A compass towards equity: a data analysis framework to capture children's behaviour in the playground context

Nasri, M.

CHAPTER 6

Modeling Group Behaviour via Spatio-temporal Data

The contents of this chapter are based on the following publications (with permission from Springer Nature):

- M. Nasri, Z. Fang, M. Baratchi, G. Englebienne, S. Wang, A. Koutamanis, and C. Rieffe, "A gnn-based architecture for group detection from spatio-temporal trajectory data," in *International Symposium on Intelligent Data Analysis*, pp. 327–339, Springer, 2023 DOI: 10.1007/978-3-031-30047-9_26.
- M. Nasri, T. Maliappis, C. Rieffe, and M. Baratchi, "T-dante: Detecting group behaviour in spatio-temporal trajectories using context information," in *International Symposium on Intelligent Data Analysis*, pp. 28–39, Springer, 2024 DOI: 10.1007/978-3-031-58553-1_3.

Abstract

Modeling group behavior can be used in schoolyards to understand children's social behavior and their interactions with the social environment. Several studies have addressed the problem of identifying group behavior through modeling spatio-temporal trajectories. In this chapter, we revisit this problem by conducting two parallel studies using dyad-based models and context-based models. In this context, children are modeled as nodes in a graph. Such a graph representation indicates, for example, the social network of children in a class or playgroup. Our proposed dyad-based model, i.e., WavenetNRI, models interactions between each pair of nodes (i.e., dyadic nodes) using their spatio-temporal trajectories. Whereas our proposed context-based model, i.e., T-DANTE, includes the context information, i.e., it models interactions of dyadic nodes by additionally considering the spatio-temporal trajectories of surrounding nodes. We conducted our experiments using two collections of datasets, namely Opentraj datasets (with five real-world pedestrian datasets) and spring simulation dataset (with five simulation datasets), and two evaluation metrics, i.e., group mitre and group correctness. Our experiments compare the performance of these two models with two other baselines. The results demonstrate that including context information can improve the accuracy of group behavior modeling in Opentraj datasets. Meanwhile, in the simulation dataset, which includes groups with larger sizes, the dyad-based model performs better than other models.

6.1 Introduction

Children constantly interact with their social environment (i.e., peer groups and teachers) in schoolyards through different games and activities [1]. The social environment provides opportunities for children to develop their social skills. Yet, these environments may include barriers (e.g., environments in which children experience ostracism by their peers) that hinder social development for certain children. Addressing the existing barriers in the social environment is crucial to prevent problems such as bullying and social exclusion and promote emotional well-being in children [216].

The primary step in identifying these barriers is understanding children's social behavior and their group formations in schoolyards. The main challenge in obtaining this understanding is the variability of children's interactions over time and space. For instance, children often change their interacting partners or groups over

recess time [35], or they often use several areas of schoolyards during group interactions [217]. Thus, accurately capturing children's group behavior in schoolyards is only possible when both spatial and temporal elements are considered in the design of the modeling framework [2].

Excessive literature in children's research has studied social behavior by measuring face-to-face interactions [218, 219]. In this form of interaction, children are physically present with one another within close proximity [77]. Similarly, in the context of the present thesis, proximity tags were initially adopted to capture children's face-to-face interactions in schoolyards (see Chapters 2 - 4). Although capturing face-to-face interactions provides an informative and straightforward measure to understand children's social behavior, it overlooks other forms of interactions that commonly happen in schoolyards. In order to create a clearer picture of children's social behavior, in addition to face-to-face interactions, we also captured parallel interactions, e.g., walking and running side-by-side, in schoolyards (see Chapter 5).

Yet, the complex nature of schoolyard activities might involve more complicated forms of interactions not captured in face-to-face contact or parallel movements. Understanding these complex group interactions requires analyzing children's behavior in their social networks in schoolyards. From the data science perspective, analyzing group behavior in a social network can turn into a mathematical problem: how to identify sub-groups (or sub-graphs) in a given community (or graph). To this end, various studies in social network analysis focused on designing community detection algorithms that identify sub-communities based on pairwise relationships among individuals [220–222]. Despite their great performance in identifying static groups, these algorithms might not be able to identify groups in scenarios where group formation dynamically changes over time and space. For example, when a group of children is playing hide and seek, depending on their role, they might be involved in social interactions that can only be revealed by analyzing the spatio-temporal dynamics of children in forming a group, i.e., how children move in space over time compared with their peers. Including these spatio-temporal dynamics is essential to analyze social interactions and group formations in a higher resolution, going beyond face-to-face interaction and parallel play.

The first attempts to address this challenge focused on classical machine learning models, which incorporate a manual feature extraction process to find the most significant features [223, 224]. Although the results show promise, the manual feature extraction and selection process is often time-consuming and might potentially introduce bias to the model. Recent studies in the field of artificial intelligence focused on developing neural network models to automatically extract features and identify sub-groups based on individual interaction graphs [225–228]. These pipelines typically incorporate spatio-temporal data to train a neural network model and re-

construct an affinity graph, i.e., a graph that represents the pairwise relationship of individuals. Applying a community detection algorithm or clustering method can identify sub-communities within this reconstructed graph. Previous studies focused specifically on spatial features by adopting multi-layer perception MLP layers [225] or 1-dimensional convolutional layers [226] to model group behavior, overlooking the temporal dependencies that might contribute to modeling group formations.

To address this gap, the current chapter revisits the problem of group behavior modeling in spatio-temporal data via neural network models by conducting two parallel studies: (1) WavenetNRI [3], built upon NRI [226], and (2) T-DANTE [4], building upon DANTE [225], to create an affinity graph via a neural network model that can be used by a community detection algorithm to identify groups in a given community. Overall, this chapter includes the following subjects:

- Discussing two novel neural network frameworks, i.e., WavenetNRI [3] and T-DANTE [4], to address group modeling tasks using spatio-temporal data.
- Discussing a trajectory simulation framework, built upon spring simulation framework [226, 229], to stimulate group and non-group interactions among particles in a physical system. This framework uniquely simulates attraction points (i.e., points where group members often mingle around) to stimulate group movements.
- Evaluating the performance of the two models, i.e., WavenetNRI and T-DANTE, using two sets of datasets, namely Opentraj dataset (includes five pedestrian datasets) and spring simulation dataset (with five simulation datasets) against two baselines (i.e., NRI [226], and DANTE [225]) via two evaluation metrics, i.e., Group Mitre and Group correctness.

The present study is organized as follows. The related literature is presented in Section 6.2. The group modeling problem is defined in Section 6.3. Section 6.4 presents the details of the proposed approach in pairwise information and context information. Section 6.5 presents the datasets, the evaluation metrics, baselines, and implementation details adopted in our experiments. Moreover, this section discusses the results of our experiments. Finally, Section 6.6 summarizes the study and points out the limitations and directions for future research.

6.2 Related Work

The spatio-temporal data adopted in modeling group behavior can be categorized into two areas: **Dyad-based modeling** and **Context-based modeling**. In dyad-

based modeling studies, the spatio-temporal data per pair of nodes in the interaction graph focuses on training their model and predicting the affinity score. Studies in context-based modeling included the spatio-temporal data of the surrounding nodes, i.e., social context, in addition to the pairwise interaction data to predict the affinity score. The following section discusses the existing literature on adopting these two strategies.

6.2.1 Dyad-based Modeling

Various studies in this area adopted graph-based neural networks (GNN) to estimate pairwise interactions among agents [226–228]. Thompson et al. [230] modeled a scene as an interaction graph, where nodes and edges represent individuals and their dyadic relationship, respectively. GNN is used afterward to predict the pairwise affinity, indicating the likelihood of pairwise interactions. In another attempt, Kipf et al. [226] proposed Neural Relational Inference (NRI), which predicts interactions between moving particles using spatio-temporal data. Both studies assumed interactions among specific pairs of agents remain constant throughout the entire timeframe. Yet, in real-world social settings, individuals often change their interaction partners.

Moreover, they both overlook the symmetric group relationships among pairs in the affinity graph. Implementing this feature satisfies the following condition in the embedding space (where the affinity graph is reconstructed): if A is in a group with B, B is also in the same group with A, to account for bidirectional relations in group memberships. The dyad-based model, WavenetNRI, adopts the dilated residual causal convolutional (GD-RCC) block [231] to capture short and long dependencies in spatio-temporal dynamics. Moreover, it uses symmetric temporal edge features and a symmetric edge updating process to address the symmetric property of group relationships.

6.2.2 Context-based Modeling

This line of research incorporates the context information, i.e., the surrounding agents, in addition to the dyad information, into the model's design. The underlying reason is that determining whether two individuals belong to the same group does not solely depend on the behavior of those two individuals. Additionally, the behavior of surrounding individuals could also impact this determination, e.g., in the context of the schoolyard, identifying two children running together as a group will be easier by including the fact that other children are playing in the sandpit as it suggests significant spatio-temporal differences between the two groups. In line

with this idea, Swofford et al. [225] introduced DANTE, a neural network model that incorporates MLP layers. DANTE adopts the information of surrounding agents in addition to the pairwise interactions to learn graph representation for a single-frame scene. In another attempt, Lu et al. [232] introduced VGDTN, which adopts 1-dimensional convolutional layers to identify group behavior using both dyad and context information in a single-frame scene. Yet, the focus of both studies on a single-frame scene might overlook the temporal features that occur over multiple timeframes. Moreover, MLP models and 1-dimensional convolutionals are known to be incapable of effectively modeling time-series data such as spatio-temporal trajectories compared with recurrent neural network (RNN) models. To address this, the context-based model, T-DANTE, enhances the context information by including scenes with multiple timeframes instead of a single scene. Moreover, by implementing RNN layers, T-DANTE aims to capture short and long dependencies in the spatio-temporal data.

6.3 Problem Formulation

Consider a dataset D that includes the movement trajectories of M agents. Each movement trajectory $X_m = \{x_1, ..., x_t, ..., x_T\}$ indicates a consecutive sequence of spatio-temporal features x_t of agent m, $m \in (1, M)$ over a timeframe with T time steps, $t \in (1, T)$. Each dyadic agent may interact with the others over the given timeframe.

Firstly, we are interested in estimating the pair-wise relationships between dyadic agents by learning the affinity score $h^2_{(i,j)}$ between all dyadic agents i and j and assembling all scores to form an affinity graph. Secondly, we are interested in identifying groups $C = \{c_j | j \in [1, K]\}$ of agents in the created affinity graph ($1 \leq K \leq N$ is the number of groups) under three main assumptions: (1) the group relationships are constant in a time window, while agents could interact with other agents from a different group. (2) Agents of the same group share similar spatial behavior over a timeframe. (3) The size of the timeframe is fixed across the measurements.

The present chapter discusses two approaches to address this problem: (1) the dyad-based modeling and (2) the context-based modeling. The dyad-based model learns interactions between dyadic nodes using their spatio-temporal trajectories. Meanwhile, the context-based model learns dyadic node interactions by considering surrounding nodes' spatio-temporal trajectories.

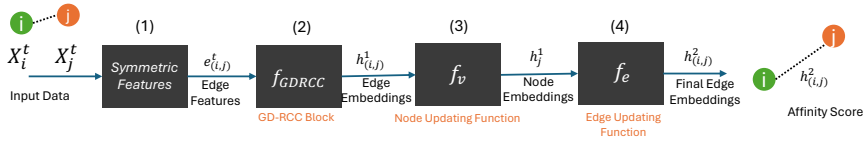

Figure 6.1: An overview of WavenetNRI framework. (1) The symmetric edge feature will be created based on the spatio-temporal of nodes i and j. (2) The edge embeddings will be created by applying a GD-RCC block to the symmetric edge feature sequences. (3) All edge embeddings will be aggregated per node j to obtain the node representation. (4) The node embedding and edge embeddings will be used in this block to create the final edge embeddings, i.e., affinity score, between node i and j.

6.4 Methodology

This section presents the design of two models, i.e., WavenetNRI and T-DANTE. These models identify group behavior by learning the affinity graph from spatio-temporal data of nodes (nodes can be agents or individuals depending on the context). While WavenetNRI focuses on extracting complex spatio-temporal dependencies based on dyad information, T-DANTE adopts an RNN model to identify group behavior using both dyad and context information. The details of these two models are described as follows:

6.4.1 WavenetNRI Framework

The WavenetNRI models group behavior solely based on dyad information. The design of this model is inspired by NRI framework [226] and Wavenet framework [231]. To satisfy the symmetric feature of group membership, WavenetNRI implements symmetric edge features and symmetric edge updating functions to account for the bidirectional nature of group memberships. Moreover, WavenetNRI adopts GD-RCC to learn short and long-term spatio-temporal dependencies in the edge feature. The overview of the WavenetNRI framework is depicted in Figure 6.1. This framework consists of four blocks, each representing one step in the training process as follows:

Step 1. Symmetric Edge Features. In the first step, the symmetric edge feature sequences $e^t_{(i,j)}$ will be created using the spatio-temporal data of dyad nodes i and j. The original NRI implements this by simply concatenating the spatio-temporal data of dyad nodes i and j (i.e., $e^t_{(i,j)} = [X^t_i, X^t_j]$). Built on this idea, WavenetNRI implements symmetric edge sequences to satisfy the symmetric nature of group relationships. The edge feature sequences in WavenetNRI are constructed by concatenating the pairwise distances and temporal increments in spatio-temporal data per dyad as follows:

$$e^t_{(i,j)} = [\|X^t_i - X^t_j\|, \Delta X^t_i \odot \Delta X^t_j], \quad t \in 1, ..., T-1, \quad \Delta X^t_i = X^{t+1}_i - X^t_i \tag{6.1}$$

Where $\|X^t_i - X^t_j\|$ denotes the Euclidean distance between dyad nodes i and j and $\Delta X^t_i \odot \Delta X^t_j$ denotes the element-wise production of the increments of dyads. Thus, edge feature sequence $e^t_{(i,j)}$ captures both spatial and temporal differences between each dyad. Moreover, the edge features are symmetric, i.e., $e^t_{(i,j)} = e^t_{(j,i)}$, corresponding to the symmetric properties of pairwise group relationships.

Step 2. GD-RCCC Block. The edge feature sequences $e^t_{(i,j)}$ obtained in the previous step will be given to the GD-RCC block to extract spatio-temporal features of the given edge (i.e., edge embeddings). The original NRI adopts one convolutional layer that may not efficiently capture the long-term interactions of edge feature sequences. In the WavenetNRI, a GD-RCC block [231] inspired by Wavenet [231] is used to transform the edge feature sequences $e^t_{(i,j)}$ into the edge embedding $h^1_{(i,j)}$ as formulated in Equation 6.2. The use of GD-RCC block has several advantages for the design of WavenetNRI: (1) The causal convolution maintains the order of the timely ordered edge sequences, (2) the dilated convolutional kernels exponentially expand the receptive fields, (3) the skip connection (i.e., 1D CNN) tackles the gradient vanishing problem, and (4) the gating activation function regulates the information flow.

$$h^1_{(i,j)} = f_{GDRCC}(e^t_{(i,j)}) \tag{6.2}$$

Step 3. Node Updating Function. This step sums up all edge embeddings $h^1_{(i,j)}$ for dyadic nodes i and j and gives them to the node updating function f_v (as proposed in the original NRI). This function generates higher level node embedding h^1_i (or h^1_j) for dyadic nodes i and j as formulated in Equation 6.3.

$$h^1_j = f_v(\sum_{i \neq j} h^1_{(i,j)}) \tag{6.3}$$

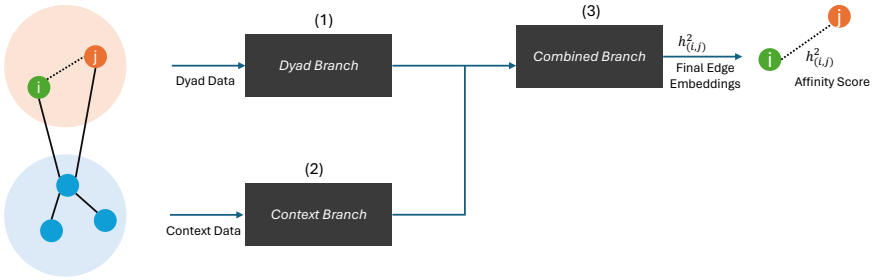

Figure 6.2: An overview of T-DANTE framework: The dyad branch extracts the spatio-temporal data of the pair of nodes of interest. The context branch extracts the spatio-temporal data of surrounding nodes, i.e., social context. The extracted features from the dyad branch and context branch will be merged in the combined branch. The output of this branch, i.e., the affinity score $h^2_{(i,j)}$, will be calculated per dyads to create the affinity graph.

Step 4. Symmetric Edge Updating Function. The element-wise production of the obtained node embeddings will be concatenated by the feature embeddings $h^1_{(i,j)}$ obtained in Step 2 and fed to the neural network f_e to get final edge embedding $h^2_{(i,j)}$ between dyadic node i and j, represented as the affinity score between the two nodes (See Equation 6.4). Through this process, the final affinity score $h^2_{(i,j)}$ captures interactions between dyadic nodes i and j and their interactions with other nodes [226].

$$h^2_{(i,j)} = f_e([h^1_{(i,j)}, h^1_i \odot h^1_j]) \tag{6.4}$$

During the supervised training phase, the ground-truth pairwise group relationships $G_{(i,j)}$ are used as labels. Since the datasets include an imbalanced distribution of the labels, the weighted cross-entropy is adopted as a loss function. This loss function assigns higher weights to the rare labels to compensate for their lower distribution. By minimizing the weighted cross-entropy, WavenetNRI is optimized to identify the "interaction" versus "no interaction" relation between nodes. After supervised training, the affinity graph will be constructed by assembling all the obtained affinity scores between pairs of nodes. The Louvain community detection algorithm [222] is applied afterward to the obtained affinity graph to find sub-groups among given nodes.

6.4.2 T-DANTE Framework

While the previous approach, WavenetNRI, solely focuses on the spatio-temporal data of the dyadic nodes, the T-DANTE framework additionally includes the spatio-temporal data of the surrounding nodes, i.e., social context. This section presents the details of T-DANTE, inspired by the DANTE framework [225]. T-DANTE extends DANTE by utilizing RNN blocks to retain temporal information and spatial features. During the training process, first, these spatio-temporal features will be extracted from the data to estimate the affinity score between pairs of nodes. Then, this affinity score will be compared with the pairwise group membership (i.e., ground truth $G_{(i,j)}$) using the log loss function. Assembling the estimated affinity score between all pairs of nodes creates an affinity graph, which can be used afterward in a community detection algorithm to detect sub-groups in the given data. The proposed T-DANTE consists of three branches: (1) Dyad Branch, (2) Context Branch, and (3) Combined Branch (See Figure 6.2). The dyad branch, similar to the WavenetNRI framework, captures local spatio-temporal information from the dyadic nodes. The context branch captures the spatio-temporal information from surrounding nodes (i.e., social context). The combined branch combines the output of these two branches and estimates the affinity score between the pair of nodes that are given in the dyad branch. The details of these branches are explained as follows:

1. Dyad Branch. The Dyad Branch extracts the spatio-temporal features of dyadic nodes using RNN layers, i.e., LSTM layers. The LSTM includes memory cells and gating mechanisms, to selectively store and retrieve information over long sequences, e.g., time series data such as movement trajectories. A series of convolutional layers are then applied to concatenate the spatio-temporal features extracted from RNN layers. Lastly, the dyad branch is followed by a Dropout layer to reduce overfitting and a Batch Normalisation layer to avoid the covariate shift and enhance the model's generalizability.

2. Context Branch. Context Branch follows the same identical design as the Dyad Branch. Yet, its given input data and role in the overall framework are different. The Context Branch extracts the spatio-temporal features of the surrounding nodes to account for context information. The number of surrounding nodes is a hyperparameter of the model (i.e., context size).

3. Combined Branch. The Combined Branch merges the extracted spatio-temporal features obtained from the Dyad Branch and Context Branch together.

Specifically, the extracted features are first flattened and passed through a series of fully connected layers, dropout layers, and batch normalization layers. Their specifications (e.g., number of layers, kernels, and filter size) depend on the characteristics of the dataset, such as the number of frames, the maximum number of nodes, and the data size. The last layer of this branch is a fully connected layer with a Sigmoid activation function to constrain the single output to the $[0, 1]$ range. This output is the affinity score for the dyadic nodes.

Assembling all the affinity values between all dyadic nodes creates the affinity graph. The group structures in the affinity graph will be identified afterward using the Dominant Sets (DS) community detection algorithm [233].

6.5 Experiments

We conducted several experiments to evaluate the models' performance. The following sections describe the datasets and evaluation metrics we used in our experiments. Furthermore, the baselines used to compare with the performance of the WavenetNRI and T-DANTE models are explained. Lastly, the implementation details of our experiments and the obtained results are presented. Our results compare the performance of the proposed frameworks, WavenetNRI and T-DANTE, with two other baselines using two evaluation metrics on two collections of datasets. The goal of this experiment is to address the following research questions:

- RQ. 1. Can symmetric edge features and GD-RCC block improve the performance of WavenetNRI compared with the original NRI model?
- RQ. 2. Can RNN block and including multiple timeframes per scene improve the performance of T-DANTE compared with the original DANTE model?
- RQ. 3. Which of the dyad-based or context-based models can perform better in identifying group behavior?

6.5.1 Datasets

In order to evaluate the performance of the models, we conducted our experiments using two publicly available datasets, i.e., Opentraj Dataset and Spring Simulation Dataset. Table 6.1 presents the characteristics of these datasets.

- **Opentraj Dataset.** Opentraj dataset [234][1] is extensively used in human trajectory prediction literature. This dataset includes the trajectories of pedestrians, location, and velocity in multiple timeframes, captured via static camera. Five pedestrian datasets, *eth*, *hotel* [209], and *zara01*, *zara02* and *students03* [210], which include the ground truth of the group membership, have been used in our experiments. This ground truth is created by annotating the pedestrians who seemed to walk in groups. The original dataset includes location data relative to the world reference W. In order to enhance the generalizability of our approach across different datasets, the trajectory of each pedestrian is transformed to a local coordinate system L_{ij}, defined as the middle point between pedestrian i and j.

- **Spring Simulation Dataset.** The spring simulation framework, built upon previous studies [226, 229], is developed to simulate group and non-group interactions among particles in a physical system. In line with the original studies, our spring simulation framework simulates the movements of groups of particles in a 2-D space. In their movements, those from the same group attract each other and distract from particles from another group. The locations, velocities, and the group membership (i.e., ground truth) of the particles are included in this simulation. Our proposed framework has made two improvements to the original framework: (1) defining group size as a simulation parameter that can be controlled over different experiments and (2) designing attraction points that stimulate particles from the same group toward certain pre-defined spots. In the proposed framework, pre-defined forces stimulate particles toward attraction points. All forces have the same strength, but their direction is different to point a particle towards a certain attraction point.

6.5.2 Evaluation Metrics

In order to evaluate the performance of the proposed model, we used two evaluation metrics, i.e., Group Mitre [235], Group Correctness [225, 233], in our experiments. In both evaluation metrics, due to the higher number of pairs from different groups compared to those from the same group, we adopted an F-1 score as it is more suitable for evaluating imbalanced datasets.

- **Group Mitre (G_M) [235]** is an evaluation metric, built upon the Mitre loss [236], has been used by several studies [229, 237, 238] to measure the

[1] https://github.com/crowdbotp/OpenTraj

Table 6.1: Characteristics of five Opentraj datasets and five spring simulation datasets used in both models, regarding the duration of measurements, pedestrian dataset in seconds and spring simulation dataset in timeframes, the number of agents, and the number of groups.

	Dataset	Duration	Agents#	Groups#
Opentraj	eth	773.4	360	58
	hotel	722.4	390	41
	zara01	360.4	148	45
	zara02	420.4	204	58
	students03	215.6	428	101
Simulation	sim_1	50	8	2
	sim_2	50	9	2
	sim_3	50	9	3
	sim_4	50	10	2
	sim_5	50	10	4

quality of the identified groups. Mitre loss adopts spanning trees to represent groups. This form of representation overlooks singletons, i.e., a group with only one node. Group Mitre solves this problem by adding a fake counterpart to each node. This fake node is considered in the same group as the original node only if the original node was singleton. The detailed implementation of G_M is presented by Solera et al. [235].

- **Group Correctness (G_c) [225, 233]** considers a group as correctly identified if at least $P * |c_j|$ of its members are correctly classified in the group, where $P \in [0, 1]$ is a threshold and $|c_j|$ indicates the size of the original group j. The $P = 1$ requires all agents in ground truth group membership data to be correctly identified in group j. Accordingly, $P < 1$ applies a milder metric in evaluating the quality of the identified groups.

6.5.3 Baselines

In the comparative study, we compared the performance of our proposed methods with two other baseline methods, namely NRI [226] as a dyad-based model and DANTE [225] as a context-based model. These baseline methods are described in Section 6.2 and are implemented based on their available source code. The original studies [4, 229] include extensive experiments where more baselines and dataset configurations have been presented. In this chapter, the most relevant approaches has been selected and presented in the result section for consistency and improving

Table 6.2: The results of Group Correctness G_C and Group Mitre G_M for WavenetNRI and T-DANTE compared with baselines using Opentraj datasets and simulation datasets. The * sign shows that this result is significantly different compared with other cases under the same evaluation metric and dataset.

	Pedestrain Dataset									
	eth		*hotel*		*zara*01		*zara*02		*students*03	
	G_C	G_M	G_C	G_M	G_C	G_M	G_C	G_M	G_C	G_M
DANTE	0.319 ±0.047	0.548 ±0.019	0.431 ±0.043	**0.586 ±0.035**	0.731 ±0.051	0.793 ±0.028	0.633 ±0.038	0.705 ±0.026	0.024 ±0.012	0.502 ±0.013
NRI	0.201 ±0.062	0.571 ±0.074	0.169 ±0.054	0.540 ±0.097	0.285 ±0.067	0.597 ±0.053	0.106 ±0.035	0.417 ±0.019	0.006 ±0.010	0.280 ±0.026
WavenetNRI	0.242 ±0.059	0.553 ±0.057	0.202 ±0.048	0.455 ±0.080	0.361 ±0.091	0.627 ±0.066	0.184 ±0.065	0.462 ±0.040	0.001 ±0.004	0.280 ±0.024
T-DANTE	**0.590* ±0.030**	**0.665 ±0.017**	**0.508* ±0.043**	0.542 ±0.023	**0.821* ±0.015**	**0.838* ±0.015**	**0.870* ±0.011**	**0.873* ±0.011**	**0.696* ±0.056**	**0.780* ±0.028**
	Simulation Dataset									
	sim_1		sim_2		sim_3		sim_4		sim_5	
	G_C	G_M	G_C	G_M	G_C	G_M	G_C	G_M	G_C	G_M
DANTE	0.215 ±0.007	0.717 ±0.004	0.198 ±0.008	0.701 ±0.003	0.095 ±0.011	0.518 ±0.011	0.199 ±0.011	0.712 ±0.005	0.041 ±0.007	0.425 ±0.009
NRI	0.984 ±0.004	0.991 ±0.002	0.983 ±0.007	0.993 ±0.002	**0.988* ±0.004**	**0.995* ±0.002**	0.996 ±0.003	0.999 ±0.001	**0.988* ±0.007**	**0.995 ±0.003**
WavenetNRI	**0.996 ±0.006**	**0.998 ±0.002**	**0.995* ±0.004**	**0.998* ±0.001**	0.977 ±0.008	0.988 ±0.004	**0.998* ±0.004**	**0.999 ±0.001**	0.953 ±0.011	0.968 ±0.009
T-DANTE	0.969 ±0.002	0.983 ±0.002	0.980 ±0.002	0.989 ±0.001	0.982 ±0.006	0.988 ±0.003	0.971 ±0.006	0.987 ±0.002	0.945 ±0.011	0.976 ±0.003

the readability.

6.5.4 Implementation Details

Our experiments were implemented via the Python programming language. The details of the implementation of WavenetNRI[2] and T-DANTE (with the spring simulation framework)[3] are available in the GitHub repository. We split each Opentraj dataset into 5 folds and evaluated the performance of each method across 5 times experiments per fold (i.e., 25 runs in total per method). Since spring simulation datasets were generated under controlled conditions, they have not been split into folds. Each spring simulation dataset was randomly split into train, test, and validation datasets. The performance of each method has been evaluated across 25 times experiments per simulation dataset. We investigated the significant differences between the top two performing models by implementing the Wilcoxon signed rank test [239]. In the following sections, the performance of the proposed models against three state-of-the-art baseline methods is presented.

[2]https://github.com/fatcatZF/WavenetNRI
[3]https://github.com/ADA-research/context-group-detection

6.5.5 Results

In this section, the performance of WavenetNRI and T-DANTE is compared with the baselines. The results of the experiments for the simulation dataset and Opentraj dataset are presented in Table 6.2.

6.5.5.1 Opentraj Datasets

According to Table 6.2, WavenetNRI has outperformed NRI in most of the cases across different Opentraj datasets. This answers RQ. 1, indicating that overall, including symmetric edge features and GD-RCC block to capture more complex dependencies in the spatio-temporal data has improved the performance of the WavenetNRI.

In order to answer RQ. 2, we compared the performance of DANTE with T-DANTE. Our results show that T-DANTE is the superior model using the Group Mitre metric in all datasets, except in the hotel dataset, in which DANTE performs better. Yet, this result is not statistically significant. The superiority of T-DANTE against DANTE in most cases demonstrates that including temporal dependencies via the LSTM layers and further enriching with multiple timeframes per scene has enhanced the performance of T-DANTE. Thus, implementing LSTM layers is more suitable compared with MLPs when using the Opentraj datasets.

Finally, RQ. 3 compares the two dyad-based models, i.e., NRI and WavenetNRI, with context-based models, i.e., DANTE and T-DANTE. The result shows that T-DANTE outperforms all baselines, i.e., DANTE, NRI, and WavenetNRI, for all Opentraj datasets using the Group Correctness metric. This shows that, indeed, including context data is beneficial for modeling group behavior.

6.5.5.2 Spring simulation datasets

According to Table 6.2, inline with the result of Opentraj datasets, the capability of WavenetNRI in learning symmetric edge features and capturing complex dependencies via GD-RCC has enhanced its performance compared with NRI, addressing RQ. 1).

To address RQ. 2, we compared the performance of DANTE with T-DANTE. The results show that similar to the findings in the Opentraj datasets, including the LSTM layers and multiple timeframes in T-DANTE, have significantly improved its performance compared with DANTE across different simulation datasets using both metrics.

Finally, to address RQ. 3, we compared the performance of dyad-based models, i.e., NRI and WavenetNRI, with context-based models, i.e., DANTE and T-DANTE.

The result of this comparison shows the superiority of dyad-based models over context-based models across all simulation datasets. This contrast with the result of Opentraj datasets can be explained by the differences in the characteristics of the Opentraj and simulation datasets. Compared with Opentraj datasets, spring simulation datasets have more scenes with group sizes larger than 3 particles. This feature makes it suitable for dyad-based models, i.e., NRI and Wavenet, to extract contextual information without being limited to the number of surrounding nodes.

Overall, the results in both dyad-based and context-based models demonstrate the positive impact of capturing temporal dynamics in the data, either with GD-RCC block or LSTM layers. Moreover, the context-based models were able to more accurately model datasets with smaller group sizes, which are mainly included in the Opentraj datasets. Whereas the dyad-based models were able to more effectively extract spatio-temporal features in larger group sizes that are mainly included in the simulation datasets. Additionally, in the Opentraj dataset, pedestrians often come to the scene from one of the two ends (of streets) and leave the scene from the other end. This structured movement might create overshared trajectories between different groups and pose challenges to modeling group behaviors. Thus, including context information has enhanced the performance of these models. In the spring simulation dataset, particles freely move in a physical box, and their movement is only directed by pre-defined attraction points. This is similar to scenarios where individuals have unstructured movements with relatively mild restrictions. In these types of scenarios, dyad information provided sufficient information to model group behaviors, and adding context information did not improve the performance. For example, on university campuses or on urban pavements where individuals appear in smaller group sizes with structured movements, context-based models can more accurately identify group behavior. Whereas in the context of individuals with unstructured movements in larger groups, such as athletes on a soccer field or children in schoolyards, dyad-based models can more accurately identify group behavior. Thus, either of these models might be useful for a specific scenario, depending on its characteristics. This leads us to the necessity of implementing dynamic context size in our model to automatically define context size based on the characteristics of the given dataset or even the specific scene.

6.6 Conclusion

Analyzing children's group behavior in schoolyards enables us to identify limitations and possibilities in social environments around the child. In dynamic social settings, such as children in schoolyards, individuals constantly change their interaction part-

ners and activities, which poses extra challenges to modeling group behavior. In these scenarios, analyzing group behavior requires including both spatial and temporal elements in individuals' movements. To address this challenge, the present study aims at modeling group behavior using spatio-temporal data by conducting two parallel studies, dyad-based modeling and context-based modeling. The first study, i.e., WavenetNRI as a dyad-based model, is built up on NRI [226] and Wavenet [231] frameworks. WavenetNRI implements two features: (1) symmetric edge features with symmetric edge updating processes to account for the symmetric nature of group membership and (2) GD-RCC block to capture complex spatio-temporal dependencies in data. This model solely adopts spatio-temporal data between dyadic nodes to train the neural network model and reconstruct the affinity graph. The second study, i.e., T-DANTE as a context-based model, is built on the DANTE framework. T-DANTE adopts LSTM layers to estimate the affinity scores using the spatio-temporal data of the surrounding nodes, i.e., context information, in addition to the data of dyadic nodes. Moreover, this framework includes multiple timeframes per scene to enrich the context data.

Our comparative study against state-of-the-art baselines demonstrates that T-DANTE is the superior model for modeling group behavior using real-world Opentraj datasets. Whilst WavenetNRI outperformed other baselines in simulation datasets. The superiority of T-DANTE versus other dyad-based models, e.g., WavenetNRI and NRI, in Opentraj datasets shows that including context information has enhanced the performance of group behavior modeling in Opentraj datasets where group sizes are relatively small. Moreover, the superiority of T-DANTE over the original DANTE shows that including RNN layers can better capture spatio-temporal dependencies compared with MLP models. On the other hand, WavenetNRI has outperformed the baselines in spring simulation datasets where larger group sizes are available. The superiority of WavenetNRI over the original NRI shows that including the symmetric edge features and GD-RCC block can better capture the spatio-temporal dependencies for modeling group behavior in larger social settings via dyad information.

This finding shows that our proposed method is capable of modeling group behavior using spatio-temporal data. Moreover, the design of our models is not limited to certain from of interactions, e.g., face-to-face interactions or parallel plays. This feature enables modeling complex group behavior in higher resolution. While in small social settings with structured movements, such as pedestrians' movement on pavements, including context information is beneficial for identifying group behavior, in larger social settings with relatively unstructured movements, such as children in schoolyards, focusing on dyad interactions is sufficient to model group behavior.

Due to the limited access to group membership data for children in schoolyards,

we have only tested the performance of our proposed models on Opentraj benchmark datasets and spring simulating datasets and not on actual schoolyards. However, the Opentraj dataset has been collected from pedestrian movements in constrained environments, e.g., university campuses, which to some degree is comparable to schoolyard scenarios where children freely move in a constrained environment. Yet, applying our proposed method to children's datasets might require further investigation. For example, since children's group dynamics constantly change over time. Thus, the implementation of dynamic context size and dynamic group membership might be required in the design of the models.

Future research can explore the incorporation of dynamic context size (based on the presented number of nodes) and dynamic group membership per scene to enhance the generalization of the proposed approach across different datasets. Another future approach could be extending the proposed models in real-time applications with online data streaming. Various applications, such as analyzing students' social behavior in schoolyards, monitoring tourists' behaviors in touristic sights, and analyzing sports teams' performances, may benefit from the presented work.