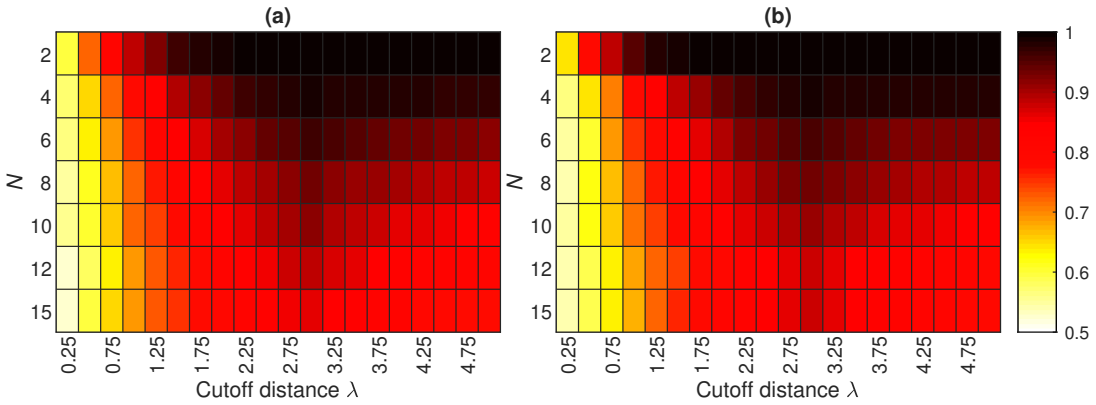

Figure 7.1: AUC landscape with respect to number of particles and cutoff distance. Actual interaction radius R used for the trajectory calculation is 3.0 and the noise is $\eta_0 = \frac{\pi}{2}$. (a) Fixed box size. The highest AUC scores and their locations are: 0.999 at $\lambda = 3.0$ arb. units ($N = 2$), 0.987 at $\lambda = 3.0$ ($N = 4$), 0.963 at $\lambda = 3.0$ ($N = 6$), 0.935 at $\lambda = 3.0$ ($N = 8$), 0.915 at $\lambda = 3.0$ ($N = 10$), 0.890 at $\lambda = 3.0$ ($N = 12$), and 0.867 at $\lambda = 3.0$ ($N = 15$). (b) Fixed density. The highest AUC scores and their locations are: 0.999 at $\lambda = 3.0$ ($N = 2$), 0.987 at $\lambda = 3.0$ ($N = 4$), 0.959 at $\lambda = 3.0$ ($N = 6$), 0.937 at $\lambda = 3.0$ ($N = 8$), 0.910 at $\lambda = 3.0$ ($N = 10$), 0.880 at $\lambda = 3.0$ ($N = 12$), and 0.877 at $\lambda = 3.0$ ($N = 15$).

7.2 Convexity score scheme in identifying interaction domain

Figure 7.1 represents the AUC landscape as a function of the number of particles N and cutoff distance λ. Figure 7.1(a) corresponds to a fixed box size (i.e. density is changing with N), whereas Fig. 7.1(b) represents the systems having same density (i.e. L is changing with N). The actual interaction radius used to simulate the trajectories of the particles was $R = 3$ and the noise was set to $\eta_0 = \frac{\pi}{2}$. Although at each N the maximum AUCs are located at the underlying interaction radius $R = 3.0$, and the maximum AUC score is higher than the conventional no-cutoff scheme corresponding to $R = 5\sqrt{2}$, the maximum AUC decreases as the number of particles increases due to the increase of indirect interactions between particles. Similar behaviors are observed at different interaction radius R and noise levels (not shown here).

The scheme proposed in Chapter 6 [[117]] to infer the underlying interaction distance from ensembles of trajectories, based on the existence of a significant decrease in averaged transfer entropy when cutoff distance λ exceeds the underlying interaction radius. The interaction distance is inferred as that where the minimum of the derivatives exists along

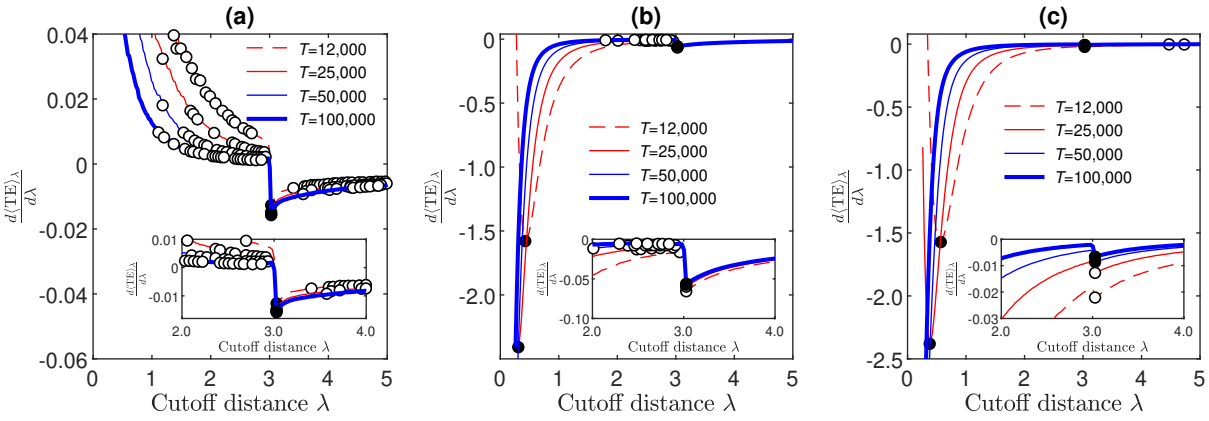

Figure 7.2: Derivative of average TE for different time length along with local minima and identified interaction radius based on global minimum scheme at (a) $\eta_0 = 0.2\pi$, (b) $\eta_0 = 1.2\pi$ and (c) $\eta_0 = 1.8\pi$. The actual interaction radius R used for ensemble of trajectories is 3. Circles represent all local minima and filled-circles represent the identified interaction radius at each time length T.

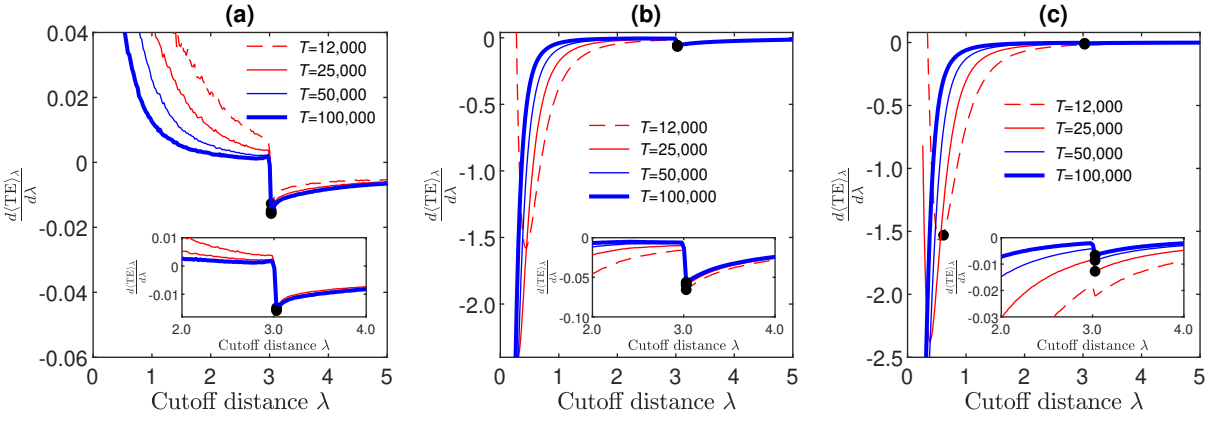

Figure 7.3: Derivative of average TE for different time lengths along with the identified interaction radius based on convexity score scheme at (a) $\eta_0 = 0.2\pi$, (b) $\eta_0 = 1.2\pi$ and (c) $\eta_0 = 1.8\pi$ for $\{M\} = \{M|2 \leq M \leq 30\}$ and $\delta = 1 \times 10^{-4}$.

the cutoff distance λ:

$$\hat{R} \equiv \text{argmin}_\lambda \frac{d\langle \text{TE} \rangle_\lambda}{d\lambda},$$

under the condition of $\frac{d^2\langle \text{TE} \rangle_\lambda}{d\lambda^2} = 0$. In practice, the length of trajectories may not be long enough and shorter length tends to result in a fluctuation in the course of TE along the cutoff distance λ, resulting in apparent minima.

Figure 7.2 shows the derivative of average TE as a function of cutoff distance λ, denoted by $\frac{d\langle TE \rangle_\lambda}{d\lambda}$ for interaction radius $R = 3$ at (a) $\eta_0 = 0.2\pi$, (b) $\eta_0 = 1.2\pi$, and (c) $\eta_0 = 1.8\pi$ for different T. Here circles represent all local minima and the filled-circles represent the global minimum identified as the interaction radius. In Figs. 7.2(b) and 7.2(c) for relatively short $T = 25,000$ or less, $\frac{d\langle TE \rangle_\lambda}{d\lambda}$ as a function of λ has the global minimum at low λ. The locations of these global minima for short time length change with T and they vanish when T is longer, and the longer $T = 50,000 - 100,000$ both result in a close value of the underlying interaction radius $R = 3$. This implies that to look for global minimum of derivative of transfer entropy may not necessarily result in an approximation of the true interaction radius especially for some short T at high noise levels.

In this chapter, a new scheme has been discussed which is expected to be robust against fluctuations of average transfer entropies along the cutoff distance by introducing a measure to quantify the degree of strong convexity at coarse-grained level, and time variance of underlying interaction radius of particles as follows.

Note in the insets of Fig. 7.2 that the shape of $\frac{d\langle \text{TE} \rangle_\lambda}{d\lambda}$ as a function of λ is (strongly) convex near the actual interaction radius irrespective of time length T. This indicates that the derivative of transfer entropy as a function of cutoff distance can shed light on the underlying spatial scale of interactions among particles. However, it is not trivial to devise a scheme to automatically infer the interaction radius. Since $\frac{d\langle \text{TE} \rangle_\lambda}{d\lambda}$ as a function of λ is convex near the interaction radius, a measure of convexity of $\frac{d\langle \text{TE} \rangle_\lambda}{d\lambda}$ is versatile in determining the interaction radius. In general, due to noise, $\frac{d\langle \text{TE} \rangle_\lambda}{d\lambda}$ can be fluctuated, producing apparent convex patterns locally. Thus in defining the convexity score, it is necessary to capture the non-local feature of $\frac{d\langle \text{TE} \rangle_\lambda}{d\lambda}$ rather than the local feature that may be subject to noise(s). We define the convexity score $\kappa(\lambda_i)$ of a function $f(\lambda)$ at a point λ_i as $\kappa(\lambda_i) = \frac{1}{M}\sum_{m=1}^{M} \sigma_i(m)$ where $\sigma_i(m) = +1$ if $f(\lambda_{i-m}) - f(\lambda_i) > \delta$ and $f(\lambda_{i+m}) - f(\lambda_i) > \delta$ and $\sigma_i(m) = -1$ if $f(\lambda_i) - f(\lambda_{i-m}) > \delta$ and $f(\lambda_i) - f(\lambda_{i+m}) > \delta$,

otherwise $\sigma_i(m) = 0$. Here δ is a non-negative small number and M is the number of neighboring points used to determine the convexity score, and $-1 \leq \kappa(\lambda_i) \leq 1$. Here the function $f(\lambda)$ represents the derivative of average TE, $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$.

Thus, around a point λ_i where $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$ is convex at some coarse-grained level $\kappa(i)$ is close to unity. Hence the point λ_i around which $\kappa(\lambda_i)$ is the maximum is considered to be an indicator of the interaction radius above which average transfer entropy between particles significantly decreases.

How to choose the optimal M and δ? We define the estimated interaction radius $\hat{R}$ as $\hat{R} \equiv \mathrm{argmax}_\lambda \kappa(M;T)$, where T represents the time length. Then the cost function is defined as $C(M) \equiv \sum_T \sum_{T'} |\hat{R}(M;T) - \hat{R}(M;T')|$ by assuming that the interaction radius is independent of time, i.e., time-invariant, and there exists (approximately) sufficient statistics for each time length in elucidating the interaction events. The parameter M is determined so that $M = \mathrm{argmin}_{M \in \{M\}} C(M)$. Here the sets of M to be searched for finding optimal M, $\{M\}$, are those users need to input *a priori*. Note that for some time length T, $\hat{R}(M;T)$ could not be chosen uniquely due to the degeneracy of $\kappa(M;T)$, and for such cases we set the cost $C(M)$ to infinity. In cases where the curves do not have any convex parts, we also cannot assign a unique $\hat{R}(M;T)$. We also varied $\delta = 0, 1 \times 10^{-8}, 1 \times 10^{-6}, 1 \times 10^{-4}$ (arb. units) and found that within this range of δ has no significant effect on the estimation of interaction radius.

Figures 7.4 and 7.5 show the relative errors (ΔR) in identifying the underlying interaction domain using the global minimum of the derivative of transfer entropy and convexity score scheme, respectively, as a function of time length and interaction radius at three different noise levels. Here relative error (ΔR) is defined as $\Delta R = \frac{|R - \hat{R}|}{R}$, where $\hat{R}$ is the identified interaction radius using either of the two schemes.

For the global minimum (of TE derivative) scheme [Fig. 7.4], there exists a clear trend such that the larger the noise level the larger the relative error, and as the time length T decreases, the relative errors are more pronounced for relatively large noise levels $\eta_0 = 1.2\pi - 1.8\pi$. Because of the appearance of a minimum at low cutoff distance for short T that ceases to exist for longer T in Figs. 7.2(b) and 7.2(c), the global minimum scheme apparently possesses higher relative error for short T [Figs. 7.4 (b) and 7.4 (c)].

Figure 7.5 shows the relative error in identifying the interaction radius at different noise levels using convexity score scheme. Although global minimum scheme possesses

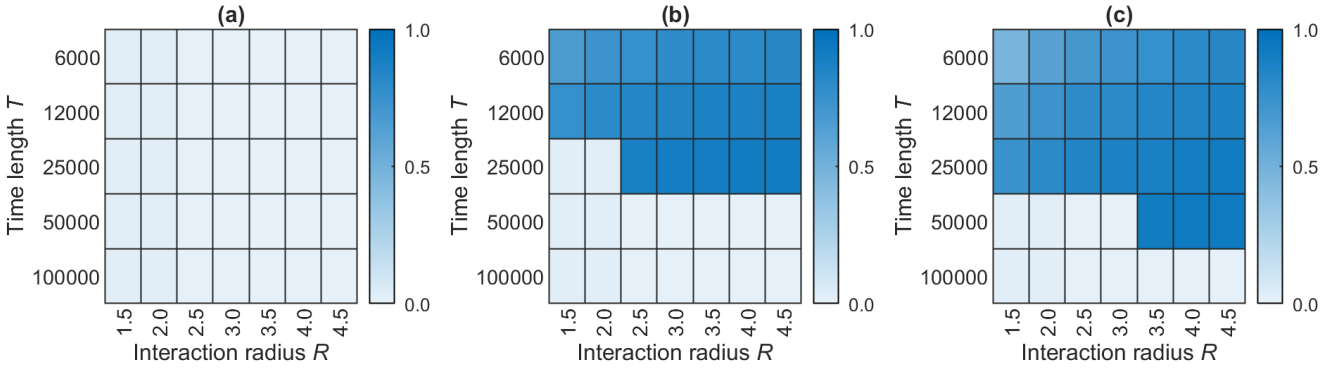

Figure 7.4: Relative error $\hat{R}$ in identifying underlying interaction domain using global minimum scheme at (a) $\eta_0 = 0.2\pi$, (b) $\eta_0 = 1.2\pi$, and (c) $\eta_0 = 1.8\pi$.

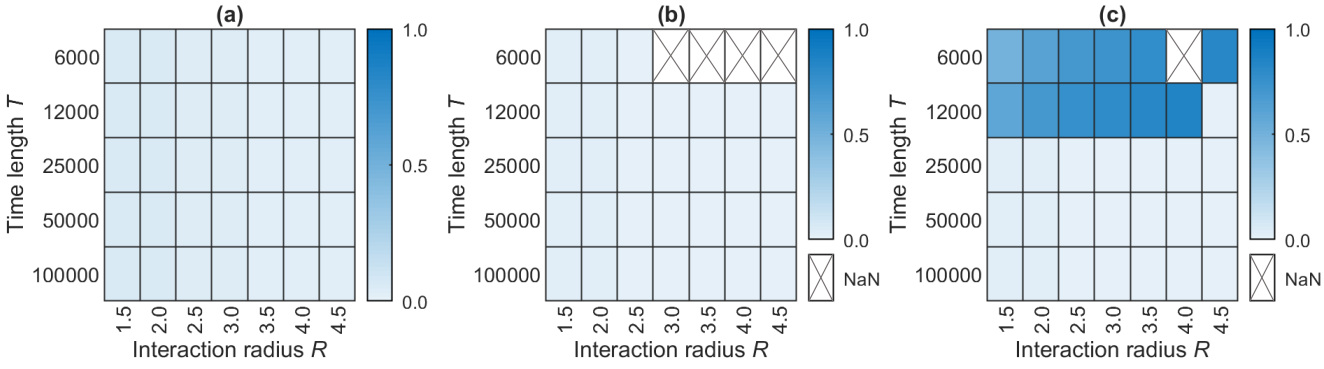

Figure 7.5: Relative error $\hat{R}$ in identifying underlying interaction domain using convexity score scheme with $\delta = 1 \times 10^{-4}$ and $\{M\} = \{M|2 \leq M \leq 30\}$ at (a) $\eta_0 = 0.2\pi$, (b) $\eta_0 = 1.2\pi$, and (c) $\eta_0 = 1.8\pi$. Cross-marked boxes 'NaN' mean that the no point was detected having non-zero convexity score κ.

high relative error at moderate noise when data length is short [Fig. 7.4(b)] but the convexity score scheme identifies the interaction radius satisfactorily [Fig. 7.5(b)] for $T \geq 12,000$. But for short T, e.g., $T \leq 6000$, the convexity score scheme fails to identify the interaction radius for $R \geq 3$. Like the global minimum scheme, the convexity score scheme possesses high relative error ΔR when noise level is very high [Fig. 7.5(c)] and T is short ($T \leq 12,000$). But when T is large ($T \geq$ 25,000), the convexity score scheme can identify the interaction radius competently [Fig. 7.5(c)] even at high noise levels.

7.3 Conclusions

In this chapter, we examined the performance of two heuristic schemes using the derivative of transfer entropy with respect to cutoff distance λ, $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$: global minimum and convexity score scheme for determining the interaction radius by using the modified VM. The striking feature -based on which we proposed the two schemes- is that $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$ exhibits a kink near the actual interaction radius. A method that is capable of determining the exact location of that kink can be used in inferring interaction radius.

For short time length (at the moderate and high level of noise) for the modified VM, the derivative of average TE exhibits a minimum at low cutoff distances that produce a relatively high error for the global minimum scheme. Moreover, in real experiments it is not necessarily possible to get sufficiently long trajectories with small noises, and the global minimum scheme may yield some non-negligible error especially for short trajectories with noise. Here an alternative scheme is presented, based on the property of convexity of $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$ at the coarse-grained level and the assumption of time-invariance of the underlying interaction domain. For the modified VM, the scheme could capture the underlying interaction radius at the low and moderate levels of noise especially for relatively short time length, ca. T=12,000-25,000 for which the global minimum scheme possesses high relative error (at moderate noise level). These two heuristic schemes are a compliment to each other and, as for appropriate usages, one should first visualize transfer entropy as a function of cutoff distance λ with its derivatives with respect to cutoff distance to confirm the existence of abrupt changes along with the cutoff distance. Significant changes in the derivative of TE with respect to cutoff distance were also observed for classical trajectories of particles interacting via Lennard-Jones potential

(not shown), and the existence of some significant change along cutoff distance around the typical length scale of interactions may be ubiquitous.

Chapter 8

Conclusions and Outlook

Our proposed scheme has demonstrated to predict which particles are more influential than the others, defined by weights w_{ij}, which corresponds to structural leadership [118]. In conceptual development on leader-follower relationship, there may exist the other types that are free from difference in the strength of fundamental interactions $\boldsymbol{w}$ [118]. In this study, interactions between particles are not necessarily pairwise via interaction circles of radius R, but they depend solely on the physical distance between them. Interactions among particles (such as cells) can be mediated via, e.g., chemical signal, and the size of interaction domain may not be necessarily invariant in time. For example, during the self- organization of an aggregation stream, Dictyostelium cells show highly heterogeneous communication signaling, i.e., some leader-like cells emit an earlier and stronger chemotactic signal to attract other follower-like cells. It is interesting to look into the problem in the aggregate of live cells on how they share the individual roles in terms of cell motility. These are some of the forthcoming subjects to be addressed along this study.

We examined the performance of two heuristic schemes using the derivative of transfer entropy with respect to cutoff distance λ, $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$: global minimum and convexity score scheme for determining the interaction radius by using the modified VM. The striking feature -based on which we proposed the two schemes- is that $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$ exhibits a kink near the actual interaction radius. A method that is capable of determining the exact location of that kink can be used in inferring interaction radius.

For short time length (at the moderate and high level of noise) for the modified VM, the derivative of average TE exhibits a minimum at low cutoff distances that produce

a relatively high error for the global minimum scheme. Moreover, in real experiments it is not necessarily possible to get sufficiently long trajectories with small noises, and the global minimum scheme may yield some non-negligible error especially for short trajectories with noise. An alternative scheme was also presented, based on the property of convexity of $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$ at the coarse-grained level and the assumption of time-invariance of the underlying interaction domain. For the modified VM, the scheme could capture the underlying interaction radius at the low and moderate levels of noise especially for relatively short time length, ca. T=12,000-25,000 for which global minimum scheme possesses high relative error (at moderate noise level). These two heuristic schemes are compliment to each other and, as for appropriate usages, one should first visualize transfer entropy as a function of cutoff distance λ with its derivatives with respect to cutoff distance to confirm the existence of abrupt changes along the cutoff distance. Significant changes in the derivative of TE with respect to cutoff distance were also observed for classical trajectories of particles interacting via Lennard-Jones potential (not shown), and the existence of some significant change along cutoff distance around the typical length scale of interactions may be ubiquitous.

In systems with many variables, identifying causal relationships is a daunting task. An important aspect of systems that should be exploited, however, is that particular variable may be only influencing another particular variable at certain time instances. We have shown that filtering out those time instances where influence does not occur greatly improves the identification of causal relationships. In the Vicsek model, for example, two agents may only interact when their distance is less than a certain threshold. To pose this as a question, we wonder at which value of interaction radius R does the motion of one agent influence the motion of another? More generally, we ponder the question: at which levels of variable X does variable Y influence variable Z? In future work, we will demonstrate the applicability of this method by applying it to data sets stemming from different fields.

Bibliography

1. P. Friedl and D. Gilmour, "Collective cell migration in morphogenesis, regeneration and cancer," *Nature reviews Molecular cell biology*, vol. 10, no. 7, p. 445, 2009.

2. A. Haeger, K. Wolf, M. M. Zegers, and P. Friedl, "Collective cell migration: guidance principles and hierarchies," *Trends in cell biology*, vol. 25, no. 9, pp. 556–566, 2015.

3. W. M. Lord, J. Sun, N. T. Ouellette, and E. M. Bollt, "Inference of causal information flow in collective animal behavior," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 1, pp. 107–116, 2016.

4. V. Mwaffo, S. Butail, and M. Porfiri, "Analysis of pairwise interactions in a maximum likelihood sense to identify leaders in a group," *Frontiers in Robotics and AI*, vol. 4, p. 35, 2017.

5. T. Omelchenko, J. Vasiliev, I. Gelfand, H. Feder, and E. Bonder, "Rho-dependent formation of epithelial "leader" cells during wound healing," *Proceedings of the National Academy of Sciences*, vol. 100, no. 19, pp. 10788–10793, 2003.

6. K. J. Cheung, E. Gabrielson, Z. Werb, and A. J. Ewald, "Collective invasion in breast cancer requires a conserved basal epithelial program," *Cell*, vol. 155, no. 7, pp. 1639–1651, 2013.

7. M. Reffay, L. Petitjean, S. Coscoy, E. Grasland-Mongrain, F. Amblard, A. Buguin, and P. Silberzan, "Orientation and polarity in collectively migrating cell structures: statics and dynamics," *Biophysical journal*, vol. 100, no. 11, pp. 2566–2575, 2011.

8. N. Yamaguchi, T. Mizutani, K. Kawabata, and H. Haga, "Leader cells regulate collective cell migration via rac activation in the downstream signaling of integrin $\beta 1$ and pi3k," *Scientific reports*, vol. 5, no. 1, pp. 1–8, 2015.

9. J. Krause, D. Hoare, S. Krause, C. Hemelrijk, and D. Rubenstein, "Leadership in fish shoals," *Fish and Fisheries*, vol. 1, no. 1, pp. 82–89, 2000.

10. C. Sueur, "Group decision-making in chacma baboons: leadership, order and communication during movement," *BMC ecology*, vol. 11, no. 1, p. 26, 2011.

11. T. D. Seeley, *The wisdom of the hive: the social physiology of honey bee colonies*. Harvard University Press, 2009.

12. M. Lindauer, "Communication in swarm-bees searching for a new home," *Nature*, vol. 179, no. 4550, pp. 63–66, 1957.

13. N. R. Franks and T. Richardson, "Teaching in tandem-running ants," *Nature*, vol. 439, no. 7073, pp. 153–153, 2006.

14. W. Jang and R. H. Gomer, "Combining experiments and modelling to understand size regulation in dictyostelium discoideum," *Journal of The Royal Society Interface*, vol. 5, no. suppl_1, pp. S49–S58, 2008.

15. Y. Ohta, T. Furuta, T. Nagai, and K. Horikawa, "Red fluorescent camp indicator with increased affinity and expanded dynamic range," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, 2018.

16. H. Beug, F. Katz, and G. Gerisch, "Dynamics of antigenic membrane sites relating to cell aggregation in dictyostelium discoideum," *The Journal of cell biology*, vol. 56, no. 3, pp. 647–658, 1973.

17. F. Alcantara and M. Monk, "Signal propagation during aggregation in the slime mould dictyostelium discoideum," *Microbiology*, vol. 85, no. 2, pp. 321–334, 1974.

18. N. Wiener, "The theory of prediction," *Modern mathematics for engineers*, 1956.

19. S. L. Bressler and A. K. Seth, "Wiener–granger causality: a well established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.

20. C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

21. D. M. Hausman, "The mathematical theory of causation," 1999.

22. F. A. Razak and H. J. Jensen, "Quantifying 'causality'in complex systems: understanding transfer entropy," *PLoS One*, vol. 9, no. 6, 2014.

23. S. Z. Chiou-Wei, C.-F. Chen, and Z. Zhu, "Economic growth and energy consumption revisited—evidence from linear and nonlinear granger causality," *Energy Economics*, vol. 30, no. 6, pp. 3063–3076, 2008.

24. R. Horváth, J. Seidler, and L. Weill, "Bank capital and liquidity creation: Granger-causality evidence," *Journal of Financial Services Research*, vol. 45, no. 3, pp. 341–361, 2014.

25. Y. Chen, G. Rangarajan, J. Feng, and M. Ding, "Analyzing multiple nonlinear time series with extended granger causality," *Physics letters A*, vol. 324, no. 1, pp. 26–35, 2004.

26. N. D. Mukhopadhyay and S. Chatterjee, "Causality and pathway search in microarray time series experiment," *Bioinformatics*, vol. 23, no. 4, pp. 442–449, 2007.

27. L. Keele, "Social capital and the dynamics of trust in government," *American Journal of Political Science*, vol. 51, no. 2, pp. 241–254, 2007.

28. S. Li, Y. Xiao, D. Zhou, and D. Cai, "Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information," *Physical Review E*, vol. 97, no. 5, p. 052216, 2018.

29. T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

30. B. Gourévitch and J. J. Eggermont, "Evaluating information transfer between auditory cortical neurons," *Journal of neurophysiology*, vol. 97, no. 3, pp. 2533–2543, 2007.

31. A. J. Majda and J. Harlim, "Information flow between subspaces of complex dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 104, no. 23, pp. 9558–9563, 2007.

32. X. San Liang and R. Kleeman, "Information transfer between dynamical system components," *Physical review letters*, vol. 95, no. 24, p. 244101, 2005.

33. B. L. Ruddell and P. Kumar, "Ecohydrologic process networks: 1. identification," *Water Resources Research*, vol. 45, no. 3, 2009.

34. S. Butail, V. Mwaffo, and M. Porfiri, "Model-free information-theoretic approach to infer leadership in pairs of zebrafish," *Physical Review E*, vol. 93, no. 4, p. 042411, 2016.

35. N. Orange and N. Abaid, "A transfer entropy analysis of leader-follower interactions in flying bats," *The European Physical Journal Special Topics*, vol. 224, no. 17-18, pp. 3279–3293, 2015.

36. K. Itoda, N. Watanabe, and Y. Takefuji, "Model-based behavioral causality analysis of handball with delayed transfer entropy," *Procedia Computer Science*, vol. 71, pp. 85–91, 2015.

37. C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 25–34, 1987.

38. H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Flocking in fixed and switching networks," *IEEE Transactions on Automatic control*, vol. 52, no. 5, pp. 863–868, 2007.

39. T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles," *Physical review letters*, vol. 75, no. 6, p. 1226, 1995.

40. I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, "Collective memory and spatial sorting in animal groups," *Journal of theoretical biology*, vol. 218, no. 1, pp. 1–12, 2002.

41. J. Buhl, D. J. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson, "From disorder to order in marching locusts," *Science*, vol. 312, no. 5778, pp. 1402–1406, 2006.

42. H. Levine, W.-J. Rappel, and I. Cohen, "Self-organization in systems of self-propelled particles," *Physical Review E*, vol. 63, no. 1, p. 017101, 2000.

43. R. A. Simha and S. Ramaswamy, "Hydrodynamic fluctuations and instabilities in ordered suspensions of self-propelled particles," *Physical review letters*, vol. 89, no. 5, p. 058101, 2002.

44. A. Czirók, H. E. Stanley, and T. Vicsek, "Spontaneously ordered motion of self-propelled particles," *Journal of Physics A: Mathematical and General*, vol. 30, no. 5, p. 1375, 1997.

45. W. Li, H.-T. Zhang, M. Z. Chen, and T. Zhou, "Singularities and symmetry breaking in swarms," *Physical Review E*, vol. 77, no. 2, p. 021920, 2008.

46. J. Agudo-Canalejo and R. Golestanian, "Active phase separation in mixtures of chemically interacting particles," *Phys. Rev. Lett.*, vol. 123, no. 1, p. 018101, 2019.

47. A. Kulkarni, S. P. Thampi, and M. V. Panchagnula, "Sparse game changers restore collective motion in panicked human crowds," *Phys. Rev. Lett.*, vol. 122, no. 4, p. 048002, 2019.

48. M. Durve and A. Sayeed, "First-order phase transition in a model of self-propelled particles with variable angular range of interaction," *Physical Review E*, vol. 93, no. 5, p. 052115, 2016.

49. W. Guo, J. Lü, S. Chen, and X. Yu, "Second-order tracking control for leader–follower multi-agent flocking in directed graphs with switching topology," *Syst. Control Lett.*, vol. 60, no. 12, p. 1051, 2011.

50. A. Murari, E. Peluso, M. Gelfusa, L. Garzotti, D. Frigione, M. Lungaroni, F. Pisano, P. Gaudio, and J. Contributors, "Application of transfer entropy to causality detection and synchronization experiments in tokamaks," *Nuclear Fusion*, vol. 56, no. 2, p. 026006, 2015.

51. K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.

52. T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

53. J. Sun and E. M. Bollt, "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings," *Physica D: Nonlinear Phenomena*, vol. 267, pp. 49–57, 2014.

54. I. Konvalinka, M. Bauer, C. Stahlhut, L. K. Hansen, A. Roepstorff, and C. D. Frith, "Frontal alpha oscillations distinguish leaders from followers: multivariate decoding of mutually interacting brains," *Neuroimage*, vol. 94, pp. 79–88, 2014.

55. S. H. Wang, X. E. Cheng, Z.-M. Qian, Y. Liu, and Y. Q. Chen, "Automated planar tracking the waving bodies of multiple zebrafish swimming in shallow water," *PloS one*, vol. 11, no. 4, 2016.

56. D. Hammer, A. Romashchenko, A. Shen, and N. Vereshchagin, "Inequalities for shannon entropy and kolmogorov complexity," *Journal of Computer and System Sciences*, vol. 60, no. 2, pp. 442–464, 2000.

57. J. Schug, W.-P. Schuller, C. Kappen, J. M. Salbaum, M. Bucan, and C. J. Stoeckert, "Promoter features related to tissue specificity as measured by shannon entropy," *Genome biology*, vol. 6, no. 4, p. R33, 2005.

58. J. Nuyts, "The use of mutual information and joint entropy for anatomical priors in emission tomography," in *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 6, pp. 4149–4154, IEEE, 2007.

59. C. Li, V. P. Singh, and A. K. Mishra, "Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy," *Water Resources Research*, vol. 48, no. 5, 2012.

60. M. J. Graham, A. J. Drake, S. Djorgovski, A. A. Mahabal, and C. Donalek, "Using conditional entropy to identify periodicity," *Monthly Notices of the Royal Astronomical Society*, vol. 434, no. 3, pp. 2629–2635, 2013.

61. A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical review E*, vol. 69, no. 6, p. 066138, 2004.

62. J. P. Pluim, J. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE transactions on medical imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

63. I. Kojadinovic, "On the use of mutual information in data analysis: an overview," in *Proc Int Symp Appl Stochastic Models Data Anal*, pp. 738–47, 2005.

64. T. Yamano, "A possible extension of shannon's information theory," *Entropy*, vol. 3, no. 4, pp. 280–292, 2001.

65. A. Kaiser and T. Schreiber, "Information transfer in continuous processes," *Physica D: Nonlinear Phenomena*, vol. 166, no. 1-2, pp. 43–62, 2002.

66. J. Jeong, J. C. Gore, and B. S. Peterson, "Mutual information analysis of the eeg in patients with alzheimer's disease," *Clinical neurophysiology*, vol. 112, no. 5, pp. 827–835, 2001.

67. S.-H. Jin, Y.-J. Kwon, J.-S. Jeong, S.-W. Kwon, and D.-H. Shin, "Differences in brain information transmission between gifted and normal children during scientific hypothesis generation," *Brain and cognition*, vol. 62, no. 3, pp. 191–197, 2006.

68. S.-H. Jin, P. Lin, and M. Hallett, "Linear and nonlinear information flow based on time-delayed mutual information method and its application to corticomuscular interaction," *Clinical Neurophysiology*, vol. 121, no. 3, pp. 392–401, 2010.

69. T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, *An introduction to transfer entropy: : Information Flow in Complex Systems*. Springer, 2016.

70. R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—a model-free measure of effective connectivity for the neurosciences," *Journal of computational neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.

71. M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy," *IEEE transactions on control systems technology*, vol. 15, no. 1, pp. 12–21, 2006.

72. M. Wibral, R. Vicente, and M. Lindner, "Transfer entropy in neuroscience," in *Directed information measures in neuroscience*, pp. 3–36, Springer, 2014.

73. R. G. James, N. Barnett, and J. P. Crutchfield, "Information flows? a critique of transfer entropies," *Physical review letters*, vol. 116, no. 23, p. 238701, 2016.

74. J. Runge, J. Heitzig, N. Marwan, and J. Kurths, "Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy," *Physical Review E*, vol. 86, no. 6, p. 061121, 2012.

75. F. Lozupone and S. Fais, "Cancer cell cannibalism: A primeval option to survive.," *Current molecular medicine*, vol. 15, no. 9, pp. 836–841, 2015.

76. C.-H. Choi, S.-J. Park, S.-Y. Jeong, H.-S. Yim, and S.-O. Kang, "Methylglyoxal accumulation by glutathione depletion leads to cell cycle arrest in dictyostelium," *Molecular microbiology*, vol. 70, no. 5, pp. 1293–1304, 2008.

77. S. D. Rosen, J. A. Kafka, D. L. Simpson, and S. H. Barondes, "Developmentally regulated, carbohydrate-binding protein in dictyostelium discoideum," *Proceedings of the National Academy of Sciences*, vol. 70, no. 9, pp. 2554–2557, 1973.

78. D. Watts and J. Ashworth, "Growth of myxamoebae of the cellular slime mould dictyostelium discoideum in axenic culture," *Biochemical Journal*, vol. 119, no. 2, p. 171, 1970.

79. Y. Maeda and J. Chida, "Control of cell differentiation by mitochondria, typically evidenced in dictyostelium development," *Biomolecules*, vol. 3, no. 4, pp. 943–966, 2013.

80. Y. Kubohara and H. Kikuchi, "Dictyostelium: An important source of structural and functional diversity in drug discovery," *Cells*, vol. 8, no. 1, p. 6, 2019.

81. A. Mesquita, E. Cardenal-Muñoz, E. Dominguez, S. Muñoz-Braceras, B. Nuñez-Corcuera, B. A. Phillips, L. C. Tábara, Q. Xiong, R. Coria, L. Eichinger, *et al.*, "Autophagy in dictyostelium: Mechanisms, regulation and disease in a simple biomedical model," *Autophagy*, vol. 13, no. 1, pp. 24–40, 2017.

82. K. Dannat, J. Tillner, T. Winckler, M. Weiss, K. Eger, and T. Dingermann, "Effects of medicinal compounds on the differentiation of the eukaryotic microorganism dictyostelium discoideum: Can this model be used as a screening test for reproductive

toxicity in humans?," *Die Pharmazie-An International Journal of Pharmaceutical Sciences*, vol. 58, no. 3, pp. 204–210, 2003.

83. F. Rivero and H. Xiong, "Rho signaling in dictyostelium discoideum," in *International review of cell and molecular biology*, vol. 322, pp. 61–181, Elsevier, 2016.

84. P. C. NEWELL, "Cellular communication during aggregation of dictyostelium: The second fleming lecture:(delivered at the general meeting of the society for general microbiology on 8 september 1977)," *Microbiology*, vol. 104, no. 1, pp. 1–13, 1978.

85. J. D. Dunn, C. Bosmani, C. Barisch, L. Raykov, L. H. Lefrançois, E. Cardenal-Muñoz, A. T. López-Jiménez, and T. Soldati, "Eat prey, live: Dictyostelium discoideum as a model for cell-autonomous defenses," *Frontiers in immunology*, vol. 8, p. 1906, 2018.

86. P. J. Van Haastert and P. N. Devreotes, "Chemotaxis: signalling the way forward," *Nature reviews Molecular cell biology*, vol. 5, no. 8, pp. 626–634, 2004.

87. V. Nanjundiah, "The evolution of communication and social behaviour indictyostelium discoideum," *Proceedings: Animal Sciences*, vol. 94, no. 6, pp. 639–653, 1985.

88. T. Gregor, K. Fujimoto, N. Masaki, and S. Sawai, "The onset of collective behavior in social amoebae," *Science*, vol. 328, no. 5981, pp. 1021–1025, 2010.

89. H. Odaka, S. Arai, T. Inoue, and T. Kitaguchi, "Genetically-encoded yellow fluorescent camp indicator with an expanded dynamic range for dual-color imaging," *PloS one*, vol. 9, no. 6, 2014.

90. C. J. Lloyd and Z. Yong, "Kernel estimators of the roc curve are better than empirical," *Statistics & Probability Letters*, vol. 44, no. 3, pp. 221–228, 1999.

91. J. V. Carter, J. Pan, S. N. Rai, and S. Galandiuk, "Roc-ing along: Evaluation and interpretation of receiver operating characteristic curves," *Surgery*, vol. 159, no. 6, pp. 1638–1645, 2016.

92. M. Leow and C. Mues, "Predicting loss given default (lgd) for residential mortgage loans: A two-stage model and empirical evidence for uk bank data," *International Journal of Forecasting*, vol. 28, no. 1, pp. 183–195, 2012.

93. S. Wang, D. Li, N. Petrick, B. Sahiner, M. G. Linguraru, and R. M. Summers, "Optimizing area under the roc curve using semi-supervised learning," *Pattern recognition*, vol. 48, no. 1, pp. 276–287, 2015.

94. J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve.," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

95. D. Faraggi and B. Reiser, "Estimation of the area under the roc curve," *Statistics in medicine*, vol. 21, no. 20, pp. 3093–3106, 2002.

96. K. Ataman, W. Streetr, and Y. Zhang, "Learning to rank by maximizing auc with linear programming," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 123–129, IEEE, 2006.

97. Z. Barutcuoglu and C. DeCoro, "Hierarchical shape classification using bayesian aggregation," in *IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)*, pp. 44–44, IEEE, 2006.

98. T. Vicsek and A. Zafeiris, "Collective motion," *Physics reports*, vol. 517, no. 3-4, pp. 71–140, 2012.

99. R. H. Turner, L. M. Killian, *et al.*, *Collective behavior*, vol. 3. Prentice-Hall Englewood Cliffs, NJ, 1957.

100. M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, *et al.*, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," *Proceedings of the national academy of sciences*, vol. 105, no. 4, pp. 1232–1237, 2008.

101. H. Delanoë-Ayari, J. Rieu, and M. Sano, "4d traction force microscopy reveals asymmetric cortical forces in migrating dictyostelium cells," *Physical Review Letters*, vol. 105, no. 24, p. 248103, 2010.

102. J. Toner and Y. Tu, "Long-range order in a two-dimensional dynamical xy model: how birds fly together," *Physical review letters*, vol. 75, no. 23, p. 4326, 1995.

103. K. Copenhagen, D. A. Quint, and A. Gopinathan, "Self-organized sorting limits behavioral variability in swarms," *Sci. Rep.*, vol. 6, p. 31808, 2016.

104. L. Barberis and F. Peruani, "Large-scale patterns in a minimal cognitive flocking model: incidental leaders, nematic patterns, and aggregates," *Physical review letters*, vol. 117, no. 24, p. 248001, 2016.

105. M. Aldana and C. Huepe, "Phase transitions in self-driven many-particle systems and related non-equilibrium models: a network approach," *Journal of Statistical Physics*, vol. 112, no. 1-2, pp. 135–153, 2003.

106. F. Ginelli, "The physics of the vicsek model," *The European Physical Journal Special Topics*, vol. 225, no. 11-12, pp. 2099–2117, 2016.

107. A. Cavagna, I. Giardina, and T. S. Grigera, "The physics of flocking: Correlation as a compass from experiments to theory," *Physics Reports*, vol. 728, pp. 1–62, 2018.

108. C. Wang and J. Han, "How does the interaction radius affect the performance of intervention on collective behavior?," *PloS one*, vol. 13, no. 2, 2018.

109. G. Chen, Z. Liu, and L. Guo, "The smallest possible interaction radius for flock synchronization," *SIAM J. Control Optim.*, vol. 50, no. 4, p. 1950, 2012.

110. D. J. Albers and G. Hripcsak, "Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 1, p. 013111, 2012.

111. S. H. Na, S.-H. Jin, S. Y. Kim, and B.-J. Ham, "Eeg in schizophrenic patients: mutual information analysis," *Clinical Neurophysiology*, vol. 113, no. 12, pp. 1954–1960, 2002.

112. P. Duan, F. Yang, T. Chen, and S. L. Shah, "Direct causality detection via the transfer entropy approach," *IEEE transactions on control systems technology*, vol. 21, no. 6, pp. 2052–2066, 2013.

113. M. Lindner, R. Vicente, V. Priesemann, and M. Wibral, "Trentool: A matlab open source toolbox to analyse information flow in time series data with transfer entropy," *BMC neuroscience*, vol. 12, no. 1, p. 119, 2011.

114. J. Sun, D. Taylor, and E. M. Bollt, "Causal network inference by optimal causation entropy," *SIAM Journal on Applied Dynamical Systems*, vol. 14, no. 1, pp. 73–106, 2015.

115. F. Hu, L.-J. Nie, and S.-J. Fu, "Information dynamics in the interaction between a prey and a predator fish," *Entropy*, vol. 17, no. 10, pp. 7230–7241, 2015.

116. U. S. Basak, S. Sattari, M. M. Hossain, K. Horikawa, and T. Komatsuzaki, "An information-theoretic approach to infer the underlying interaction domain among elements from finite length trajectories in a noisy environment," *The Journal of Chemical Physics*, vol. 154, no. 3, p. 034901, 2021.

117. U. S. Basak, S. Sattari, K. Horikawa, and T. Komatsuzaki, "Inferring domain of interactions among particles from ensemble of trajectories," *Phys. Rev. E*, vol. 102, p. 012404, Jul 2020.

118. J. Garland, A. M. Berdahl, J. Sun, and E. M. Bollt, "Anatomy of leadership in collective behaviour," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 7, p. 075308, 2018.

Appendix A

A.1 Equation

The velocity of particle i at time $t+1$ is determined by the weighted average velocity of its neighboring particles as follows:

$$\vec{v}_i^{t+1} = \frac{1}{\sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} w_{ij}} \sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} \left[w_{ii}\vec{v}_i^t + w_{ij}\vec{v}_j^{t-\kappa+1} \right]. \tag{A.1}$$

After splitting the velocity vector in (Eq. A.1) into its X and Y components we have:

$$\begin{aligned} \vec{v}_{xi}^{t+1} &= \frac{1}{\sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} w_{ij}} \sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} \left[w_{ii}v_{xi}^t + w_{ij}v_{xj}^{t-\kappa+1} \right], \\ &= \frac{1}{\sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} w_{ij}} \sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} \left[w_{ii}v_0 \cos\theta_i^t + w_{ij}v_0 \cos\theta_j^{t-\kappa+1} \right], \end{aligned} \tag{A.2}$$

and

$$\begin{aligned} \vec{v}_{yi}^{t+1} &= \frac{1}{\sum_{j:|\vec{r}_i^t - \vec{r}_j t-\kappa+1| \leq R} w_{ij}} \sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} \left[w_{ii}v_{yi}^t + w_{ij}v_{yj}^{t-\kappa+1} \right], \\ &= \frac{1}{\sum_{j:|\vec{r}_i^t - \vec{r}_j t-\kappa+1| \leq R} w_{ij}} \sum_{j:|\vec{r}_i^t - \vec{r}_j^{t-\kappa+1}| \leq R} \left[w_{ii}v_0 \sin\theta_i^t + w_{ij}v_0 \sin\theta_j^{t-\kappa+1} \right]. \end{aligned} \tag{A.3}$$

The ratio of the velocity components is then used to obtain the direction of motion:

$$\begin{aligned}\tan\theta_i^{t+1} &= \frac{v_{yi}^{t+1}}{v_{xi}^{t+1}},\\ &= \frac{\frac{1}{\sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R} w_{ij}} \sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R}\left[w_{ii}v_0\sin\theta_i^t + w_{ij}v_0\sin\theta_j^{t-\kappa+1}\right]}{\frac{1}{\sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R} w_{ij}} \sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R}\left[w_{ii}v_0\cos\theta_i^t + w_{ij}v_0\cos\theta_j^{t-\kappa+1}\right]},\\ &= \frac{\sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R}\left[w_{ii}\sin\theta_i^t + w_{ij}\sin\theta_j^{t-\kappa+1}\right]}{\sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R}\left[w_{ii}\cos\theta_i^t + w_{ij}\cos\theta_j^{t-\kappa+1}\right]}. \end{aligned} \tag{A.4}$$

Thus,

$$\begin{aligned}\theta_i^{t+1} &= \tan^{-1}\left[\frac{\sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R}\left[w_{ii}\sin\theta_i^t + w_{ij}\sin\theta_j^{t-\kappa+1}\right]}{\sum_{j:|\vec{r}_i^t-\vec{r}_j^{t-\kappa+1}|\le R}\left[w_{ii}\cos\theta_i^t + w_{ij}\cos\theta_j^{t-\kappa+1}\right]}\right],\\ &\equiv \langle\boldsymbol{\theta}(t-\kappa+1)\rangle_{R,\boldsymbol{w},\vec{r}_i^t}.\end{aligned}$$

A.2 Application to 50 particles system

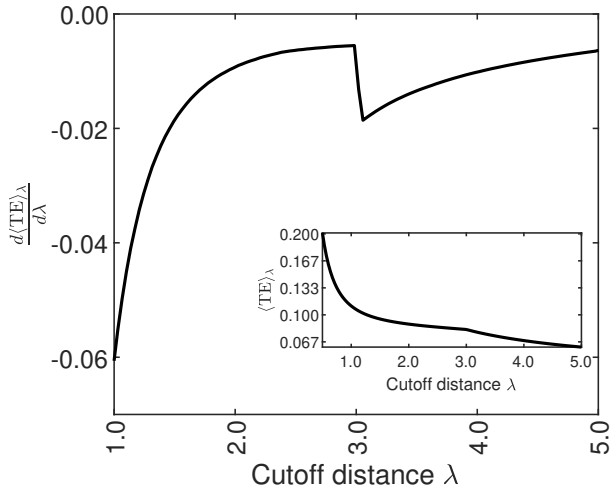

Figure A.1: The derivative of $\langle \mathrm{TE}\rangle_\lambda$ with respect to λ. Inset: $\langle \mathrm{TE}\rangle_\lambda$. In this figure the number of particles is $N = 50$, interaction radius $R = 3$ and noise $\eta_0 = 1.2\pi$. It indicates that minimum of the derivative approach can be used to identify interaction radius for larger number of particles.

A.3 Leader-follower ratio

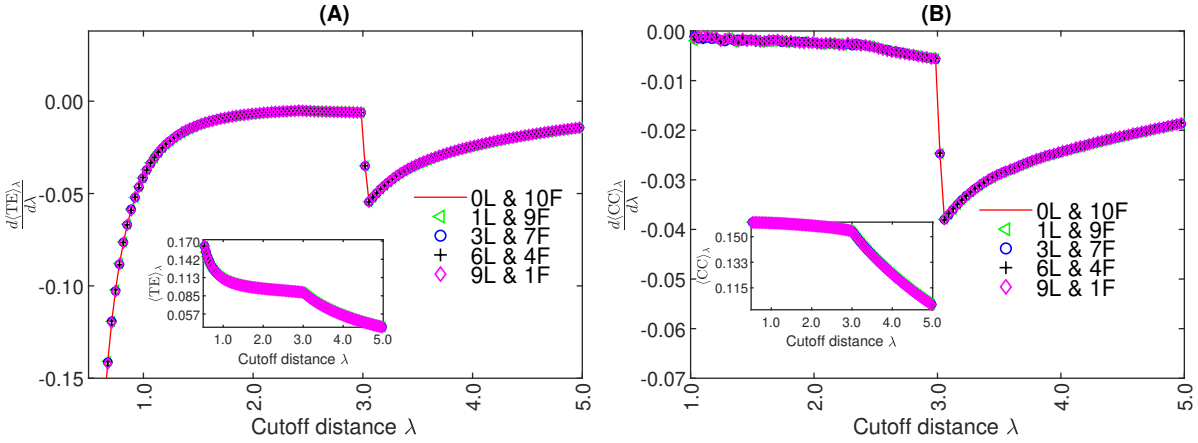

Figure A.2: (A) The derivative of $\langle \mathrm{TE} \rangle_\lambda$ with respect to λ for different leader-follower ratio. Inset: $\langle \mathrm{TE} \rangle_\lambda$. (B) The derivative of $\langle \mathrm{CC} \rangle_\lambda$ with respect to λ for different leader-follower ratio. Inset: $\langle \mathrm{CC} \rangle_\lambda$. The actual interaction radius used for these figures is 3 and noise $\eta_0 = 1.2\pi$. It can be seen that leader-follower ratio does not affect the average information flow, hence the estimation of the interaction radius.

A.4 Estimated interaction radius at $R = 2.0$

Figure (A.3) shows the derivative of average TE for different time lengths along with the identified interaction radius based on convexity score scheme at different noise levels. It has been found that at low and moderate noise levels the convexity score scheme identifies the interaction radius satisfactorily [Figs. A.3(a) and A.3(b)]. But at very high level of noise ($\eta_0 \approx 1.8\pi$) this scheme performs satisfactorily for longer T but possesses higher relative error for shorter T [Fig. A.3(c)].

A.5 Convexity score

According to the definition, convexity score κ is maximum at the point for which the $\frac{d\langle \mathrm{TE} \rangle_\lambda}{d\lambda}$ is strongly convex, and κ is minimum where the curve is concave. Figure (A.4) shows the convexity score for moderate noise level ($\eta_0 = 1.2\pi$) at different interaction radii R. For $R = 2.0$ [Fig. A.4(a)] unique maximum convexity scores have been identified for each T ($T = 6,000 - 100,000$). Hence the convexity score scheme identifies the interaction radius correctly. For $R = 3.0$ and $R = 4.0$, no unique λ was identified for

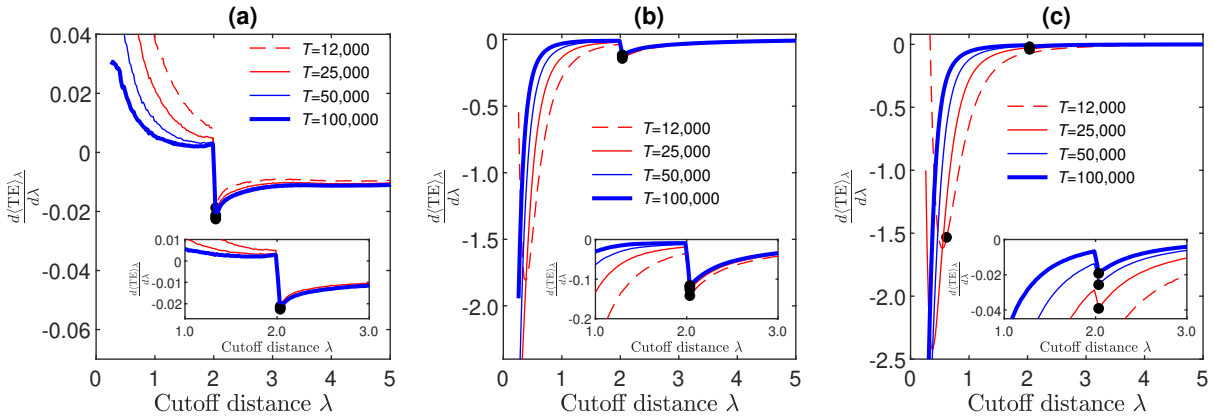

Figure A.3: Derivative of average TE for different time lengths along with the identified interaction radius based on convexity score scheme at (a) $\eta_0 = 0.2\pi$, (b) $\eta_0 = 1.2, \pi$ and (c) $\eta_0 = 1.8\pi$ for $R = 2.0$, $\delta = 1 \times 10^{-4}$ and $\{M\} = \{M|2 \leq M \leq 30\}$.

$T = 6,000$ [Figs. A.4(b) and A.4 (c)]. However, the convexity score scheme can identify the interaction radius perfectly for longer T ($T = 12,000 - 100,000$).

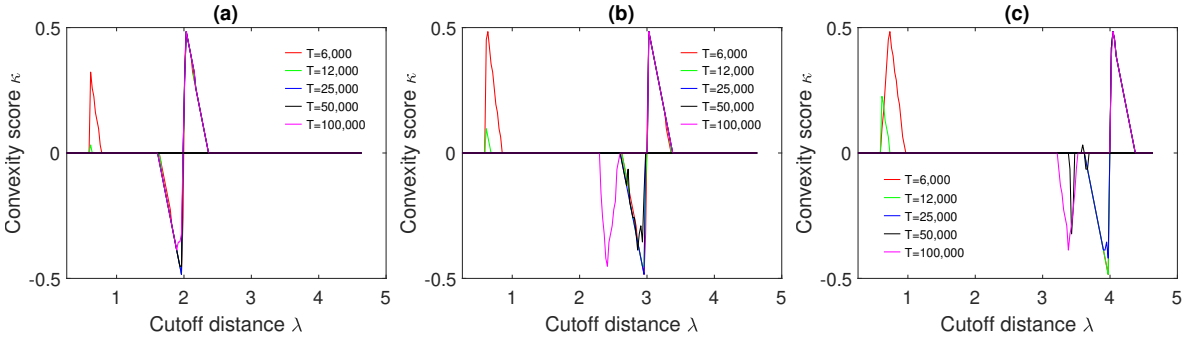

Figure A.4: Convexity score κ at (a) $R = 2.0$, (b) $R = 3.0$, and (c) $R = 4.0$ at $\eta_0 = 1.2\pi$.