Revealing Causal Graph Structure in Pigeon Flocks Using Attentions in Temporal ConvNet

Jimson Paulo Layacan, Aengus Martin Donaire, Harvey Felipe, Clark Kendrick Go
Collaborative Analytics Group, Department of Mathematics, Ateneo de Manila University
Quezon City, Philippines

2023 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-8867-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IJCNN54540.2023.10191157

***Abstract*—The study of collective animal behavior has long been of interest to researchers across various disciplines. In this study, a modified version of the Attention-based Dilated Depthwise Separable Temporal Convolutional Networks (AD-DSTCN) architecture was used to analyze high-resolution spatiotemporal data of moving pigeons with the goal of producing causal graph structures in pigeon flocks. The modification allowed for causality inference on data where individuals have multidimensional feature representations. The study revealed several important findings. Firstly, the results showed that, although weak external influences may still be present, each pigeon primarily relies on its own knowledge and decision-making. Secondly, the results also indicated that edges in the graph correspond to pairs of pigeons that are close to each other, suggesting that pigeons that are physically near each other tend to have stronger social connections. Thirdly, the study suggested that there is no single persistent pigeon that holds the most influence in the flock. Instead, the most influential node appears to shift at various time intervals, a concept known as intermittent switching. These findings offer new insights into the complexity of animal behavior and its underlying mechanisms, as well as a deeper understanding of how pigeons make decisions and interact within a flock. Ultimately, the use of the modified AD-DSTCN architecture offers a promising approach for understanding collective animal behavior.**

***Index Terms*—AD-DSTCN, attention mechanism, causal inference, collective behavior**

I. Introduction

Collective behavior is a well-studied phenomenon in various biological systems, such as migrating geese [1], bacterial colonies [2], locust swarms [3], fish schools [4], and bird flocks [5]. It originates from local actions taken by individuals that can influence the overall movement of a socially interacting group [6]. Numerous biological processes rely on such social behavior including resource acquisition, reproduction, and predator avoidance [7]–[10].

Beyond these biological functions, the study of collective behavior has also led to many practical applications such as in robotics [11] and autonomous systems [12]. For instance, swarm intelligence researchers have built robots that autonomously operate without centralized control by mimicking swarming behaviors [13]. This approach has proven effective in solving complex problems, including sensing in micromachinery and the human body as well as search and rescue operations [14].

On the other hand, pigeons make great subjects for studying collective behavior due to their complex social movements [15] and ability to thrive in diverse environments [16], [17]. They are also highly social animals that form strong social bonds and communicate with each other through a variety of vocalizations and body language [18].

Previous studies have explored the factors that may influence decision-making in pigeon flocks. Nagy et al. [19] proposed that hierarchy drives the flocking behavior of pigeons. However, other studies have suggested that a balance between compromise and leadership may also play a role in pigeons' movements. For instance, Xu et al. [20] found that both reciprocal relationships and stratified hierarchical leadership contribute to their motion. Similarly, Zhang et al. [21] examined how pigeon flocks obey hierarchical and egalitarian interaction patterns and proposed that pigeons follow the average of their neighbors on smooth trajectories but switch to following leaders during sudden turns.

Chen et al. [22] took their analysis a step further and discovered a switching hierarchical mechanism where a long-term leader guides the flock's movement, but a temporary leader takes over during turns. They also found that the causal relationships of pigeons follow a local interaction mode, with pigeons closer to the center of mass being more influential than others [23].

Apparently, there has not been a clear consensus on the causal relationships in the collective behavior of pigeons. This can be attributed to the fact that different studies have presented varying methodologies for understanding the dynamics of pigeon flocks. Factors such as group size, hierarchy, and individual personalities have been proposed as explanations. Traditionally, researchers have analyzed spatial positions and identified initiators of group movements to understand leader-follower relationships in animal groups [24].

Meanwhile, statistical methods like cross-correlation [25] and Granger Causality [26] (model-based) and event synchronization [27] and transfer entropy [28] (model-free) have also been used to study collective behavior. However, model-based methods often rely on assumptions about the system that can limit their accuracy [29], while model-free methods do not require a model but may be less interpretable and require more data for comparable performance [30].

To address the limitations of such methods, deep learning techniques have emerged as a powerful tool for causal inference. Deep learning does not rely on assumptions about the underlying system, making it well-suited for complex, nonlinear systems. Furthermore, recent advances in interpretation tools for deep learning have also made it possible to identify causal links with greater accuracy and precision.

One such example is the causal discovery approach proposed by Nauta et al. [31]. This method employs Convolutional Neural Networks (ConvNets) with attention layers to extract valuable insights into causal associations in complex decision-making systems. Despite its success in identifying causal linkages on benchmark datasets in finance and neuroscience, it may require further adjustments and modifications to accurately model and interpret collective animal behavior.

In this paper, the researchers explored causal relationships in pigeon flocks through deep interpretable learning. Specifically, they adapted Nauta's Temporal Causal Discovery Framework (TCDF) for use with pigeons. The modification involved adapting the predictive architecture, Attention-based Dilated Depthwise Separable Temporal Convolutional Network (AD-DSTCN), to enable multidimensional feature representations. The attention scores obtained from the architecture were analyzed and used to construct causal graph structures, which were then examined for any possible relation to pigeons' pairwise distances and intermittent leadership switching.

II. Data Description

The researchers used pre-processed high-resolution GPS datasets obtained from a study by Nagy et al. [32], which consisted of time series data on homing pigeons' position, velocity, and acceleration in the x, y, and z dimensions during several flights. The datasets involved three pigeon flocks, A, B, and C, with ten pigeons in each group, and five free flights for each group. To ensure accurate comparison of the flight patterns, the data was carefully prepared by removing time steps containing null values and matching the time steps to have the same number of pigeons in each flight interval.

III. Mathematical Formulation

Using multivariate spatiotemporal data, the primary aim of the study was to identify causal associations between pigeons based on their velocity in the x, y, and z dimensions, and construct causal graphs based on these relationships.

Consider the set of n observational multivariate time series represented as

$$\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\} \in \mathbb{R}^{n\times m\times t} \tag{1}$$

where each $\mathbf{X}_i$ consists of m features of length t, i.e.,

$$\mathbf{X}_i = (\mathbf{x}_i^1, \mathbf{x}_i^2, \ldots, \mathbf{x}_i^m)^T \in \mathbb{R}^{m\times t}, \tag{2}$$

and each feature $\mathbf{x}_i^j$ contains numerical values across t time steps, i.e.,

$$\mathbf{x}_i^j = (x_{i,1}^j, x_{i,2}^j, \ldots, x_{i,t}^j)^T \in \mathbb{R}^t. \tag{3}$$

The inferred causal relationships are represented by a directed causal graph

$$\mathbf{G} = (\mathbf{X}, \mathbf{E}), \tag{4}$$

where each directed $e_{i,k} \in \mathbf{E}$ represents a causal relationship between the cause $\mathbf{X}_i$ and effect $\mathbf{X}_k$. Given the notation above,

$$\text{Flight} := \mathbf{X}, \quad \text{Pigeon}_i := \mathbf{X}_i,$$
$$\text{Feature}_j := \mathbf{x}_i^j, \quad \text{Time step}_t := x_{i,t}^j.$$

IV. Methodology

This study employed a three-stage approach: model training, causal inference, and analyses of the causal graph structures.

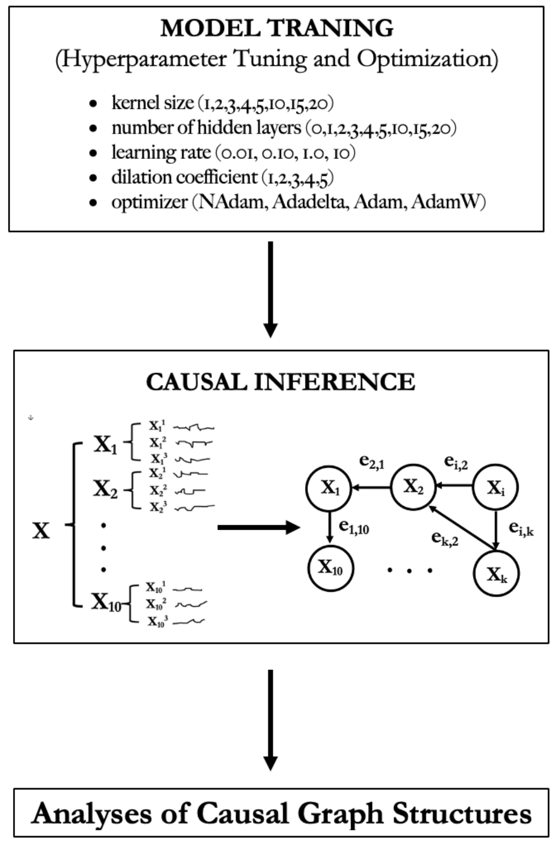

Fig. 1. Methodology Flowchart

First, the researchers trained a predictive model to forecast the velocities of pigeons. Being a spatiotemporal function, the velocity was used as it provides more accurate information than steady global positional data [19]. After training the model, the attention scores were used in conjunction with Nauta's (TCDF) [31] to build causal graph structures. This framework allowed for the identification of causal relationships between the pigeons and an understanding of their social dynamics. Finally, the researchers analyzed the connection between the graphs produced and the pairwise distances of each pigeon in order to gain a deeper understanding of the factors influencing their movements. Moreover, they also investigated the so-called intermittent leadership switching in pigeons.

A. Model Training

The AD-DSTCN is a deep learning architecture for discovering causal relationships in observational time series data. This architecture uses multiple Convolutional Neural Networks (CNNs) that receive all observed time series as input. One of its key features is the attention mechanism, which allows the model to assign different weights to different time series based on their relative importance for the prediction. The attention scores are calculated through an attention layer at the top of the architecture, as described by Nauta et al. [31].

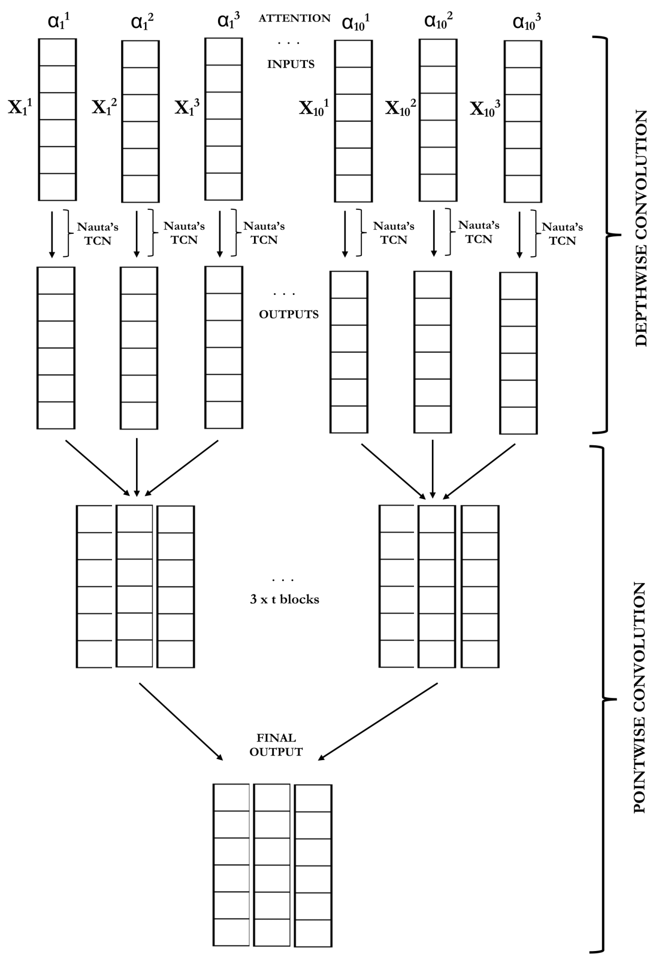

Fig. 2. Modified Attention-based Dilated Depthwise Separable Temporal Convolutional Networks (AD-DSTCN) $(n = 10, m = 3)$

The core of the AD-DSTCN architecture is composed of depthwise separable convolutions [33]. These convolutions are a fusion of depthwise and pointwise convolutions, where the depthwise convolutions use a kernel to process each input channel individually, and the pointwise convolution merges the individual results using a 1×1 convolution.

In this case, the depthwise convolutions used are called Temporal Convolutional Networks (TCNs) [34]. They are a type of deep neural network that are specifically designed for time series or sequence modeling tasks. TCNs have the advantage of being causal, meaning that information from the future does not leak into the past. Furthermore, TCNs can handle sequences of any length and produce an output sequence of the same length, thus allowing for the prediction of a pigeon's velocity throughout a flight, taking into account not only that pigeon's velocity, but also the velocities of other pigeons within the flock.

To handle multi-dimensional feature representations such as velocities, the original AD-DSTCN architecture was modified. The modification involved stacking the outputs of the depthwise convolution into groups of three, followed by applying a 1×1 pointwise convolution to the multiple stacked time series to obtain the final prediction. This modification, illustrated in Figure 2, enables the handling of multi-dimensional feature representations, which is crucial for predicting the velocities of pigeons in three-dimensional space.

Since there are n pigeons in a flight, n independent modified AD-DSTCNs were trained, having a common structure but different target time series. The purpose of each network $\mathcal{N}_j$ from $j = 1, 2, \ldots, n$ was to predict its target time series $\mathbf{x}_j^1$, $\mathbf{x}_j^2$, and $\mathbf{x}_j^3$ $(x, y, z$ dimensions). The input to the jth network comprised a dataset $\mathbf{X}$ of $3n \times t$, consisting of $3n$ time series of equal size and length t.

To calibrate the modified architecture, an extensive grid search was performed to find the optimal hyperparameters. Various sets of hyperparameters were tested on three groups of pigeons (A, B, and C) using each of their first free flight data. Multiple predictions for each pigeon were made for each flight, and the mean squared error (MSE) was computed for each prediction. To find the average MSE for a flight, the average of the MSEs for all pigeons in that flight, which is referred to as MSE_{ave}, was then calculated. Finally, the average of MSE_{ave}'s across all first free flights for each group of pigeons was obtained to compare the different sets of hyperparameters and determine the best model.

TABLE I
RANGE OF VALUES OF HYPERPARAMETERS

Hyperparameter	Values
Kernel Size	1, 2, 3, 4, 5, 10, 15, 20
Number of Hidden Layers	0, 1, 2, 3, 4, 5, 10, 15, 20
Learning Rate	0.01, 0.10, 1.0, 10
Dilation coefficient	1, 2, 3, 4, 5
Loss Function Optimizer	NAdam, Adadelta, Adam, AdamW

In particular, the test involved defining a range of values for each hyperparameter (Table I) (including kernel size, number of hidden layers, learning rate, optimizer, and dilation coefficient) and testing all combinations of values to find the one that resulted in the best performance.

B. Causal Inference

To model the attention mechanism, a framework where each network $\mathcal{N}_j$ has its own attention vector was implemented where $j = 1, 2, \ldots, n$. In notation,

$$\mathbf{a}_j = [a_{1,j}^1, a_{1,j}^2, a_{1,j}^3, ..., a_{i,j}^1, a_{i,j}^2, a_{i,j}^3, ..., a_{n,j}^1, a_{n,j}^2, a_{n,j}^3]$$

TABLE II
TOP COMBINATIONS OF HYPERPARAMETERS

Kernel Size	# Hidden Layers	Learning Rate	Dilation Coefficient	Loss Function Optimizer	Averaged MSE$_{ave}$
2	0	0.01	4	Adam	0.00825
2	0	0.01	3	Adam	0.00847
2	0	0.01	1	Adam	0.00893
2	0	0.10	4	AdamW	0.00932
2	0	0.10	5	AdamW	0.00932

are attention scores $a^1_{i,j}, a^2_{i,j}, a^3_{i,j} \in a_j$ which were multiplied with input time series $\mathbf{x}^1_i, \mathbf{x}^2_i$, and $\mathbf{x}^3_i$, respectively, in network $\mathcal{N}_j$.

Consequently, the attention scores $a^1_{i,j}, a^2_{i,j}$, and $a^3_{i,j}$ indicate the degree to which $\mathcal{N}_j$ focuses on input time series $\mathbf{x}^1_i, \mathbf{x}^2_i$, and $\mathbf{x}^3_i$, (corresponding to the ith pigeon) when predicting targets $\mathbf{x}^1_j, \mathbf{x}^2_j$, and $\mathbf{x}^3_j$, (corresponding to the jth pigeon). Higher values for $a^1_{i,j}, a^2_{i,j}$, and $a^3_{i,j}$ suggest that $\mathbf{X}_i$ (ith pigeon) may influence $\mathbf{X}_j$ (jth pigeon).

Due to the grouping of attention scores, a direct interpretation of attention is not possible. To obtain a single score for each pigeon, a weighted average of their attention scores was calculated. This method incorporates the influence of each velocity dimension, resulting in a representative score for each pigeon. For example, when predicting velocities in network $\mathcal{N}_j$ (jth pigeon), the weighted average of attention scores $a^1_{i,j}, a^2_{i,j}$, and $a^3_{i,j}$ is obtained to determine a single representative score for feature i (ith pigeon).

During the network training, all attention scores are initialized as 1, $\mathbf{a}_j$ = [1, 1, ..., 1]. As the network predicts its target time series using backpropagation, the attention scores are also updated in every training epoch by increasing or decreasing each score.

Nauta proposed a simple algorithm [31] for identifying potential causes, which determines the threshold by finding the largest gap τ_j between the weighted attention scores. The algorithm sorts the scores in descending order and looks for the largest gap g between two adjacent scores. The threshold τ_j is then set to the attention score on the left side of the gap.

Originally, the algorithm requires that $\tau_j \geq 1$ since all scores were initialized at 1, and a score would only increase through backpropagation if the network paid attention to that time series. However, a reduced threshold of 0.5 was needed to consider a larger number of potential causes, including weaker influences. To ensure that the algorithm does not include low attention scores in the selection, the gap selected for τ_j must be in the first half of the sorted list of gaps. This restriction limits the number of potential causes of the target to at most 50% of the input time series. It is also required that the gap for τ_j cannot be in the first position, between the highest and second-highest attention scores.

Moreover, this study used removal interventions for the causal validation step, as opposed to the permutation importance described in Nauta's causal framework [31]. This approach involves removing each potential cause one by one and re-running the prediction with the modified input data. In this manner, the validation provides a more direct measure of feature importance by directly observing the effect of removing each feature on the model's performance. Specifically, the MSE$_{ave}$ of the velocity predictions with interventions was compared to the original. If the new MSE$_{ave}$ was higher, the potential cause was considered validated, indicating that the cause (pigeon) is deemed computationally important for predicting a specific pigeon.

C. Analyses of Causal Graph Structures

After generating causal graph structures for multiple flights, the subsequent step was to analyze these graphs. One of the objectives was to investigate the relationship between the graphs and the physical proximity of pigeons, and so the researchers calculated the pairwise distances between each pair of pigeons at every time step. The distances were then averaged over the entire flight, and the resulting average distances were used to determine whether two pigeons were near to each other. In particular, the average of all pairwise distances of all pigeon pairs was used as the threshold across all time steps. Pairs with a distance less than this threshold were considered to be near. This represents a proximity range within which the birds can influence each other.

In the analysis, this study not only determined the overall structure of the network but also identified the most influential node within the causal graph structures. To achieve this, the metric that uses the number of descendants was employed. The number of descendants represents the quantity of nodes that can be reached through the edges originating from a particular node. This metric provides an indication of the node's level of influence within the network, as a node with a larger number of descendants holds a stronger position of influence.

Lastly, the researchers also studied the switching of leaders. To do this, they analyzed multiple determined causal graphs at various time windows, each of which was one minute long. For each time window, a causal graph structure was created and the number of descendants was used to determine the most influential pigeon in the causal graph. This process was repeated over multiple time windows to track changes in each pigeon's level of influence. The researchers evaluated the consistency of leadership roles throughout the flight by observing if the same pigeon consistently emerged as the leader across different time windows.

V. RESULTS AND DISCUSSION

The results of the experiment, summarized in Table II, showed that the combination of a kernel size of 5, no hidden layers, a learning rate of 0.01, the Adam optimizer, and a dilation coefficient of 4 produced the lowest average of MSE_{ave}'s. These hyperparameters were integrated into the predictive model's architecture to forecast the velocity of moving pigeons. The researchers then analyzed the model's internal parameters, particularly the weighted attention scores.

The heat map presented in Figure 3 shows the attention scores assigned by the model for a particular flight and group of pigeons. However, it can be noted that similar patterns are observed across all flights and groups, where the diagonal of the heat map is consistently almost blue, indicating a high attention score assigned to each pigeon for itself. This suggests that in general, pigeons tend to rely primarily on their own information when making decisions.

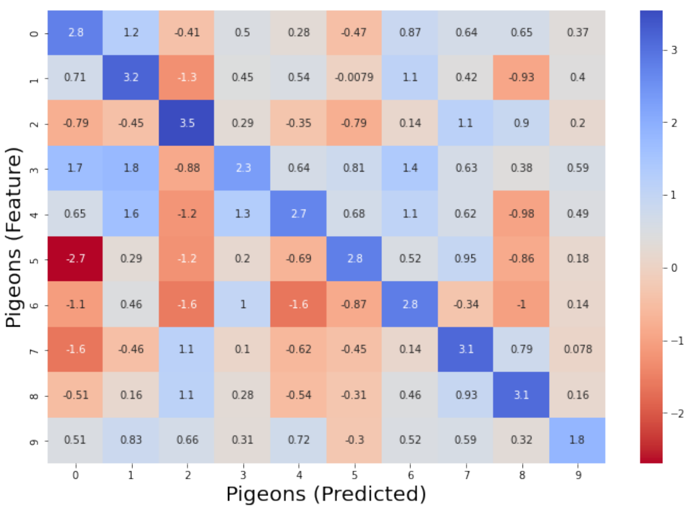

Fig. 3. Weighted Attention Scores of the First Free Flight of Group B

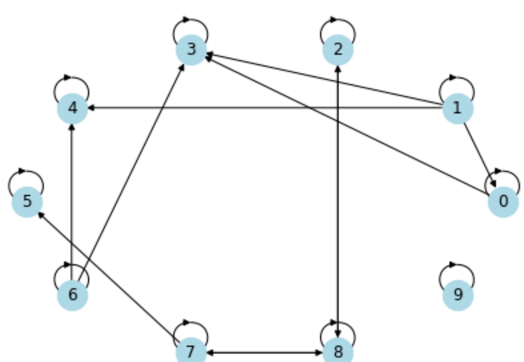

Fig. 4. Causal Graph Structure of the First Free Flight of Group B

This finding has implications for the causal inference stage of the study. According to Nauta's TCDF, a feature must have an attention score greater than 1 to be considered a potential cause. However, applying this threshold in this study would result in the pigeon's velocity being the only cause attributed to the pigeon itself. To overcome this limitation, the researchers decided to lower the threshold to 0.5, an arbitrary value, to produce causal graph structures. This allows for the discovery of weak influences among the pigeons' decision-making processes, even if the majority of the decision-making is based on the pigeon's own information.

Moreover, the researchers noted a potential relationship between the physical proximity of pigeons and their connectivity in the causal graph structures. Specifically, they observed that pigeon pairs that were connected by edges were also closer in physical distance. While this relationship is yet to be mathematically proven, it suggests that proximity of pigeons may play a role in the dynamics of their interactions. This observation could imply that pigeons that are closer together may be more likely to interact and potentially form stronger social connections. The causal graph structure in Figure 4 depicts the relationships of pigeons from Group B's flight one. However, it is not necessarily the case that the physical proximity of pigeons always implies their connectivity in the causal graph.

The adjacency matrix in Table III shows the proximity between the pigeons by indicating whether they were close to each other or not, based on the overall average pairwise distances. The matrix confirms that the graph formed during this flight has edges between nodes (pigeons) that are close to each other. In this matrix, a value of 1 indicates that two pigeons are close in distance with each other, meaning that their distance is less than the average of all pairwise distances, implying close proximity between the pigeons. Conversely, a value of 0 indicates that the pair is not close in distance.

TABLE III
DISTANCE INDICATOR FOR THE FIRST FREE FLIGHT OF GROUP B

Pigeon	0	1	2	3	4	5	6	7	8	9
0	1	1	0	1	1	1	1	1	0	0
1	1	1	0	1	1	1	1	1	0	0
2	0	0	1	0	0	0	0	1	1	0
3	1	1	0	1	1	1	1	1	0	0
4	1	1	0	1	1	1	1	1	0	0
5	1	1	0	1	1	1	1	1	0	0
6	1	1	0	1	1	1	1	1	0	0
7	1	1	1	1	1	1	1	1	1	0
8	0	0	1	0	0	0	0	1	1	0
9	0	0	0	0	0	0	0	0	0	1

Furthermore, it was found that a consistent most influential pigeon could not be identified by using different time windows. This was determined by analyzing the causal graph structures obtained in each time window. The concept of intermittent switching was used to support this finding, as it suggests that leadership may not be consistently held by a single entity, but rather can be switched between multiple entities in a dynamic and unpredictable manner [22], [35]. This coincides with the observed lack of a consistent leader in this study. For example, in Figure 5, it can be observed that in a span of around 28 minutes, leaders were found to be different 8 times throughout the flight supporting the concept of intermittent switching.

The intermittence and self-reliance of pigeons suggest that

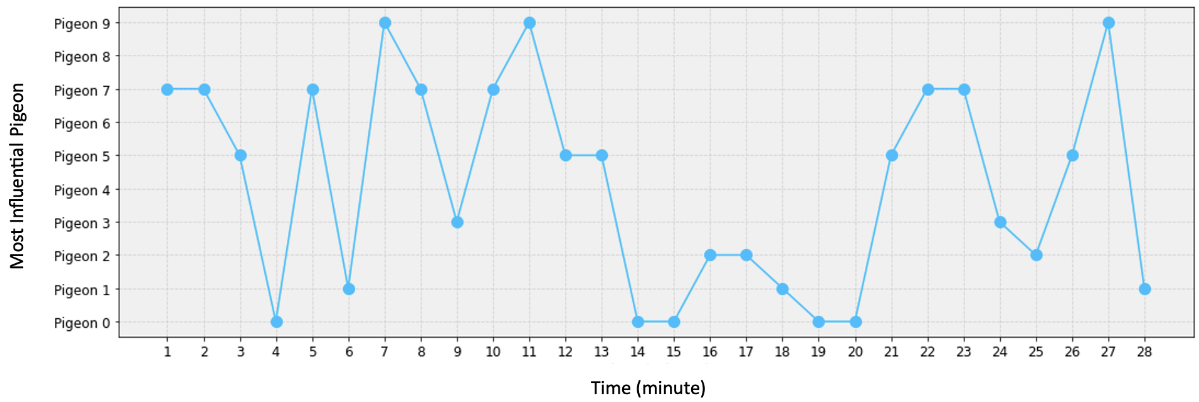

Fig. 5. Intermittent Switching Illustration of First Free Flight of Group B

collective behavior among pigeon flocks may not be driven by a dominant leader, but rather by individual interactions among the pigeons. This indicates that pigeons exhibit a high degree of autonomy and decision-making ability within flocks. They have been observed to be capable of complex problem-solving and decision-making, as well as communication , which allows them to collectively make decisions and in a coordinated manner.

Ultimately, these results provide important insights into the dynamics of pigeon flocks and may challenge current understanding of leader-follower relationships in animal groups. This finding is in line with other studies that also found that the concept of leadership may not be as prevalent as previously thought [36]. This highlights the need for further research to better understand the complex dynamics of animal groups and to consider multiple factors when studying causal relationships in collective animal behavior.

VI. Conclusion and Recommendations

The use of modified AD-DSTCN for causal inference of collective animal behavior has shown promising results, such as the observations of self-reliant decision making, connectivity, and intermittent switching. However, it is acknowledged that these observations require further verification through the use of more sophisticated mathematical methods in order to generalize the findings. It is also important to note that there are many ways to analyze the causal graph structures. For instance, although edges represent causal influences, it is possible to find the most influential node in different ways using other graph theory concepts. Hence, the need to further verify the results with other existing hypotheses to provide additional validation of the findings. Finally, the researchers recommend to apply the causality inference method presented in this paper to other animal species.

Acknowledgment

The authors are grateful to colleagues M. Nagy, Z. Akos, D. Biro, and T. Vicsek for providing the data used in this study. The authors likewise thank the various offices of the Ateneo de Manila University: Office of the Associate Dean for Research and Creative Work, School of Science and Engineering, Loyola Schools, and the Department of Mathematics for all their support.

References

[1] J. Ge, C. Fan, C. Yan, and L. Wang, "Multi-uavs close formation control based on wild geese behavior mechanism," in *2019 Chinese Automation Congress (CAC)*. IEEE, 2019, pp. 967–972.

[2] A. Sokolov, I. S. Aranson, J. O. Kessler, and R. E. Goldstein, "Concentration dependence of the collective dynamics of swimming bacteria," *Physical review letters*, vol. 98, no. 15, p. 158102, 2007.

[3] C. A. Yates, R. Erban, C. Escudero, I. D. Couzin, J. Buhl, I. G. Kevrekidis, P. K. Maini, and D. J. Sumpter, "Inherent noise can facilitate coherence in collective swarm motion," *Proceedings of the National Academy of Sciences*, vol. 106, no. 14, pp. 5464–5469, 2009.

[4] B. L. Partridge, "The structure and function of fish schools," *Scientific american*, vol. 246, no. 6, pp. 114–123, 1982.

[5] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, "Statistical mechanics for natural flocks of birds," *Proceedings of the National Academy of Sciences*, vol. 109, no. 13, pp. 4786–4791, 2012.

[6] G. Dieck Kattas, X.-K. Xu, and M. Small, "Dynamical modeling of collective behavior from pigeon flight data: Flock cohesion and dispersion," *PLoS computational biology*, vol. 8, no. 3, p. e1002449, 2012.

[7] M. Milinski and R. Heller, "Influence of a predator on the optimal foraging behaviour of sticklebacks (gasterosteus aculeatus l.)," *Nature*, vol. 275, no. 5681, pp. 642–644, 1978.

[8] T. Pitcher, A. Magurran, and I. Winfield, "Fish in larger shoals find food faster," *Behavioral Ecology and Sociobiology*, vol. 10, no. 2, pp. 149–151, 1982.

[9] D. Biro, D. J. Sumpter, J. Meade, and T. Guilford, "From compromise to leadership in pigeon homing," *Current biology*, vol. 16, no. 21, pp. 2123–2128, 2006.

[10] J. K. Parrish, S. V. Viscido, and D. Grunbaum, "Self-organized fish schools: an examination of emergent properties," *The biological bulletin*, vol. 202, no. 3, pp. 296–305, 2002.

[11] S. Yin, H. Li, and J. Liu, "New pigeon flocks behavior used for multiple robots autonomous formation," *Journal of Software Engineering*, vol. 11, pp. 210–216, 2017.

[12] J. Werfel, K. Petersen, and R. Nagpal, "Designing collective behavior in a termite-inspired robot construction team," *Science*, vol. 343, no. 6172, pp. 754–758, 2014.

[13] E. Bonabeau, "Swarm intelligence," *A Primer on Multiple Intelligences*, p. 211, 2003.

[14] Y. Tan and Z.-y. Zheng, "Research advance in swarm robotics," *Defence Technology*, vol. 9, no. 1, pp. 18–39, 2013.

[15] D. Goodwin, "The significance of some behaviour patterns of pigeons," *Bird Study*, vol. 3, no. 1, pp. 25–37, 1956.

[16] Z. Arad, I. Gavrieli-Levin, and J. Marder, "Adaptation of the pigeon egg to incubation in dry hot environments," *Physiological Zoology*, vol. 61, no. 4, pp. 293–300, 1988.
[17] E. Schleucher, "Heterothermia in pigeons and doves reduces energetic costs," *Journal of Thermal Biology*, vol. 26, no. 4-5, pp. 287–293, 2001.
[18] S. Watanabe and N. F. Troje, "Towards a "virtual pigeon": a new technique for investigating avian social perception," *Animal Cognition*, vol. 9, pp. 271–279, 2006.
[19] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, "Hierarchical group dynamics in pigeon flocks," *Nature*, vol. 464, no. 7290, pp. 890–893, 2010.
[20] X.-K. Xu, G. D. Kattas, and M. Small, "Reciprocal relationships in collective flights of homing pigeons," *Physical Review E*, vol. 85, no. 2, p. 026120, 2012.
[21] H.-T. Zhang, Z. Chen, T. Vicsek, G. Feng, L. Sun, R. Su, and T. Zhou, "Route-dependent switch between hierarchical and egalitarian strategies in pigeon flocks," *Scientific reports*, vol. 4, no. 1, pp. 1–7, 2014.
[22] D. Chen, T. Vicsek, X. Liu, T. Zhou, and H.-T. Zhang, "Switching hierarchical leadership mechanism in homing flight of pigeon flocks," *EPL (Europhysics Letters)*, vol. 114, no. 6, p. 60008, 2016.
[23] D. Chen, Y. Wang, G. Wu, M. Kang, Y. Sun, and W. Yu, "Inferring causal relationship in coordinated flight of pigeon flocks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 11, p. 113118, 2019.
[24] A. Strandburg-Peshkin, D. Papageorgiou, M. C. Crofoot, and D. R. Farine, "Inferring influence and leadership in moving animal groups," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1746, p. 20170006, 2018.
[25] A. K. Engel, P. König, C. M. Gray, and W. Singer, "Stimulus-dependent neuronal oscillations in cat visual cortex: inter-columnar interaction as determined by cross-correlation analysis," *European Journal of Neuroscience*, vol. 2, no. 7, pp. 588–606, 1990.
[26] C. W. Granger, "Testing for causality: A personal viewpoint," *Journal of Economic Dynamics and control*, vol. 2, pp. 329–352, 1980.
[27] R. Q. Quiroga, T. Kreuz, and P. Grassberger, "Event synchronization: a simple and fast method to measure synchronicity and time delay patterns," *Physical review E*, vol. 66, no. 4, p. 041904, 2002.
[28] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.
[29] V. Mwaffo, J. Keshavan, T. Hedrick, and S. Humbert, "A data-driven method to dissect the dynamics of the causal influence in complex dynamical systems," in *2018 IEEE Workshop on Complexity in Engineering (COMPENG)*. IEEE, 2018, pp. 1–5.
[30] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery." *Queue*, vol. 16, no. 3, pp. 31–57, 2018.
[31] M. Nauta, D. Bucur, and C. Seifert, "Causal discovery with attention-based convolutional neural networks," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 312–340, 2019.
[32] M. Nagy, G. Vásárhelyi, B. Pettit, I. Roberts-Mariani, T. Vicsek, and D. Biro, "Context-dependent hierarchies in pigeons," *Proceedings of the National Academy of Sciences*, vol. 110, no. 32, pp. 13 049–13 054, 2013.
[33] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
[34] S. Bai, J. Z. Kolter, and V. Koltun, "Convolutional sequence modeling revisited," 2018.
[35] V. Mwaffo, J. Keshavan, T. L. Hedrick, and S. Humbert, "Detecting intermittent switching leadership in coupled dynamical systems," *Scientific Reports*, vol. 8, no. 1, p. 10338, 2018.
[36] M. Bourjade, B. Thierry, M. Hausberger, and O. Petit, "Is leadership a reliable concept in animals? an empirical study in the horse," *PLoS One*, vol. 10, no. 5, p. e0126344, 2015.